



Balkan Journal of Electrical & Computer Engineering

An International Peer Reviewed, Referred, Indexed and Open Access Journal

www.bajece.com

Vol : 10
No : 3
Year : 2022
ISSN : 2147 - 284X



It is abstracted and indexed in, Index Google Scholarship, the PSCR, Cross ref, DOAJ, Research Bible, Indian Open Access Journals (OAJ), Institutional Repositories (IR), J-Gate (Informatics India), Ulrich's, International Society of Universal Research in Sciences, DRJI, EyeSource, Cosmos Impact Factor, Cite Factor, SIS Scientific Indexing Service, IJIF, iifactor. ULAKBİM-TR Dizin.

General Publication Director & Editor-in-Chief
Musa Yılmaz, Batman University, Turkey.

Vice Editor
Hamidreza Nazarpouya, University of California Riverside, USA

Scientific Committee
Abhishek Shukla (India)
Abraham Lomi (Indonesia)
Aleksandar Georgiev (Bulgaria)
Arunas Lipnickas (Lithuania)
Audrius Senulis (Lithuania)
Belle R. Upadhyaya (USA)
Brijender Kahanwal (India)
Chandar Kumar Chanda (India)
Daniela Dzhonova-Atanasova (Bulgaria)
Deris Stiawan (Indonesia)
Emel Onal (Turkey)
Emine Ayaz (Turkey)
Enver Hatimi (Kosovo)
Ferhat Sahin (USA)
Gursel Alici (Australia)
Hakan Temeltaş (Turkey)
Ibrahim Akduman (Turkey)
Jan Izykowski (Poland)
Javier Bilbao Landatxe (Spain)
Jelena Dikun (Lithuania)
Karol Kyslan (Slovakia)
Kunihiko Nabeshima (Japan)
Lambros Ekonomou (Greece)
Lazhar Rahmani (Algerie)
Marcel Istrate (Romania)
Marija Eidukeviciute (Lithuania)
Milena Lazarova (Bulgaria)
Muhammad Hadi (Australia)
Muhamed Turkanović (Slovenia)
Mourad Houabes (Algerie)
Murari Mohan Saha (Sweden)
Nick Papanikolaou (Greece)
Okyay Kaynak (Turkey)
Osman Nuri Ucan (Turkey)
Ozgun E. Mustecaplioglu (Turkey)
Padmanaban Sanjeevikumar (India)
Ramazan Caglar (Turkey)
Rumen Popov (Bulgaria)
Tarek Bouktir (Algeria)
Sead Berberovic (Croatia)
Seta Bogosyan (USA)
Savvas G. Vassiliadis (Greece)
Suwarno (Indonesia)
Tulay Adali (USA)
Yogeshwarsing Calleecharan (Mauritius)
YangQuan Chen (USA)
Youcef Soufi (Algeria)

Aim & Scope

The journal publishes original papers in the extensive field of Electrical-Electronics and Computer engineering. It accepts contributions which are fundamental for the development of electrical engineering, computer engineering and its applications, including overlaps to physics. Manuscripts on both theoretical and experimental work are welcome. Review articles and letters to the editors are also included.

Application areas include (but are not limited to): Electrical & Electronics Engineering, Computer Engineering, Software Engineering, Biomedical Engineering, Electrical Power Engineering, Control Engineering, Signal and Image Processing, Communications & Networking, Sensors, Actuators, Remote Sensing, Consumer Electronics, Fiber-Optics, Radar and Sonar Systems, Artificial Intelligence and its applications, Expert Systems, Medical Imaging, Biomedical Analysis and its applications, Computer Vision, Pattern Recognition, Robotics, Industrial Automation.



ISSN: 2147- 284X
Vol: 10
No : 3
Year: July 2022

CONTENTS

B. Küçükcaraca, B. Barutcu; Life Cycle Assessment of Wind Turbine in Turkey,.....	230 - 236
C. Andiç, A. Öztürk, S. Tosun; Dynamic Economic Dispatch with Valve-Point Effect Using Crow Search Algorithm,.....	237 - 244
A. T. Çakmak; 4 Wheel Steering System Control Unit Design,.....	245 - 251
B. Ağgöl, G. Erdemir; Development of a Counterfeit Vehicle License Plate Detection System by Using Deep Learning.....	252 - 257
M. Öztürk, A. Küçükmanisa, O. Urhan; Drowsiness Detection System Based on Machine Learning Using Eye State,	258 - 263
M. Uzun, O. Abul; Proactive Metering of Mobile Internet User Experience,.....	264 - 272
N. Ekren, A. S. Sarkin; Semi-conductor Applications to Printed Circuits on Flexible Surfaces,.....	273 - 277
H. A. Yıldız, O. Aydın; Design Consideration for Active-Only Memcapacitor Emulator Circuits,.....	278 - 285
E. Kizilkaplan, F. Yalçinkaya; The New HEMS Modelling of Human Heart,.....	286 - 294
T. Kocaer, Y. Öner; Axial Flux Motor Design for Ventilation Fans Used in The Automotive Industry,.....	295 - 299
M. Sülü, R. Daş; Graph Visualization of Cyber Threat Intelligence Data for Analysis of Cyber Attacks,.....	300 - 306
Y. Canbay, Ş. Sağıroğlu, Y. Vural; A New Anonymization Model for Privacy Preserving Data Publishing: CANON.....	307 - 316
F. Asadi; Reduction of Output Impedance of Buck Converter with Genetic Algorithm,.....	317 - 322
B. Gecer, İ. Kiyak; Wavelength tune of InGaN based blue LEDs by changing indium percentage and operational voltage variables,.....	323 - 327
E. Gündüzalp, G. Yildirim, Y. Tatar; Efficient Task Scheduling in Cloud Systems with Adaptive Discrete Chimp Algorithm,.....	328 - 336
T. Inan, U. Kacar; Ear semantic segmentation in natural images with Tversky loss function supported DeepLabv3+ convolutional neural network,.....	337 - 346
Z. Işcan; Comparison of Deep Learning and Traditional Machine Learning Classification Performance in a SSVEP Based Brain Computer Interface,.....	347 - 355

BALKAN JOURNAL OF ELECTRICAL & COMPUTER ENGINEERING
(An International Peer Reviewed, Indexed and Open Access Journal)

Contact

Batman University
Department of Electrical-Electronics Engineering
Bati Raman Campus Batman-Turkey

Web: <http://dergipark.gov.tr/bajece>
<https://www.bajece.com>
e-mail: bajece@hotmail.com

Life Cycle Assessment of Wind Turbine in Turkey

Buket Küçükkaraca and Burak Barutçu*


Abstract—This article aims to assess a life cycle analysis of a wind turbine in Turkey regarding production, transport, construction, operation, and disposal processes in terms of energy and environment. In this context, a 2-MW three-bladed horizontal axis wind turbine has been selected. Two different scenarios have been studied in this paper, which comprises wind turbines using Al-Conductor cable and wind turbines using Cu-Conductor cable. Although the cost of Al-Conductor cables is lower, their joule losses are higher. The life cycle assessment of a wind turbine includes production, transportation, construction, operation and disposal, and the energy has been generated in its operation time which is selected as 20 years. Iron, cast iron, steel, copper, aluminum, and oil have been sent to a recycling facility and the composite material has been sent for incineration at the end of its life. The energy payback period has been calculated as 10 months in both scenarios and the embodied energy is 2858.2 MWh and 2830.3 MWh for a wind turbine using Al-Conductor cable and a wind turbine using Cu-Conductor cable during its lifetime, respectively. Air emission and wastewater generation have been calculated to assess environmental impacts (global warming, acidification, eutrophication, etc.) whose consequence provides an evaluation of the life cycle assessment of a wind turbine. For example, global warming is 14.44 g eq./kWh and 15.24 g CO₂ eq./kWh for a wind turbine using Al-Conductor cable and a wind turbine using Cu-Conductor cable, respectively. As a result, the life cycle analysis of the wind turbines has been evaluated and compared regarding two different scenarios according to the definition of ISO 14040 standard throughout its life from production to disposal.

Index Terms—Al-Conductor Cable, Cu-Conductor Cable, Life Cycle Assessment, Wind Turbine


I. INTRODUCTION

ONE OF the branches of renewable energy is wind energy has worked the kinetic energy of wind turns into electrical energy by wind turbines. The wind has been used for

BUKET KÜÇÜKKARACA, Istanbul, Turkey, (e-mail: buketkkaraca@gmail.com).

 <https://orcid.org/0000-0002-7480-2060>

BURAK BARUTÇU, is with Energy Institute of Istanbul Technical University, Istanbul, Turkey, (e-mail: barutcb@itu.edu.tr).

 <https://orcid.org/0000-0002-8834-2317>

*Corresponding Author

Manuscript received December 03, 2021; accepted June 27, 2022.

DOI: [10.17694/bajece.1032172](https://doi.org/10.17694/bajece.1032172)

various purposes such as milling, pumping water, sailing since ancient times. [1] Due to an increase in CO₂ emission and water consumption and the decrease in fossil fuels, the wind has been one of the energy sources with high growth recently [2]. Wind, solar, biomass, ocean, geothermal, hydropower energies are among renewable energies and are therefore called clean energy and green [3].

The purpose of the paper is to examine the life cycle assessment of a wind turbine according to its energy indexes and environmental impacts. In this context, energy consumption, energy payback time, energy intensity, and CO₂ intensity have been assessed in detail. During its life cycle, the wind turbine has five stages including production, transportation, construction, operation, and disposal. Two types of cables have been used in the electrical infrastructure of wind farms; thus, the results have been evaluated in two scenarios for wind turbines using Al-Conductor and Cu-Conductor cables which aim connection of turbines with each other. On the other hand, Cu-Conductor cables are used inside of the turbines.

A. Life Cycle Assessment

The issue of better understanding and reducing the environmental impacts caused by the production and consumption of products have gained importance in recent years. Therefore, one of the enhanced techniques has been life cycle assessment which aimed that assess environmental impacts and improve the environmental performance of a product throughout its lifetime. In this study, it has been maintained the approaches for life cycle assessment that have been presented in ISO 14040 series standards which have four phases which are goal and scope definition, life cycle inventory analysis, life cycle impact assessment and life cycle interpretation [4]. And this association according to phases has been shown in Fig. 1.

B. Environmental Impacts

Environmental impacts have been evaluated as a result of the life cycle assessment, such as global warming, acidification, eutrophication, human toxicity, freshwater aquatic eco-toxicity and photochemical oxidation. Environmental impacts have been determined based on characterization factor and amounts of chemicals. Each chemical has different characterization factor according to environmental impact. Global warming, which is caused by CO₂, NO₂, CH₄ and CFCs, has a global effect. Global warming gives rise to poles to melt, seasons to change, lose to soil moisture and forests to disappear.

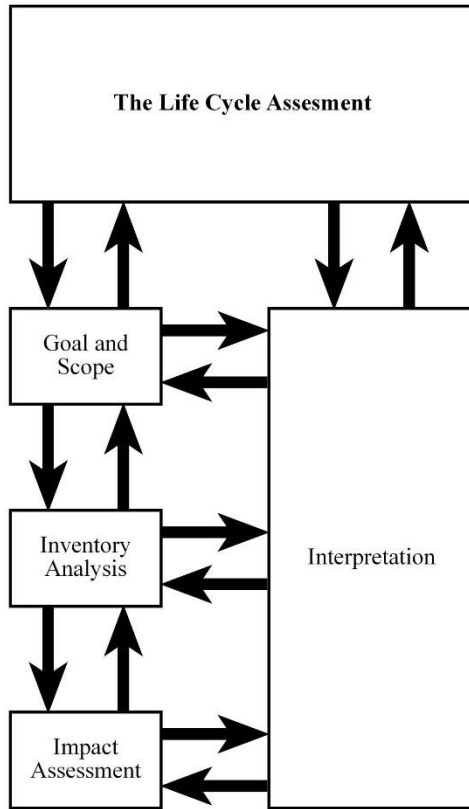


Fig. 1. The Life Cycle Assessment [4]

Acidification has both regional and local effects, commonly caused by SO_x , NO_x , HCL , HF , and NH_4 . Acidification leads to corrosion of building, the body of water acidification, vegetation, and soil effects. Eutrophication, is caused by nutrients which are PO_4 , NO , NO_2 , nitrates, and NH_4 , has local effects. Human toxicity stems from chemicals released into the soil, water, and air, increasing the effects of morbidity and mortality at global, regional, and local scales. Aquatic ecotoxicity stems from toxic chemicals at the local scale, which induces decreasing biodiversity in the waterbody. Photochemical oxidation is caused by non-methane hydrocarbon, which leads to smog and vegetation damage [5].

C. Wind Turbine

Wind turbines may be classified on the basis of their rotational axis into horizontal and vertical wind turbines. Nowadays, horizontal axis wind turbines are used more for commercial purposes due to higher efficiency and high power density. On the other hand, the vertical axis wind turbines may be divided into two major types which are the Savonius and the Darrieus. In vertical axis wind turbines design, generator and gearbox are located on the ground, which means that they do not need any tower causing low efficiency due to low wind speed. Vertical axis wind turbines do not need yaw systems since they do not follow the wind direction [6], [7].

In this paper, energy and environmental examinations have been conducted for horizontal axis wind turbines, including life cycle analysis has been evaluated. Thus, components of the horizontal axis wind turbine were examined to analyze the

life cycle assessment. The components of the wind turbine are basically divided into four parts which are foundation, tower, rotor, and nacelle [8].

1) Foundation

The most important role of foundation in wind turbines is to provide stability by fixing the wind turbines to the ground. The broadest foundation classifications are slabs foundations and pile foundations [9], [10].

2) Tower

The tower, which is generally made of steel with a tubular form, enables that carry rotor and nacelle. Large static loads due to changing wind power are captivated by the tower. The tower comprises one of the significant components in terms of the cost of the wind turbine [9], [10].

3) Rotor

The rotor can mainly be assessed under two headings: blades and hub. The geometry of blades has been designed by examining aerodynamics; in this context, it resembles the airplane wings which means its cross-section is asymmetric. In order to minimize the loads and fatigue, lightweight materials are chosen, in these circumstances, the blades generally are made of synthetics reinforced with carbon fibers and fiberglass, and epoxy resin. Nowadays, three-bladed horizontal axis wind turbines, which provide the highest efficiency, are preferred. The hub provides that the blades are fixed to the rotor shaft, and it is generally made of cast iron due to its complex structure. The enormously significant function of the hub is that the energy coming from blades transmitted to the generator [9], [10].

4) Nacelle

The nacelle comprises wind turbine machinery which are the main shaft, main bearing, main gear, couplings, mechanical brake, hydraulics systems, generator, machine support frame, nacelle enclosure, and yaw system. Rotational energy coming from the hub is directed to the gearbox or the generator by the main shaft which also hands over the loads to the fixed system nacelle. Main gear, which transports energy among the rotor and the generator, provides that increase speed. Brakes in wind turbines can be categorized into two groups: aerodynamic brake systems and mechanical systems. The aim of the generator is that the mechanical energy is converted to electrical energy. The yaw system provides that the rotor adjusts the right position according to the wind direction [9], [10].

D. Wind Energy in Turkey

In 2019, 8 percent of electricity produced in Turkey provided by wind energy. The estimated wind power installation potential is 48 GW in Turkey, where has suitable areas in the Aegean and Mediterranean region [11]. 1.30% of the area of Turkey has high wind energy potential [12]. The data obtained from the Turkish Wind Energy Association about the cumulative installations of wind energy in Turkey have been shown in Fig. 2.

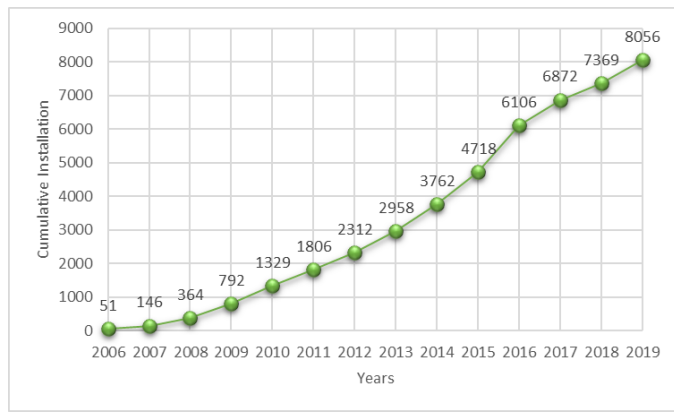


Fig. 2. Cumulative Installation of wind energy in Turkey [13]

II. STAGES

The ISO 14040 series standards have defined Life Cycle Assessment (LCA), giving the requirements for its execution. The definition of the ISO 14040 series standards has been used in the determination of environmental impacts of the wind turbine during its entire lifetime. The wind turbine Gamesa G8X model with 2 MW has been investigated in this study in order to calculate the energy consumption and environmental impacts of the wind turbine and its components [8]. Gamesa G8X-2.0 MW wind turbine is a three-bladed horizontal design, active yaw system, upwind and pitch regulated [14]. To start with, necessary materials for the production of the wind turbine have been sorted according to their weights and then their energy requirement has been found to calculate the energy requirement of production of the wind turbine and its components. These materials of the wind turbine are transported and constructed afterward. The operation time of the wind turbine is estimated as 20 years [8]. After complete its life, components of the wind turbine have been sent to proper disposal facilities. As a result, the life cycle assessment of the wind turbine is conducted in five different stages which are production, transportation, construction, operation, and disposal. Fig. 3 represents the stages of a wind turbine life cycle.

TABLE I. WEIGHT OF WIND TURBINE MATERIALS [8]

Material	Value	Unit
Resin	13	t
Fibre Glass	9	t
Cast Iron	14	t
Iron	44	t
Steel	180	t
Silica	0.3	t
Copper	4	t
Concrete	270	m ³

A. Production of Wind Turbine

The wind turbine consists of foundation, tower, rotor, and nacelle. Concrete, steel and iron have been used for foundation which comprises of ferrule and footing. The tower is made of steel and 67 m tall. The nacelle is made of iron, steel, silica, copper, fiber glass and resin. Nacelle includes a bed frame,

main shaft, transformer, generator, gearbox, and nacelle cover. Rotor is made of resin, fiber glass and cast iron. Rotor includes three blades, blade hub and nose-cone. The length of a blades is 39 m [8]. The necessary materials and their weights have been given in Table I.

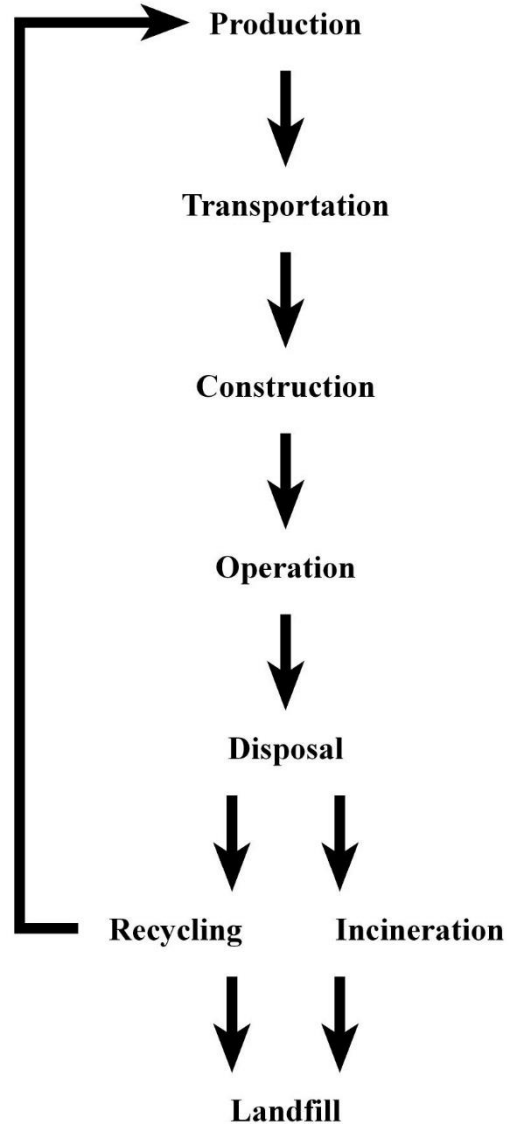


Fig. 3. Summary Life Cycle of a Wind Turbine

Furthermore, cabling has been one of the most significant components of the production of wind turbine. Converting wind energy into electrical energy is made by wind turbines which need an electrical supply system used medium voltage generally. All turbines are connected to the main transformer by cables which transmit the energy. Cables play an important role in terms of their cross-sections and lengths that varies with distance and wind farm structure [15]. XLPE cable has used both internal and connection cabling. XLPE cable in the wind turbine may be divided into two types of cables that Al-

Conductor and Cu-Conductor cable. Al-Conductor cable is made of aluminum and copper; however, Cu-Conductor cable is only made of copper, and both have a triple formation [16]. 50 mm² cross section Cu-Conductor cables have been used in internal cabling. In contrast to internal cabling, both Al-Conductor cable and Cu-Conductor cables have been used in connection cabling. It is assumed that 95 mm², 120 mm² and 185 mm² cross-section Al-Conductor cables have been used in the connection. In the past, Cu-Conductor cables were used in wind farms. Al-Conductor cables have been used in recent years in Turkey instead of Cu-Conductor cables. Al-Conductor cables have high unit length resistance, notwithstanding it costs less compared with Cu-Conductor cables. Despite the high joule losses, Al-Conductor cables are selected in Turkey because of the cost. The optimization of the position of the cables has not been implemented in Turkey. After the location of wind turbines is determined, the road is determined for construction equipment according to the slope and wind turbine location. Therefore, cables have laid parallel to the road. However, the length of the cables can be shortened owing to optimization. The location of the wind farm has not been determined specifically in this study. Because of this reason, the calculation has been made according to 1 km of cable length.

B. Transportation of Wind Turbine

Transportation is one of the most vital stakeholders in life cycle assessment of wind turbines. A specific wind farm has been not specified in this study. However, average distances have been determined from Europe to Turkey via trucks. It is assumed that the tower and foundation have been manufactured in Turkey. The distances from tower and foundation manufacturers to the wind farm are 500 km and 150 km via trucks, respectively. The rotor and nacelle have been transported from Europe to the wind farm, whose distances have been measured as 3500 km and 4000 km, respectively [17].

C. Construction of Wind Turbine

Construction equipment, which is necessary for both construction and deconstruction, cause energy consumption and emissions. In this context, crane, forklift and excavation digger have been used both construction and deconstruction. Working hours that seemed appropriate have been decided for construction equipment. The crane has worked for 2 days, the forklift and excavation digger have worked for 1 day and the same working hours have been accepted meanwhile in deconstruction. The fuel consumption of hourly crane, forklift, and excavation digger is 620.1 L, 64 L, and 44.1 L, respectively [18].

D. Operation of Wind Turbine

It is predicted that the wind turbine has been 20 years in operation. The wind turbine should be maintained during its operation period. In this context, the 300.8 ton lubricant in the gearbox and the cooling systems of the wind turbine should be changed [19]. The generator, which is made of steel, copper, and silica, should be changed in the operation stage [8].

E. Disposal of Wind Turbine

After the wind turbine has completed its life, the components will be sent to appropriate disposal facilities. Iron, cast iron, steel, copper, aluminum and lubricant have sent to recycling facilities. The issue of composite materials has been a controversial subject within the field of disposal methods. In contrast to other materials, composite materials have been sent to the incineration facility in this study [20], [21]. Moreover, it has been taken into account the transportation calculation from the wind farm to disposal facilities.

III. RESULTS

The result part has defined and demonstrated key parameters, which are energy payback time, energy intensity, CO₂ intensity, and environmental impacts. There are two different scenarios currently being adapted in research into life cycle assessment. One is the wind turbine using Al-Conductor cable for connection and the other one is the wind turbine using Cu-Conductor. The reason why it is made two scenarios is two different types of cables used in wind farms. Therefore, it has been calculated in two different scenarios, mentioned and compared their results in detail.

A. Capacity Factor, Lifetime and Electricity Generation

Capacity factor varying between 10 and 50% has presented the actual energy production of the wind turbine. However, the capacity factor relevant to modern wind turbines is ranged from 20 to 35%. Reasons for changing the capacity factor can be location, size of the wind turbine, wind reliability, and wind availability. On the other hand, lifetime of a wind turbine can be varied between 20 and 30 years. For many years, the tower and foundation can last, but the generator, gearbox and blades may constitute a problem, so they should be changed relevantly [22].

It is assumed that the 2 MW wind turbine has been used for a duration of 20 years, having a capacity factor of 20%. The assumptions of the present study demonstrate that the annual electricity generation of the wind turbine is 3504 MWh.

B. Embodied Energy

One of the energy indexes that may be significant to take into consideration is embodied energy, which may variate according to materials in the production stage, the weight or type of materials, and the choice of the wind turbine, as well as affect the life cycle analysis from wind turbine manufacturing to its disposal stage [22].

Embodied energy has been calculated for five different stages in both scenarios. The total embodied energy of both scenarios has been shown in Table II.

C. Energy Payback Time, Energy Intensity and CO₂ Intensity

The term "Energy payback time" refers to a quantitative parameter that recovers initial investment [23]. Eq. (1) below illustrates the calculation of energy payback time [18]:

$$\text{Energy Payback Time} = \frac{\text{Embodied Energy}}{\text{Generated Energy}} \quad (1)$$

According to this formula, energy payback time has been calculated for each scenario, and the results have been given in Table II.

Two other indexes are energy intensity and CO₂ intensity, which have a significant role in the life cycle assessment of

the wind turbine [24]. The results obtained from both scenarios have been compared in Table II.

TABLE II
ENERGY INDEXES

Wind Turbine	Embodied Energy (MWh)	Energy Payback Time (months)	Energy Intensity (MWh/MWh)	CO ₂ Intensity (kg CO ₂ /kWh)
Wind Turbine using Al-Conductor Cable	2858.2	10	0.0408	0.016
Wind Turbine using Cu-Conductor Cable	2830.3	10	0.0404	0.017

A comparison of the two embodied energy results reveals that there is no strict difference between Al-Conductor and Cu-Conductor wind turbines. The reason why the high result in the embodied energy of the wind turbine using Al-Conductor cable is energy consumption in both production and disposal stages. Comparing the two results of energy payback time, it can be seen that these two results are the same

in both scenarios. No important difference has been found between both wind turbines in terms of energy and CO₂ intensity. These parameters have played such a significant role in the life cycle assessment of wind turbines that they have compared them with other results obtained from the literature, and it has been shown in Table III.

TABLE III
LITERATURE COMPARISON

Wind Turbine	Lifetime (years)	Onshore/Offshore	Power (MW)	Energy Payback Time (years)	Energy Intensity (MWh/MWh)	CO ₂ Intensity (gCO ₂ /kWh)
Wind Turbine using Al-Conductor	20	Onshore	2	0.82	0.041	15.92
Wind Turbine using Cu-Conductor	20	Onshore	2	0.81	0.040	16.71
France [24]	20	Onshore	3	1.03	0.051	11.77
France [24]	40	Onshore	3	0.81	0.04	8.87
Italy [25]	20	Onshore	0.66	-	-	14.8*
Vestas V110 [26]	20	Onshore	2	0.67	-	7.2*
Vestas V112 [27]	20	Onshore	3	0.67	-	7*
Siemens Wind Power [28]	-	Onshore	2.3	0.52	-	6*
Siemens Wind Power [28]	-	Onshore	3.2	0.43	-	5*
Siemens Wind Power [28]	-	Offshore	4	0.93	-	10.9*
Siemens Wind Power [28]	-	Offshore	6	0.83	-	7.8*

*given in terms of gCO₂-eq/kWh

It can be seen from the data in Table III that the lifetime of the wind turbine generally evaluates as 20 years. It is apparent from this table that all wind turbines except one have less than a year of energy payback time. It has been reported similar results for energy intensity. However, it has been listed various results for CO₂ intensity due to different acceptance.

D. Air Emission and Wastewater Generation

Air emissions and wastewater generation have been calculated in order to conclude a better understanding of the life cycle assessment of the wind turbine and predict the environmental impacts of the wind turbine. It is assumed that wastewater has been generated during the production of the

materials. Table III illustrates some of the main parameters of the wastewater generated by wind turbine production. Unlike wastewater generation, all stages have resulted in air emissions. Air pollutants may be classified two categories which are primary air pollutants and secondary air pollutants. Primary air pollutants have been released directly to the atmosphere from a source, that pollutants are NO_x, CO, SO_x, PM and Hydrocarbons. Primary pollutants cause diseases that have a great risk for human health such as cardiovascular disease, cancer, chronic obstructive pulmonary disease and

asthma [29]. Apart from hydrocarbons, all primary air pollutants have been listed in Table IV.

TABLE IV
WASTEWATER GENERATION

Pollutant	Wind Turbine using Al-Conductor	Wind Turbine using Cu-Conductor	Unit
Chemical Oxygen Demand (COD)	0.047	0.047	g/kWh
Suspended Solid	0.005	0.005	g/kWh
Total Organic Carbon (TOC)	0.004	0.004	g/kWh
Biochemical Oxygen Demand (BOD ₅)	0.002	0.002	g/kWh

TABLE V
AIR EMISSION

Pollutant	Wind Turbine using Al-Conductor	Wind Turbine using Cu-Conductor	Unit
SO ₂	0.01	0.01	gSO ₂ /kWh
NO _x	0.04	0.04	gNO _x /kWh
CO	0.02	0.01	gCO/kWh
PM _{2.5}	0.002	0.002	gPM _{2.5} /kWh

E. Life Cycle Assessment

Wastewater and air emissions give rise to environmental impacts, and the resulting substances have been calculated by characterization factors. Environmental impacts are listed in Table V as a result of the entire life cycle of the wind turbine.

TABLE VI
LIFE CYCLE ASSESSMENT

Environmental Impact	Wind Turbine using Al-Conductor	Wind Turbine using Cu-Conductor	Unit
Global Warming (GWP100)	14.44	15.24	g CO ₂ eq./kWh
Eutrophication (EP)	5.39	5.39	mg PO ₄ - eq./kWh
Acidification (AP)	14.43	14.43	mg SO ₂ eq./kWh
Human Toxicity (HT)	699.86	700.02	mg 1,4-dichlorobenzene eq./kWh
Freshwater Aquatic Eco toxicity (FWAE)	86.30	85.03	mg 1,4-dichlorobenzene eq./kWh
Photochemical Oxidation (POCP)	1.17	1.02	mg ethylene eq./kWh

It is seen obviously that both scenarios are quite similar in terms of environmental impacts. However, it is necessary to express which scenario causes more environmental impacts. Wind turbines using Al-Conductor cable have a greater effect on global warming, eutrophication, human toxicity; although, wind turbines using Cu-Conductor cable have a higher effect on photochemical oxidation and freshwater aquatic ecotoxicity. And both scenarios have the same acidification effect on the environment. As shown in Fig. 4 and 5 that transportation, construction, and operation stages have the same rate in both cases, and production and disposal stages present a difference in these two situations because of cables.

F. Disposal and Total Energy Consumption

On the other hand, the share of disposal in the energy consumption has a great importance. 27.0% of the total energy consumption has been spent on a wind turbine using Cu-Conductor cable and 26.6% of the total energy consumption has been spent on a wind turbine using Al-Conductor cable.

This situation indicates that disposal is a significant indicator in energy consumption due to its higher than one-quarter portion.

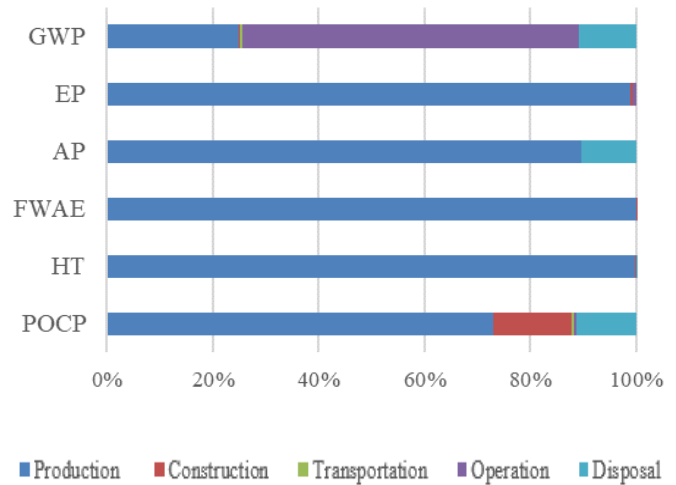


Fig. 4. Wind Turbine with Al-Conductor Cable

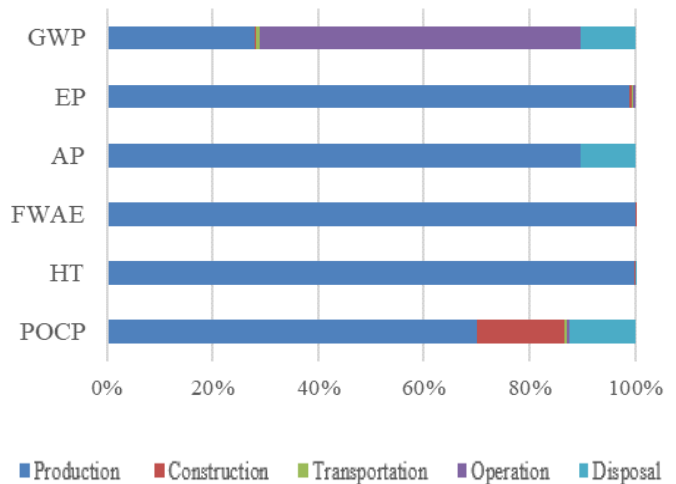


Fig. 5. Wind Turbine with Cu-Conductor Cable

IV. DISCUSSIONS

This paper has set out to determine the life cycle assessment of a wind turbine in Turkey, consisting of a comparison between wind turbine using Al-Conductor and wind turbine using Cu-Conductor. This paper has shown the energy indexes and environmental impacts of a wind turbine, and the results have been given based on comparative analysis. Considering the comparison of the energy payback times of the scenarios, the wind turbine using the Al-Conductor cable is longer than the wind turbine using the Cu-Conductor cable; as a result, the wind turbine using the Cu-Conductor cable provides shortening the energy payback time. Al-Conductor cable is more widely used nowadays because of its cost. Although a wind turbine using an Al-Conductor cable is higher in embodied energy, its impact on global warming is less. The

data given in this article for Al-Conductor cable is of great importance for many countries. The calculations could change according to the geography of the wind farm, the location of wind turbines and cabling and the design of cabling; therefore, it cannot be generalized.

V. CONCLUSION

Further research in the disposal method of the blades would be useful to clarify the subject. In this study, it was determined that the blades are sent to the incineration facility since their recycling efficiency is low. On the other hand, the foundation has been left in the wind farm after the wind turbine completes its lifetime, and this situation should be evaluated in further research. In this study, cables were sent to a recycling facility, but cable sales are another option for this situation. Further research has needed to be done to establish whether sales of cable or recycling of cable is more efficient.

REFERENCES

- [1] Elhoury, S. A. (2018). Using Wind Power Plants as Alternative Energy. *Global Journal of Engineering Science and Researches*, 5 (2).
- [2] Saidur, R., Rahim, N. A., Islam, M. R., and Solangi, K. H. (2011). Environmental impact of wind energy. *Renewable and Sustainable Energy Reviews*, 15 (2011), 2423–2430. doi:10.1016/j.rser.2011.02.024.
- [3] NREL. (2001). Renewable Energy: An Overview. doi:10.1049/ic.2008.0789.
- [4] Lee, K.-M., and Inaba, A. (2004). Life Cycle Assessment: Best Practices of International Organization for Standardization (ISO) 14040 Series. *Committee on Trade and Investment*, (February), 99. Retrieved from http://publications.apec.org/publication-detail.php?pub_id=453
- [5] Agency, U. S. E. P. (2006). Life Cycle Assessment: Principles and Practice. *Global Shadows: Africa in the Neoliberal World Order*, 44 (2), 8–10.
- [6] Tong, W. (2010). Fundamentals of wind energy. *WIT Transactions on State of the Art in Science and Engineering*, 44. doi:10.2495/978-1-84564.
- [7] Makhlas, K. Al, and Alsehli, F. (2014). Wind Power. Retrieved from <http://linkinghub.elsevier.com/retrieve/pii/B978012014901850005X>
- [8] Martinez, E., Sanz, F., Pellegrini, S., Jimenez, E., and Blanco, J. (2008). Life cycle assessment of a multi-megawatt wind turbine. *Renewable Energy*, 34 (2009), 667–673. doi:10.1016/j.renene.2008.05.020.
- [9] DNV/Risø. (2013). *Design of Wind Turbines*. doi:10.1201/b15566-7.
- [10] BWE. (2016). The Structure of a Modern Wind Turbine – An Overview. *German Wind Energy Association*, 1–14. Retrieved from http://www.wwindea.org/technology/ch01/en/1_2.html
- [11] The Trade Council of Denmark in Istanbul. (2020). WIND ENERGY MARKET Prepared by The Trade Council of Denmark in Istanbul wind ENERGY IN DENMARK.
- [12] Ulu, E. Y., and Dombayci, O. A. (2018). Wind Energy in Turkey: Potential and Development. *Technology, Engineering & Mathematics (EPSTEM)*, 4, 132–136. Retrieved from www.isres.org
- [13] Turkish Wind Energy Association. (2020). Turkish Wind Energy Statistics Report January 2020.
- [14] Gamesa Eólica. (2007). *Drawings and Specifications of Gamesa Eolica Wind Turbines*.
- [15] Deutsches Windenergie-Institut, Tech-wise, and DM Energy. (2001). Wind Turbine Grid Connection and Interaction.
- [16] Nexans. (n.d.). Nominal cross section, 30 (36).
- [17] SGRE. (2019). Location Finder I Siemens Gamesa.
- [18] Chipindula, J., Sai, V., and Botlaguduru, V. (2018). Life Cycle Environmental Impact of Onshore and Offshore Wind Farms in Texas, 1–18. doi:10.3390/su10062022.
- [19] Haapala, K. R., and Prempreeda, P. (2014). Comparative life cycle assessment of 2.0 MW wind turbines, 3 (2), 170–185.
- [20] Jensen, J. P. (2018). Evaluating the environmental impacts of recycling wind turbines. *Wind Energy*, 22 (2019), 316–326. doi:10.1002/we.

- [21] BIR. (2008). Report on the Environmental Benefits of Recycling. *October*, (April), 49. Retrieved from <http://www.thenbs.com/topics/DesignSpecification/articles/benefitsMasterSpecifications.asp>
- [22] Crawford, R. H. (2009). Life cycle energy and greenhouse emissions analysis of wind turbines and the effect of size on energy yield. *Renewable and Sustainable Energy Reviews*, 13 (2009), 2653–2660. doi:10.1016/j.rser.2009.07.008.
- [23] Ardente, F., Beccali, G., and Æ, M. C. (2004). Life cycle assessment of a solar thermal collector: sensitivity analysis, energy and environmental balances. *Renewable Energy*, 30 (2005), 109–130. doi:10.1016/j.renene.2004.05.006.
- [24] Palomo, B., and Gaillardon, B. (n.d.). Life Cycle Assessment of a French Wind Plant.
- [25] Ardente, F., Beccali, M., Æ, M. C., and Brano, V. Lo. (2006). Energy performances and life cycle assessment of an Italian wind farm. *Renewable and Sustainable Energy Reviews*, 12 (2008), 200–217. doi:10.1016/j.rser.2006.05.013.
- [26] Razdan, P., and Garrett, P. (2015). Life Cycle Assessment of Electricity Production from an onshore V110-2.0 MW Wind Plant, (December).
- [27] D'Souza, N., Gbegbaje-Das, E., and Shonfield, P. (2011). Life Cycle Assessment of Electricity Production from a V112 Turbine Wind Plant. doi:10.1201/noe0415375528.ch3.
- [28] Bonou, A., Laurent, A., and Olsen, S. I. (2016). Life cycle assessment of onshore and offshore wind energy: from theory to application. *Applied Energy*, 180, 327–337. doi:10.1016/j.apenergy.2016.07.058.
- [29] Sule, T. U. N. (2014). The Influence of Primary Air Pollutants on Human Health Related Risk. *Journal of Environment and Earth Science*, 3 (8).

BIOGRAPHIES



the Energy Institute within the same university in 2020.

BUKET KÜÇÜKKARACA ŞİŞLİ, İstanbul, in 1995. She graduated from the Environmental Engineering Department of the Civil Engineering Faculty of Istanbul Technical University in 2018. She also recently graduated with a Master's degree in Energy Science and Technologies from the Energy Institute within the same university in 2020.



BURAK BARUTÇU Uskudar, İstanbul, in 1971. He received the B.S. degree in Electrical Engineering (1994), M.S. (1997) and PhD. (2005) Degrees in Nuclear Engineering from the Istanbul Technical University. From 1997 to 2003 he was a Research Assistant in Nuclear Energy Institute and between 2003 – 2007 Energy Institute of ITU. he has been an Assistant Professor in Renewable Energy Department of Energy Institute since 2007. His research interests include wind energy systems, photovoltaic systems, wind and solar energy prediction, signal processing.

Dynamic Economic Dispatch with Valve-Point Effect Using Crow Search Algorithm

Cenk Andic, Ali Ozturk and Salih Tosun

Abstract—This paper presents a method based on meta-heuristic to solve Dynamic Economic Dispatch (DED) problem in a power system. In this paper, Crow Search Algorithm (CSA), which is one of the heuristic methods is proposed to solve the DED problem in a power system. In this study, line losses, generation limit values of generators, generation-consumption balance, valve-point effect and ramp rate limits of generator are included as constraints. The CSA is implemented on two different test cases. Finally, the CSA results are compared with the results of well-known heuristics in the literature such as Particle Swarm Optimization (PSO), Genetic Algorithm (GA), Symbiotic Organism Search (SOS) algorithm, Artificial Bee Colony (ABC) algorithm, Simulated Annealing (SA), Imperial Competitive Algorithm (ICA), Modified Ant Colony Optimization (MACO) algorithm. The results show that the proposed algorithm has a better operating cost. With the results of the algorithm proposed in the test system 1, a profit of \$2,056,5931 per day and \$751,751,4815 per year is obtained. It is seen that with the results of the algorithm proposed in the test system 2, a daily profit of \$12,279,7328 and a yearly profit of \$4,482,102,472 are obtained. Test systems are operated by using less fuel with the results of the proposed algorithm and thus the harmful gas emissions released by thermal production units to the environment are also reduced.


Index Terms—Crow search algorithm, dynamic economic dispatch, thermal power units, valve-point effect

I. INTRODUCTION


THE DEMAND for electrical energy is increasing day by day. On the contrary, fossil energy resources are decreasing. Thus, it is an important issue to operate the electrical energy with the least cost [1-3].

Thermal power plants that provide electrical energy


CENK ANDIC, is with Department of Electrical Engineering University of Istanbul Technical University, Istanbul, Turkey, (e-mail: andic18@itu.edu.tr).

 <https://orcid.org/0000-0003-1123-899X>

ALI OZTURK, is with Department of Electrical-Electronics Engineering University of Duzce, Duzce, Turkey, (e-mail: aliozturk@duzce.edu.tr).

 <https://orcid.org/0000-0002-3609-3603>

SALIH TOSUN, is with Department of Electrical-Electronics Engineering University of Duzce, Duzce, Turkey, (e-mail: salih Tosun@duzce.edu.tr).

 <https://orcid.org/0000-0002-5698-6628>

Manuscript received February 18, 2022; accepted May 08, 2022.

DOI: [10.17694/bajece.1075860](https://doi.org/10.17694/bajece.1075860)

production have production costs so the generators are spending fuel. These fuel they spend are modeled with a quadratic equation called fuel cost function. The purpose of the Economic Dispatch (ED) problem is to distribute the power required by the loads on the power systems taking into account the fuel costs of the generators. Thus, more electric power generation is expected from the generator, which has a low fuel cost. However, when performing these loads to the generators, the lower and upper bound of parameters will be taken to produce. The ED problem can be solved by traditional methods [2-19]. A more realistic analysis is made by including the valve-point effect of the generators. In this case, the fuel cost function transforms from a quadratic equation to a non-convex equation [3-6]. In this case, it becomes difficult to find the global best solution to the problem. Therefore, heuristic methods are used to solve this problem, which is a non-convex structure [7-31]. [13] used the method of Artificial Neural Networks (ANN) to solve the ED problem.

In the literature, when we examine the dissolution of this problem by heuristic methods, we see that the ED problem is solved by Genetic Algorithm (GA) in the study of [14]. [15] Solve the ED problem with Particle Swarm Optimization (PSO) algorithm. [16] solves with Artificial Bee Colony (ABC) algorithm. [17] solves the ED problem with Crow Search Algorithm (CSA). [18] has solved the ED problem using Grey Wolf Optimization (GWO) algorithm.

The ED problem aims to distribute the load value given for a only one hour to the generators. However, in operating conditions, the load value changes continuously [11-33]. In this case, the Dynamic Economic Dispatch (DED) problem is solved which takes into account the hourly load change. The load sharing is provided to the generators according to the 24-hour load change [35]. However, the part to be considered here should be paid to the production increase or decrease limits (ramp limit) of generators for hourly load changes [19]. [20] has solved the DED problem using the Symbiotic Organisms Search (SOS) algorithm. [21] has solved the DED problem using Harmony Search Algorithm (HSA). [22] has solved the DED problem using Simulated Annealing (SA) technique.

In this study, the CSA, which is an heuristic methods is proposed to solve the DED problem. The CSA was run on two different test systems. The results obtained were compared with other results in the literature such as the SA [22], the GA [24], the PSO [25], Imperial Competitive Algorithm (ICA) [26], Sequential Quadratic Programming (SQP) [27], the PSO

[28], Evolutionary Programming–Sequential Quadratic Programming (EP-SQP) [29], Particle Swarm Optimization–Bacterial Foraging (PSO-BF) [30], Chaotic Sequence–Differential Evolution (CS-DE) algorithm [31], Artificial Immune System (AIS) [24], the ABC [24], CDBCO [32], the SOS [20], Modified Ant Colony Optimization (MACO) [33], combining Biogeography–Based Optimization with Brain Storm Optimization (BBOSB) [34] and it was seen that the CSA results found best operating conditions in terms of fuel cost.

II. PROBLEM FORMULATION OF DED PROBLEM

A large part of electrical energy production is met by thermal generation units. In order for thermal power plants to produce electrical energy, they need to burn fuel. The power demanded by the loads given for any given hour is shared between the thermal power plants in the power system with the ED problem. It is based on the idea that the thermal generation unit with less fuel cost should produce more electrical energy. In addition, The DED problem allocates the thermal power plants according to the 24-hour (one-day) changing load values. The fuel cost function of m thermal generation units is shown in Eq. (1) below.

$$\min \sum F_i(P_{Gi}) = \sum_{i=1}^m (a_i + b_i \cdot P_{Gi} + c_i \cdot P_{Gi}^2) \quad (1)$$

where, P_{Gi} : it is the active power value produced by the i th thermal power plant. a_i , b_i and c_i are the fuel cost function coefficients of i th thermal power plant. The fuel cost function of the thermal generation units is expressed as a second order function as shown in Eq. (1). However, in reality, valves are used to adjust the output power of steam turbines. If there is an increase in the output power, the valves are activated in order to increase the input power of the plant. However, while the control mechanism trying to achieve this balance tries to maximize efficiency, a ripple occurs in the output power and this effect is called the valve-point effect [16-17]. In this study, valve-point effects of steam turbines are included in the fuel cost function. With the valve-point effect, the DED problem turns into a non-convex problem and becomes a complex problem without a global solution point. The fuel cost function of a thermal power plant with the valve-point effect included is shown in Eq. (2) below.

$$\min \sum F_i(P_{Gi}) = \sum_{i=1}^m (a_i + b_i \cdot P_{Gi} + c_i \cdot P_{Gi}^2) + |d_i \cdot \sin(e_i \cdot (P_{Gi}^{min} - P_{Gi}))| \quad (2)$$

where, d_i and e_i are the fuel cost function co-efficients showing the valve-point effect. P_{Gi}^{min} is the minimum limit value that the i th thermal power plant can produce. In this study, the DED problem with valve-point effects is subjected to constraints given as follows:

A. Demand Constraint

$$\sum_{i=1}^N P_i = P_D + P_L \quad (3)$$

The sum of the power demanded by the loads (P_D) and the lost power (P_L) must be equal to the sum of the power produced by N thermal power plants (P_i). There must be a balance of power. Power loss can be calculated using B-coefficients, given as follows:

$$P_L = \sum_{i=1}^N \sum_{j=1}^N P_i B_{ij} P_j \quad (4)$$

B. Real Power Operating Limits

The active power produced by thermal power plants has a certain minimum–maximum limit range. Thermal power plants should be operated within this limit range.

$$P_{Gi}^{min} \leq P_{Gi} \leq P_{Gi}^{max} \quad (5)$$

C. Generating Unit Ramp Rate Limits

$$P_{Gi,t} - P_{Gi,t-1} \leq UR_i \quad (6)$$

$$P_{Gi,t-1} - P_{Gi,t} \leq DR_i \quad (7)$$

where $P_{Gi,t}$ and $P_{Gi,t-1}$ are power outputs of the i th generating units at time t and $(t-1)$, UR_i and DR_i are up and down ramp-rate limits of i th generating units.

III. CROW SEARCH ALGORITHM

The CSA is one of the metaheuristic optimization algorithms introduced by Askarzadeh in 2016 [23]. The CSA is inspired by the intelligent behavior of crows and is a population-based optimization algorithm. Crows live in flocks and each crow has a home in its memory. Crows store their excess food in their nests. While one crow takes its excess food to its nest, it is followed by another crow. Two different situations can occur. i) The crow flies to its nest without noticing that it is being followed and steal it once the owner leaves. ii) The crow realizes that it is being followed and flies to a random place to protect the food in its nest. One of these two situations occurs. We can think of each crow in the flock as a solution to an optimization problem. The steps of the crow search algorithm are given below.

- (a) Initialize problem and adjustable parameters
- (b) Initialize position and memory of crows
- (c) Evaluate fitness function
- (d) Generate new position
- (e) Check the feasibility of new positions
- (f) Evaluate fitness function of new positions
- (g) Update memory
- (h) Check termination criterion

The flowchart of the crow search algorithm is shown in Fig. 1.

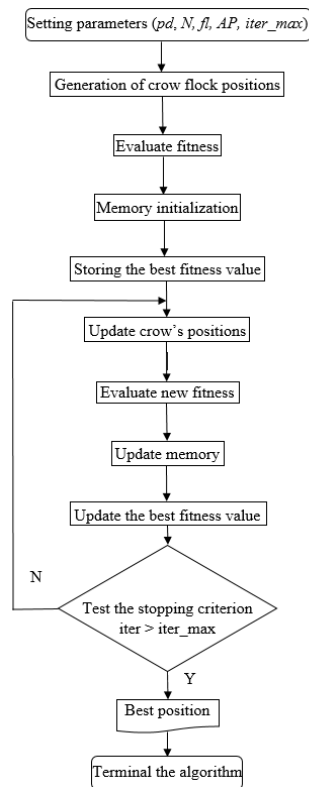


Fig. 1. Flowchart of the proposed method

IV. CSA IMPLEMENTATION FOR DED PROBLEM

This section shows the steps in implementation the CSA to the DED problem with valve-point effect.

A. Generate The Initial Population of The Flock

In the crow flock population, there are as many crows as the problem size in each flock (flock size is N). And each crow represents the solution to the DED problem. For example, since there are five variables in test case 1, the d -dimensional should be five. Initially, the positions of the crows in the flock are randomly generated to fit the lower and upper bounds of the variable in the problem.

$$x^{i,iter} \quad (i = 1, 2, \dots, N) \quad (8)$$

$$x^{i,iter} = [x_1^{i,iter}, x_2^{i,iter}, x_3^{i,iter}, \dots, x_d^{i,iter}] \quad (9)$$

B. Calculate The Fitness Value of Flock

The fitness values of each crow in the flock are calculated. Fitness function is shown in Eq. (2).

C. Memory Initialization

In the memory of each crow in the flock there is the location of their nest. Crow's positions represent the solutions to the problem. At first, the locations of the randomly generated crow flock are stored in memory as the best value in the first iteration.

$$m^{i,iter} = [m_1^{i,iter}, m_2^{i,iter}, m_3^{i,iter}, \dots, m_d^{i,iter}] \quad (10)$$

D. Generate A New Positions

As a mentioned, crows follow each other and there are two

situations. If crow i notices that crow j is following him, crow i flies to a random place in search space. If crow i does not notice that crow j is following him, crow i flies to his nest. It is the Awareness Probability (AP) that determines which of these two situations will occur. A random number is generated between 0 and 1, and if this number is greater than the AP, the condition occurs that the crow does not notice that it is being followed. The condition that the crow does not notice that it is following is realized. If the randomly generated number is less than the AP, the crow realizes that it is being followed and flies to a random place in the search space.

$$x^{i,iter+1} = \begin{cases} x^{i,iter} + r_j \times fl^{i,iter} \times (m^{i,iter} - x^{i,iter}) & r_j \geq AP^{i,iter} \\ \text{a random position} & \text{otherwise} \end{cases} \quad (11)$$

where r_j is a random number with uniform distribution between 0 and 1. In the CSA, the parameter of awareness probability is primarily responsible for intensification and diversification (AP). By lowering the awareness probability value, the CSA is more likely to focus its search on a local area where a current good answer can be located.

E. Update Position and Memory

The positions of the crows in the flock change according to the above conditions and are updated in each iteration. Thus, the fitness values of crows also change.

$$x(i,:) = x_{new}(i,:) \quad (12)$$

In each iteration, a comparison is made between the crows' new fitness value and the best fitness value in memory. If the new fitness value is less than the fitness value in the memory, the fitness value in the memory is updated with the new fitness value.

$$mem(i,:) = x_{new}(i,:) \quad (13)$$

$$fit_mem(i) = ft(i) \quad (14)$$

Else, the fitness value in the memory continues to be iterated without updating.

F. Stopping Criterion Testing

Iteration continues until it reaches the maximum number of iterations ($iter_{max}$).

$$iter = 1, 2, 3, \dots, iter_{max} \quad (15)$$

G. The Sample Problem Solution with CSA

The function given in Eq. (2) is taken as a sample problem. The number of problem dimension (variables) is assigned to 3, the flock (population) size is assigned 2, the awareness probability is assigned 0.03, the flight length is assigned 2, while the iteration number of the algorithm is assigned to 10.

At the first step of the algorithm, the crows positions is created randomly. This initial population is given in Table I.

TABLE I
THE INITIAL POPULATION OF CROWS CREATED RANDOMLY AND CALCULATED FITNESS VALUE FOR EACH FLOCK

Hour	Demand	Flock1				Flock2			
		Crow1	Crow2	Crow3	Cost	Crow1	Crow2	Crow3	Cost
1	210	161,323	112,687	159,448	2240495	153,838	134,393	115,797	1945802,8
2	200	190,93	140,522	157,873	2899900	151,367	116,055	178,423	2464525
3	190	178,170	84,947	169,862	2435740	178,140	84,182	170,096	2430118
4	185	176,716	149,925	141,714	2839928	164,319	122,589	165,849	2683737
5	190	85,281	102,629	172,330	1707372	184,784	92,273	114,779	2023760,5
6	215	165,579	140,375	170,903	2625049	157,520	76,486	113,896	1333852
7	220	188,945	113,189	122,894	2056104	195,549	96,863	134,994	2079942
8	230	127,142	132,716	164,797	1952345	111,826	133,939	172,949	1892855
9	235	127,624	100,191	137,913	1312305	160,952	85,957	146,416	1588656,5
10	240	178,407	54,306	134,387	1276128	96,930	134,120	110,868	1023892
11	245	77,258	130,910	157,964	1216344	150,175	123,819	148,483	1780530
12	250	190,439	102,133	125,414	1685595	166,093	109,386	136,184	1622263
13	245	194,946	102,271	113,852	1666344	187,306	84,504	159,335	1867377,7
14	255	164,186	93,547	144,415	1476999	165,604	129,499	157,542	1982606
15	260	175,062	123,981	172,951	2126367	167,132	50,729	177,316	1357274
16	255	156,005	144,196	111,942	1577065	182,675	140,968	131,336	2005997
17	250	175,846	101,357	111,813	1395523	152,122	78,995	129,501	1111173,4
18	245	136,916	118,067	173,495	1840623	177,520	149,674	147,016	2298544
19	240	156,562	111,207	160,417	1887705	142,872	78,609	130,964	1129338
20	235	157,574	126,917	120,716	1707607	178,321	108,320	95,977	1481451
21	230	199,419	92,452	88,407	1508064	194,891	117,766	169,552	2528667,2
22	220	125,233	144,511	137,632	1879316	193,117	122,271	90,801	1867484
23	215	178,391	79,761	132,861	1765529	181,223	91,416	132,517	1907126
24	210	181,547	77,373	130,274	1797323	157,890	103,619	131,102	1831493
Total Cost (\$/24h)					44,875,770				44,238,465

As seen in Table 1, it represents the positions of the crows in the flocks. The positions of the crows also represent the solution to the problem. The fitness value of each flock is calculated according to Eq. (2) for each hour. As a result of the 24-hour dynamic economic dispatch, the least total cost value that gives the best fitness value in the flock is thrown into memory. Thus, the best initial value is stored in memory. The value in memory actually shows the best result up to the current iteration. This is the value for the initial population 44,238,465 \$/24h as has been shown in Table 1. Crows store their food in their nests. The positions of the crows indicate the position of their nests. Crows try to reach more food by following each other. By paying attention to the flight length of the crows in the search space, they update the positions of the crows in each iteration. The values of the positions of the crows in the flock in the 9th and 10th iterations are shown in Table 2.

TABLE II
THE CHANGE OF POSITIONS OF THE CROWS ACCORDING TO ITERATIONS

9th Iteration			10th Iteration								
Flock1			Flock2			Flock1			Flock2		
Crow1	Crow2	Crow3	Crow1	Crow2	Crow3	Crow1	Crow2	Crow3	Crow1	Crow2	Crow3
87,647	93,312	92,680	52,864	117,75	71,889	80,797	88,698	87,359	83,244	79,607	78,845
54,232	38,852	63,340	67,064	45,249	154,24	92,930	50,504	82,684	65,052	64,209	60,864
65,378	93,789	96,936	89,324	82,442	137,10	105,582	60,429	86,689	81,497	70,006	73,643
109,60	72,975	148,89	65,606	92,020	106,48	141,102	118,158	67,268	65,243	68,05	125,077
111,99	75,358	78,835	54,752	107,38	122,90	108,314	53,725	89,442	53,051	105,113	120,574
94,387	89,541	63,819	71,060	146,89	58,613	93,239	51,808	53,369	54,072	65,794	103,341
62,609	78,046	146,14	57,173	85,021	66,761	120,953	49,398	74,921	98,117	42,069	75,329
54,478	125,79	77,353	57,374	39,525	110,66	68,335	75,485	61,819	91,004	90,403	68,074
113,21	47,008	59,094	152,10	47,534	60,495	59,497	53,228	126,374	54,929	62,947	124,018
72,229	58,884	154,78	94,087	83,024	78,361	51,442	93,644	135,180	100,56	58,173	72,826
98,341	59,343	108,28	57,386	105,60	101,09	73,722	131,224	58,081	92,015	63,966	97,555
118,85	76,829	75,123	143,49	93,263	76,928	79,056	125,003	45,030	86,437	50,288	116,249
106,97	80,280	79,289	100,20	41,521	89,394	68,079	134,084	55,447	147,812	41,803	51,268
60,971	149,06	68,675	54,333	116,52	109,11	84,899	37,822	149,266	70,649	49,154	154,483
103,18	92,228	70,052	110,13	82,323	92,328	64,736	60,754	129,643	136,239	42,735	102,041
60,383	40,001	167,36	78,708	53,995	126,25	51,915	142,650	65,446	132,819	70,165	52,515
84,088	76,398	91,046	95,703	41,896	106,24	127,86	67,967	53,429	104,513	85,774	54,839
137,16	107,10	60,286	104,61	44,220	49,508	134,457	105,659	49,816	67,041	74,112	67,253
53,534	91,128	121,96	80,327	78,720	93,448	72,620	106,076	80,313	61,134	73,085	103,440
88,644	44,536	110,69	50,367	91,798	46,280	101,658	44,452	96,242	63,734	92,63	96,988
55,556	89,923	115,25	64,914	54,756	126,58	109,763	47,270	87,860	63,202	42,965	133,289
132,40	74,362	64,425	139,09	69,415	55,619	81,113	61,942	108,126	70,144	59,891	105,077
75,531	111,96	89,399	113,21	43,843	122,36	93,289	61,866	95,770	95,136	41,095	138,648
97,956	53,186	53,222	94,969	121,48	63,415	65,137	88,706	58,348	67,446	43,645	127,242
Total Cost (\$/24h)	9,449,683			9,934,455				6,725,579			5,261,095

It can be seen from Table 1 and 2 that the positions of the crows change in each iteration. In each iteration, the crow's positions are updated and the fitness value of the crow flocks is calculated according to these new values. The best fitness value among the flocks is compared with the best value in the memory, and if it is a better value than the value in the memory, the memory value is updated with this new fitness value. This comparison process is performed in each iteration. Table 3 shows the variation of the memory value in each iteration.

TABLE III
THE CHANGE OF THE VALUE IN MEMORY IN EACH ITERATION

Iter 1	Iter 2	Iter 3	Iter 4	Iter 5	Iter 6	Iter 7	Iter 8	Iter 9	Iter 10
44,238,46	36,887,14	30,403,10	26,294,20	22,989,27	19,485,01	16,247,78	12,957,75	9,449,68	5,261,09

At the end of the 10th iteration, the fitness function value is seen as \$5,261,09 per day. In this way, the dynamic economic dispatch problem is solved by using the crow search algorithm.

V. SIMULATION RESULTS

In this study, the CSA was used to solve the DED problem. The DED problem includes some constraints such as line losses, ramp-rate limits, valve-point effects and power balance. Thus, the DED problem become a non-convex problem. The proposed algorithm has been tested on two different test systems which are five units with loss and ten units without loss. Simulations were done using a 2.50 GHz Windows 10 personal computer with 16 GB-RAM and MATLAB program package and the CSA was run at least 30 times for two test cases. The results obtained from the CSA were compared with the results of other previously reported methods in the literature. Test cases and results have been given subsections. The parameters of the CSA used in this study are given in Table 4.

TABLE IV
SETTING PARAMETERS OF THE CSA FOR THE DED PROBLEM

Test Cases	pd (Problem Dimension)	AP (Awareness Probability)	fl (Flight Length)	Flock_size	Iter_max
Test Case 1	5	0.3	2	30	3000
Test Case 2	10	0.3	2	40	3000

A. Test Case 1

There are five thermal generation units in the test case 1. In this case, losses are taken into account and the B co-efficient is given in Table 5.

TABLE V
TRANSMISSION LOSS CO-EFFICIENTS FOR FIVE-UNIT TEST SYSTEM [22]

$$B = \begin{bmatrix} 0.000049 & 0.000014 & 0.000015 & 0.000015 & 0.000020 \\ 0.000014 & 0.000045 & 0.000016 & 0.000020 & 0.000018 \\ 0.000015 & 0.000016 & 0.000039 & 0.000010 & 0.000012 \\ 0.000015 & 0.000020 & 0.000010 & 0.000040 & 0.000014 \\ 0.000020 & 0.000018 & 0.000012 & 0.000014 & 0.000035 \end{bmatrix}$$

System data for five thermal units is given in Table 6.

TABLE VI
SYSTEM PARAMETERS FOR FIVE-UNIT TEST SYSTEM [20]

Parameters	P ₁	P ₂	P ₃	P ₄	P ₅
a _i (\$/h)	25	60	100	120	40
b _i (\$/MWh)	2.0	1.8	2.1	2.0	1.8
c _i (\$/(MW) ² h)	0.0080	0.0030	0.0012	0.0010	0.0015
d _i (\$/h)	100	140	160	180	200
e _i (1/MW)	0.042	0.040	0.038	0.037	0.035
P _{i min} (MW)	10	20	30	40	50
P _{i max} (MW)	75	125	175	250	300
UR (MW/h)	30	30	40	50	50
DR (MW/h)	30	30	40	50	50

The 24-hour changing load data for solving the DED problem on test case 1 are given in Table 7.

TABLE VII
24-HOUR LOAD CHANGE DATA FOR FIVE-UNIT TEST SYSTEM [20]

Hour (h)	Load (MW)	Hour (h)	Load (MW)	Hour (h)	Load (MW)	Hour (h)	Load (MW)
1	410	7	626	13	704	19	654
2	435	8	654	14	690	20	704
3	475	9	690	15	654	21	680
4	530	10	704	16	580	22	605
5	558	11	720	17	558	23	527
6	608	12	740	18	608	24	463

The results show a total operating cost of 41,030.9994 \$/24h. In Table 8, the results (minimum, average and maximum value) of the CSA are compared with other results in the previously reported literature and it is seen that the proposed algorithm gives the best result.

TABLE VIII
FUEL COST VALUES FOR TEST CASE 1

Method	Min. (\$/24h)	Average (\$/24h)	Max. (\$/24h)
SA [22]	47,356.0000	-	-
GA [24]	44,862.4200	44,921.7600	45,893.9500
PSO [25]	44,253.2400	45,657.0600	46,402.5200
ABC [25]	44,045.8300	44,064.7300	44,218.6400
ICA [26]	43,117.0500	43,144.4700	43,302.2300
SOS [20]	43,090.5925	43,103.0828	43,162.2146
CSA	41,030.9994	41,079.5397	41,103.6947

TABLE IX
RESULTS FROM THE CROW SEARCH ALGORITHM FOR TEST CASE 2

Hour	P1 (MW)	P2 (MW)	P3 (MW)	P4 (MW)	P5 (MW)	P6 (MW)	P7 (MW)	P8 (MW)	P9 (MW)	P10 (MW)	Cost (\$/h)
1	153.5076	235.6846	81.6252	60.1069	88.9610	159.9993	129.7902	51.2789	20.0463	55	28.1963
2	174.1104	220.5821	158.5912	60.0262	84.7006	159.9844	129.9999	47.0064	20.0004	55	29.7468
3	207.0047	272.3180	227.2118	60.0006	78.8044	159.9958	129.6627	47.0054	21.0429	55	32.9361
4	224.8680	338.2997	295.7740	60.0042	73.0179	159.9943	129.6720	48.0839	21.2884	55	36.1262
5	194.4103	368.0764	339.9999	60.0000	102.9533	159.9999	129.9999	49.3601	20.2221	55	37.7134
6	273.1290	416.5603	339.9959	70.9246	114.2374	157.6254	129.9905	50.5381	20.0004	55	40.9957
7	348.4823	421.9858	339.9998	62.5551	119.1232	158.0220	129.9992	47.0290	20.0021	55	42.6297
8	413.1924	459.9158	339.2464	62.0455	85.2900	159.9999	129.9999	51.2597	20.0501	55	44.2668
9	469.6148	459.9999	339.9999	75.3451	131.9493	159.9999	129.9999	79.5731	22.5176	55	47.6461
10	469.7190	459.4386	339.3061	124.9983	180.6288	159.9999	129.9758	106.6304	46.3029	55	51.2269
11	469.4233	459.9995	339.9998	122.7567	227.7601	154.6573	129.9998	92.4028	20.0005	55	51.0756
12	468.3962	459.8852	339.9926	169.0843	242.7376	159.9998	129.9989	119.9861	34.9188	55	53.6800
13	449.4786	459.7576	339.9988	141.0363	223.0026	159.9983	129.9988	89.0004	24.7282	55	51.0958
14	408.6655	446.9922	339.2466	91.0022	211.2103	159.8815	129.9997	59.0053	22.9962	55	47.6328
15	355.1128	442.8886	329.4883	73.0002	163.7124	159.9999	129.3275	47.4679	20.0021	55	44.2889
16	275.0022	362.0988	289.7085	66.7447	146.8688	159.9909	129.9944	48.1494	20.4419	55	39.4097
17	195.0020	333.8780	321.4467	60.0063	159.5385	159.9194	128.2062	47.0001	20.0027	55	37.7423
18	203.1714	399.4876	339.9952	63.9597	208.0304	159.9653	129.9993	47.0211	21.3697	55	40.9917
19	282.9999	454.4229	339.7089	76.5863	209.1755	159.9995	129.5644	48.5425	20.0002	55	44.2861
20	361.9901	459.5435	339.9986	117.1854	242.9992	159.5332	129.9970	77.9877	22.7651	55	48.8086
21	399.9707	459.9998	338.1435	68.2714	242.9999	159.9249	129.9999	48.3217	21.3680	55	47.5604
22	323.2210	391.9479	265.8824	61.5723	192.0673	150.8171	120.3835	47.1076	20.0006	55	41.1339
23	248.1792	311.0990	185.8063	60.0043	122.1070	159.9983	112.7484	54.7652	22.2921	55	34.6970
24	168.0059	257.7098	167.4774	60.0001	16.6554	159.9993	129.9676	49.1781	20.0063	55	31.3638
Total Cost (\$/24h)											1,005,250.6

B. Test Case 2

There are ten thermal generation units in the test case 2. In this case, the DED problem was considered without a transmission losses. System data for ten thermal units is given in Table 10.

TABLE X
SYSTEMS PARAMETERS FOR TEN-UNIT TEST SYSTEM [32]

Parameters	P ₁	P ₂	P ₃	P ₄	P ₅	P ₆	P ₇	P ₈	P ₉	P ₁₀
a _i (\$/h)	958.2	1313.6	604.97	471.6	480.29	601.75	502.7	639.4	455.6	692.4
b _i (\$/MWh)	21.6	21.05	20.81	23.9	21.62	17.87	16.51	23.23	19.58	22.54
c _i (\$/(MW ² h))	.00043	.00063	.00039	.0007	.00079	.00056	.00211	.0048	.10908	.00951
d _i (\$/h)	450	600	320	260	280	310	300	340	270	380
e _i (1/MW)	.041	.036	.028	.052	.063	.048	.086	.082	.098	.094
P _{i,max} (MW)	150	135	73	60	73	57	20	47	20	55
P _{i,min} (MW)	470	460	340	300	243	160	130	120	80	55
UR (MW/h)	80	80	80	50	50	50	30	30	30	30
DR (MW/h)	80	80	80	50	50	50	30	30	30	30

The 24-hour changing load data for solving the DED problem on test case 2 are given in Table 11.

TABLE XI
24-HOUR LOAD CHANGE DATA FOR TEN-UNIT TEST SYSTEM [20]

Hour (h)	Load (MW)	Hour (h)	Load (MW)	Hour (h)	Load (MW)	Hour (h)	Load (MW)
1	1036	7	1702	13	2072	19	1776
2	1110	8	1776	14	1924	20	2072
3	1258	9	1924	15	1776	21	1924
4	1406	10	2072	16	1554	22	1628
5	1480	11	2146	17	1480	23	1332
6	1628	12	2220	18	1628	24	1184

The operating value results of the generators are shown in

Table 9. The results show a total operating cost of 1,005.250.6 \$/24h. In Table 12, the results (min., average and max.) of the CSA are compared with other results in the previously reported literature and it is seen that proposed algorithm gives the best result.

TABLE XII
FUEL COST VALUES FOR TEST CASE 2

Method	Min (\$/24h)	Average (\$/24h)	Max (\$/24h)
SQP [27]	1,051,163.0000	-	-
PSO [28]	1,036,506.0000	1,040,496.0000	-
EP-SQP [29]	1,034,100.0000	-	-
PSO-BF [30]	1,026,537.2600	1,028,826.7400	1,033,565.2700
CS-DE [31]	1,023,432.0000	1,026,475.0000	1,027,634.0000
AIS [24]	1,021,980.0000	1,023,156.0000	1,024,973.0000
ABC [24]	1,021,576.0000	1,022,686.0000	1,024,316.0000
CDBCO [32]	1,021,500.0000	1,024,300.0000	-
SOS [20]	1,020,894.0757	1,021,072.6846	1,021,194.9972
MACO [33]	1,019,093.1700	1,019,254.2100	1,024,310.8000
BBOSB [34]	1,017,530.3328	1,018,487.8504	1,031,843.5673
CSA	1,005,250.6000	1,007,546.0000	1,009,304.0000

VI. CONCLUSION

The minimum fuel cost and operating conditions of a power system are found by the economic dispatch problem. However, for a more realistic analysis, this problem should be solved by considering the 24-hour load change. While making hourly load sharing of generators, ramp rate limits were taken into consideration. With the proposed algorithm, the DED

problem was solved on two different test systems specified in the literature. The CSA results show that it is operated with the lowest fuel cost in the conditions determined in the test case 1. With the results of the CSA, the test system 1 is operated at a cost of \$41,030,9994 per day. With the results of the SOS algorithm, which is the best result in the literature, the test system 1 is operated at a cost of \$43,090,5925 per day. With the proposed algorithm, a daily profit of \$2,059.5931 is obtained. The CSA results show the operation of the system with the lowest fuel cost under the conditions specified in the initial test system (including losses). In the test case 2, the number of thermal generation units increases, but the losses are neglected. In this test system, the CSA found the lowest operating conditions (min., average and max.). The results of the CSA show that test system 2 is operated at a cost of \$1,005,250,6000 per day. With the results of the BBOSB algorithm, which is the best result in the literature, this system is operated with a daily cost of \$1,017,530,3328. With the proposed algorithm, a daily profit of \$12,279.7328 is obtained. The results show that the CSA gives better results than other algorithms mentioned in the literature. By using the CSA, a yearly profit of \$751,751.4815 and \$4,482,102.472 are provided in test system 1 and 2, respectively.

With the CSA algorithm, multi-objective DED problems can also be successfully solved, for example environment emission dispatch and including photovoltaic (PV) and energy storage systems and electric vehicles (EV).

REFERENCES

- [1] Vlachogiannis, J. G., & Lee, K. Y. (2009). Economic load dispatch—A comparative study on heuristic optimization techniques with an improved coordinated aggregation-based PSO. *IEEE Transactions on Power Systems*, 24(2), 991-1001.
- [2] Jayabarathi, T., Jayaprakash, K., Jeyakumar, D. N., & Raghunathan, T. (2005). Evolutionary programming techniques for different kinds of economic dispatch problems. *Electric power systems research*, 73(2), 169-176.
- [3] Mandal, B., & Roy, P. K. (2021). Dynamic economic dispatch problem in hybrid wind based power systems using oppositional based chaotic grasshopper optimization algorithm. *Journal of Renewable and Sustainable Energy*, 13(1), 013306.
- [4] Xiong, G., & Shi, D. (2018). Hybrid biogeography-based optimization with brain storm optimization for non-convex dynamic economic dispatch with valve-point effects. *Energy*, 157, 424-435.
- [5] Zheng, Z., Li, J., & Han, Y. (2020). An improved invasive weed optimization algorithm for solving dynamic economic dispatch problems with valve-point effects. *Journal of Experimental & Theoretical Artificial Intelligence*, 32(5), 805-829.
- [6] Dai, C., Hu, Z., & Su, Q. (2021). An adaptive hybrid backtracking search optimization algorithm for dynamic economic dispatch with valve-point effects. *Energy*, 122461.
- [7] Younes, Z., Alhamrouni, I., Mekhilef, S., & Reyasudin, M. (2021). A memory-based gravitational search algorithm for solving economic dispatch problem in micro-grid. *Ain Shams Engineering Journal*, 12(2), 1985-1994.
- [8] Sinha, N., Chakrabarti, R., & Chattopadhyay, P. K., (2003). Evolutionary programming techniques for economic load dispatch. *IEEE Transactions on Evolutionary Computation*, 7(1), 83-94.
- [9] Irving, M.R., & Sterling, M.J.H., (1983). Economic dispatch of active power with constraint relaxation. *IEEE Proceedings C - Generation, Transmission and Distribution*, 130(4), 172-177.
- [10] Chowdhury, B. H., & Rahman, S., (1990). A Review of recent advances in economic dispatch. *IEEE Transactions on Power Systems*, 5(4), 1248-1259.
- [11] Yang, H. T., Yang, P. C., & Huang, C. L. (1996). Evolutionary programming based economic dispatch for units with non-smooth fuel cost functions. *IEEE transactions on Power Systems*, 11(1), 112-118.
- [12] Ross, D. W., & Kim, S. (1980). Dynamic economic dispatch of generation. *IEEE transactions on power apparatus and systems*, 6(1), 2060-2068.
- [13] Yalcinöz, T., & Altun, H. (2000). Comparison of simulation algorithms for the Hopfield neural network: an application of economic dispatch. *Turkish Journal of Electrical Engineering & Computer Sciences*, 8(1), 67-80.
- [14] Zou, D., Li, S., Kong, X., Ouyang, H., & Li, Z. (2019). Solving the combined heat and power economic dispatch problems by an improved genetic algorithm and a new constraint handling strategy. *Applied energy*, 237, 646-670.
- [15] Mahor, A., Prasad, V., & Rangnekar, S. (2009). Economic dispatch using particle swarm optimization: A review. *Renewable and sustainable energy reviews*, 13(8), 2134-2141.
- [16] Sonmez, Y. (2011). Multi-objective environmental/economic dispatch solution with penalty factor using Artificial Bee Colony algorithm. *Scientific Research and Essays*, 6(13), 2824-2831.
- [17] Andic, C., Ozturk, A., & Tosun, S. (2020). Türkiye'deki güc sisteminde karga arama algoritması kullanılarak ekonomik yük dağıtımı. *Düzce Üniversitesi Bilim ve Teknoloji Dergisi*, 8(1), 428-436.
- [18] Pradhan, M., Roy, P. K., & Pal, T. (2018). Oppositional based grey wolf optimization algorithm for economic dispatch problem of power system. *Ain Shams Engineering Journal*, 9(4), 2015-2025.
- [19] Al-Bahrani, L., Seyedmahmoudian, M., Horan, B., & Stojcevski, A. (2021). Solving the real power limitations in the dynamic economic dispatch of large-scale thermal power units under the effects of valve-point loading and ramp-rate limitations. *Sustainability*, 13(3), 1274.
- [20] Sonmez, Y., Kahraman, H. T., Dosoglu, M. K., Guvenc, U., & Duman, S. (2017). Symbiotic organisms search algorithm for dynamic economic dispatch with valve-point effects. *Journal of Experimental & Theoretical Artificial Intelligence*, 29(3), 495-515.
- [21] Sahoo, A. K., Panigrahi, T. K., Paramguru, J., & Hota, A. P. (2021). Dynamic economic dispatch using harmony search algorithm. In *Advances in machine learning and computational intelligence*, 425-435.
- [22] Panigrahi, C. K., Chattopadhyay, P. K., Chakrabarti, R. N., & Basu, M. (2006). Simulated annealing technique for dynamic economic dispatch. *Electric Power Components and Systems*, 34, 577-586.
- [23] Askarzadeh, A. (2016). A novel metaheuristic method for solving constrained engineering optimization problems: crow search algorithm. *Computers & Structures*, 169, 1-12.
- [24] Hemamalini, S., & Simon, S. P. (2011a). Dynamic economic dispatch using artificial immune system for units with valve-point effect. *International Journal of Electrical Power & Energy Systems*, 33, 868-874.
- [25] Hemamalini, S., & Simon, S. P. (2011b). Dynamic economic dispatch using artificial bee colony algorithm for units with valve-point effect. *European Transactions on Electrical Power*, 21, 70-81.
- [26] Mohammadi-Ivatloo, B., Rabiee, A., Soroudi, A., & Ehsan, M. (2012). Imperialist competitive algorithm for solving non-convex dynamic economic power dispatch. *Energy*, 44, 228-240.
- [27] Attaviriyapap, P., Kita, H., Tanaka, E., & Hasegawa, J. (2002). A hybrid EP and SQP for dynamic economic dispatch with nonsmooth fuel cost function. *IEEE Transactions on Power Systems*, 17, 411-416.
- [28] Yuan, X., Su, A., Yuan, Y., Nie, H., & Wang, L. (2009). An improved PSO for dynamic load dispatch of generators with valve-point effects. *Energy*, 34, 67-74.
- [29] Basu, M. (2013). Hybridization of bee colony optimization and sequential quadratic programming for dynamic economic dispatch. *International Journal of Electrical Power & Energy Systems*, 44, 591-596.
- [30] Saber, A. Y. (2012). Economic dispatch using particle swarm optimization with bacterial foraging effect. *International Journal of Electrical Power & Energy Systems*, 34, 38-46.
- [31] He, D., Dong, G., Wang, F., & Mao, Z. (2011). Optimization of dynamic economic dispatch with valve-point effect using chaotic sequence based differential evolution algorithms. *Energy Conversion and Management*, 52, 1026-1032.
- [32] Lu, P., Zhou, J., Zhang, H., Zhang, R., & Wang, C. (2014). Chaotic differential bee colony optimization algorithm for dynamic economic dispatch problem with valve-point effects. *International Journal of Electrical Power & Energy Systems*, 62, 130-143.

- [33] Secui, D. C. (2015). A method based on the ant colony optimization algorithm for dynamic economic dispatch with valve-point effects. *International Transactions on Electrical Energy Systems*, 25(2), 262-287.
- [34] Xiong, G., & Shi, D. (2018). Hybrid biogeography-based optimization with brain storm optimization for non-convex dynamic economic dispatch with valve-point effects. *Energy*, 157, 424-435.
- [35] Gharehchopogh, F. S., Shayanfar, H., & Gholizadeh, H. (2020). A comprehensive survey on symbiotic organisms search algorithms. *Artificial Intelligence Review*, 53(3), 2265-2312.
- [36] Zou, D., & Gong, D. (2022). Differential evolution based on migrating variables for the combined heat and power dynamic economic dispatch. *Energy*, 238, 121664.
- [37] Ali Shaabani, Y., Seifi, A. R., & Kouhanjani, M. J. (2017). Stochastic multi-objective optimization of combined heat and power economic/emission dispatch. *Energy*, 141, 1892-1904.
- [38] Mohammadi-Ivatloo, B., Moradi-Dalvand, M., & Rabiee, A. (2013). Combined heat and power economic dispatch problem solution using particle swarm optimization with time varying acceleration coefficients. *Electric Power Systems Research*, 95, 9-18.

BIOGRAPHIES



research interests include economic dispatch and power system, state estimation and renewable energy.

CENK ANDIC received the B.S. degrees in electrical-electronics engineering from the University of Duzce, Turkey in 2018 where he is currently pursuing the Ph.D. degree in electrical-electronics engineering. He has been an Research Assistant with the Electrical Engineering Department, Istanbul Technical University (ITU). His



ALI OZTURK received his B.S. in Electrical Engineering Yildiz Technical University, Turkey, in 1996, his M.Sc. in Electrical Engineering from Sakarya University, Sakarya, Turkey, in 2001 and his Ph.D. in 2007 from the same university. His research interests include voltage stability and optimization in electric power systems.



SALIH TOSUN received his B.S. in Electrical Education Marmara University, Turkey, in 1986, his M.Sc. in Electrical Engineering from Marmara University, Istanbul, Turkey, in 1991. His research interests include voltage stability and optimization in electric power systems.

4 Wheel Steering System Control Unit Design

Aynur Tuba Çakmak, Ergun Erçelesi


Abstract— Ease of driving and flexibility in movements are very important for the developing automobile technology today. The aim of the study is to provide convenience to the driver by trying different modes in the steering system of a vehicle. In this study, an electronic board has been designed for the steering system for the 4 wheels in the vehicle. The card developed in the study provides choosing from 3 different modes as standard, circle and crab to driver. The driver will have the advantage of both comfort and safe driving with the mode chosen according to the road and load conditions. Also, the 4-wheel steering system improves handling stability and active safety at high speeds. Using 4-wheel steering at low speeds can reduce the turning radius on the steering wheel for more convenient maneuvers. Positive feedback was received from drivers who tried the card on heavy-duty vehicles. Tests continue on different heavy-duty vehicles. If the electronic board and software do not perform as expected in different heavy vehicles, improvements will be made in the design. However, the results obtained so far are satisfactory. It shows good response performances, also improves ride quality and stability.

Index Terms— Electronics control unit design, Steering system, 4-Wheel steering.


I. INTRODUCTION

NOWADAYS the competition among automobile manufacturers all over the world and the awareness of consumers have led to the need to produce more perfect vehicles. In the design of a vehicle, frame features (wind resistance coefficient, shape, viewing angles, etc.), interior comfort (air conditioning, electric windows, electric control seats, etc.) workmanship quality (application of leather upholstery, plastic quality used in the console, trim success, etc. and safety accessories (airbag, inside door protection bars, ABS, etc.) constitute important criteria. Many cars produced today have most of the features listed above. In past, the variety of accessories is a dominant factor in the sale of a car. Nowadays the health and safety of drivers and passengers has become crucial and a step further. Automobile road safety criteria have been established by independent organizations and decisions are made by these organizations, whether the automobile will be produced or not due to tests [1].

AYNUR TUBA ÇAKMAK, is with Department of Electrical Engineering University of Gaziantep University, Gaziantep, Turkey, (e-mail: aynurtubacakmak@gmail.com).

 <https://orcid.org/0000-0003-0083-3371>

ERGUN ERÇELEBİ, is with Department of Electrical Engineering University of Gaziantep University, Gaziantep, Turkey, (e-mail: ercelebi@gantep.edu.tr).

 <https://orcid.org/0000-0002-4289-7026>

Manuscript received July 10, 2021; accepted May 23, 2022.
DOI: [10.17694/bajece.969569](https://doi.org/10.17694/bajece.969569)

In the developing technology age, it is seen that many studies have been carried out to develop mechanical and electronic systems. Wheel steering systems have had their share in this development process. Many methods have been developed and studied to increase the flexibility of movements in wheel steering systems.

Wheel steering systems are a defining part for whole vehicle styles. It shows the link among vehicle and driver. Its primary mission is to direct a vehicle in the wanted path securely and under exact control. Owing to the significance of the steering system, it must be secure and basic [2]. Conventional steering system design is usually done mechanically.

Thanks to growth of control and electronic technologies, the four-wheel steering (4WS) system has been researched as a vehicle maneuvering technique that increases the maneuverability of vehicles at low velocities and improves their determination at high velocities. On account of increase control efficiency, since the first 4WS system was important, much work has been given to the control strategies of the 4WS system, especially the steering stability controller project, which has been an investigate subject of interest in recent years. It is also well known that processing stability control with different unknowns is a complicated and nonlinear process [3].

To reduce the cornering radius, the rear wheels should be turned in the opposite direction to the front wheels at low speeds. If the rear wheels turn in identical direction as the front wheel, it improves the vehicle's ability to change lanes and stability at medium and high speeds. One of the major advantages of all-wheel steering is the resetting of the lean angle and the other is to increase the dynamic response capability. Theoretical studies show that when the steering angle of the rear wheels is the only control data, the four-wheel steering cannot properly follow the yaw and lateral acceleration. Particularly high speed cornering, with the constant yaw of the vehicle steered by four wheels, its front wheels are less than the yaw of the steered vehicle. Therefore, a greater wheel steering angle is required for the similar bend radius, which enhances the user's work.

Stanislav et al. [4] examined the simulation and modeling of a land-road vehicle with a 4WS system. In the study, it was aimed to develop a steering system that provides driving safety, maneuverability and driving quality. Their result is that the vehicle with the 4WS has preferable maneuverability when the rear wheels are steered in the reverse direction to the front than when steering from the front.

Li et al. [5] worked by testing the absence of classic 4WS. With the help of CarSim's and Matlab / Simulink programs, the simulation of the proposed control system was created and

evaluated. The simulation conclusions indicate that with the designed 4-wheel active steering (4WAS) controller, good response performances could be achieved, and could be improved quality and stability of the ride.

Barec et al. [6] developed a control system for four-wheel steering vehicles. Then, they created a sample of the vehicle with 4WAS on a scale of 1: 5 to test this system. As a result of these samples, the yaw velocity and lateral acceleration can be adjusted.

Yi et al. [7] developed a rear-wheel steering (RWS) control algorithm to improve vehicle usage success with no need for tire features. The algorithm has been examined through computer simulations. The simulation has been tested for step steering and waiting for sinus scenarios under different road friction cases. Simulation results showed that the tried algorithm caused an increase in vehicle driving performance.

When the studies in the literature are examined, it is seen that the studies are generally done on simulation programs or scaled models. The results obtained indicate that it increases driving comfort and security.

In this study, the 4WS electronic card, which can be used especially in work machines, has been designed and tested on work machines. In the literature, the advantages and disadvantages of using the 4-wheel steering process, which is usually done on simulation and used in cars, on work machines have been investigated.

II. PROPOSED METHOD

The turn control system receives feed from the oil of the P2 pump through the priority valve shown in Fig. 1 The priority valve also acts as a flow control valve, managed by the load signal (LS) pressure from the turn orbital valve. With the increase of the load in the turn circuit, the load signal pressure affecting the priority valve increases and more oil is sent to the turn circuit. With the decrease of the load signal pressure, less goes from the priority valve to the turn circuit.

This valve is used to turn the rear wheels to the right / left in order to increase manoeuvrability in machines with large front wheels. With the multi-purpose turn control valve, the operator can operate the machine in the following turn modes with the

selection made from the turn mode selection switch in the cabin:

- Standard Turn Mode: In this mode, only the front axle moves either clockwise or counter clockwise and the rear wheel does not. This is the drive that we see in day to day life in all the four wheelers. It is generally used at moderate speed.
- Circle Turn Mode: In this mode, both the front and rear axles move in opposite directions relative to each other. Since both axles move in different directions, the radius of curvature decreases during turning. This drive is mainly used during parking of the vehicle. This means the vehicle will require less space for parking and this will be helpful in places where traffic and parking is a major problem. [8]
- Crab Turn Mode: In this mode, the front and rear axle wheels of the machine turn to the same side. In this mode, the machine crab walks. This drive enabled vehicle to change the lane during highway driving. It is generally applied at higher speed. [9]

Fig. 2. shows the positions of the wheels according to the modes as described above.

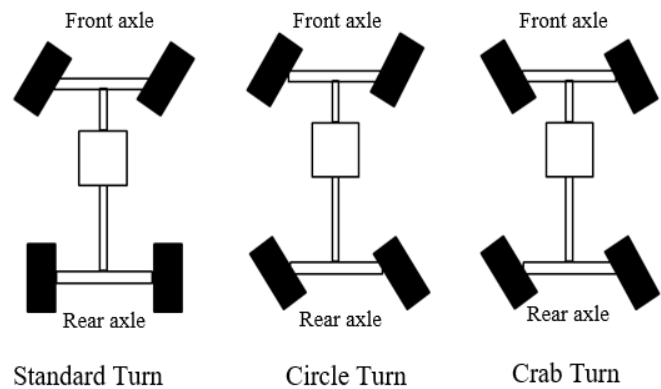


Fig.2. System with turn circuit 4WS

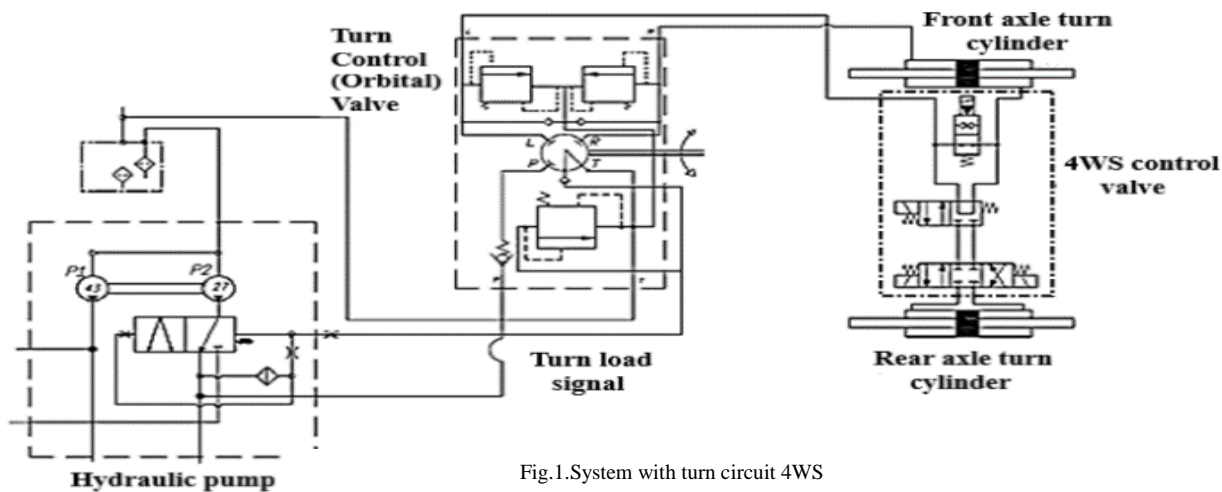


Fig.1.System with turn circuit 4WS

III. DESIGN OF CONTROL UNIT 4WS ELECTRONIC

In Fig.3., turn system electric circuit shows the Y20 turn solenoid, the Y21 rear axle turn solenoid, the Y22 circle turn solenoid and the Y24 crab turn solenoid.

understandable and suitable for future developments. AMP-346350-5 (PCB Connector) [10] has been utilized to communicate between the gearbox and the circuit board.

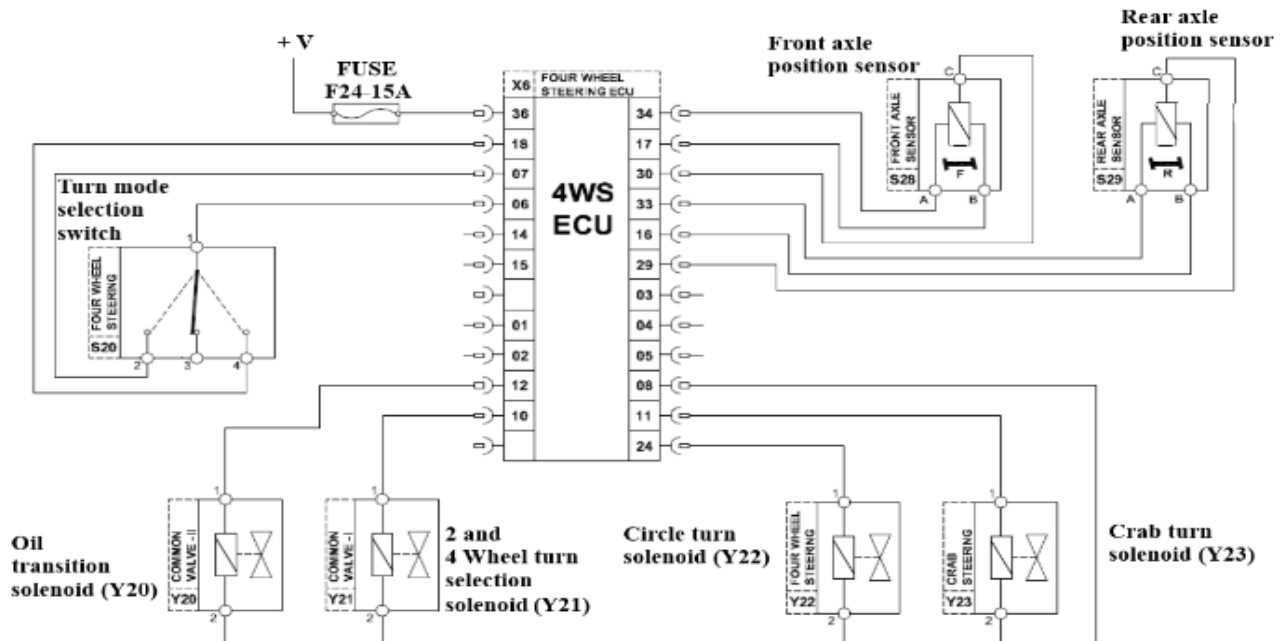


Fig.3. Electrical circuit diagram of turn system

Standard Turn Mode: In this mode, all solenoids are in the off state. Oil from the turn control valve goes directly to one side of the front axle cylinder and to the other through the Y20 and Y21 solenoid valves. No oil goes to the rear axle turn cylinder. Thus, in this mode, only the front axle wheels turn when the operator turns the steering wheel.

Circle Turn Mode: In this mode, Y20, Y21 and Y22 solenoid valves are energized (ON). With the energization of the Y20 and Y21 solenoids, the left side of the front axle turn cylinder is connected to the Y22, Y23 solenoid valve group. In this mode, when the machine turns right with the energization of the circle turn Y22 solenoid, the oil coming out from the left side of the front axle cylinder goes to the left side of the rear axle cylinder. While the machine rotates to the left, the oil flow between the axle cylinders occurs in the opposite direction.

Crab Turn Mode: In this mode, Y20, Y21 and Y23 solenoid valves are energized (ON). With the energization of the Y20 and Y21 solenoids, the left side of the front axle turn cylinder is connected to the Y22, Y23 solenoid valve group. In this mode, with the energization of the crab turn Y23 solenoid, the front wheels turn right, while the oil from the left side of the front axle cylinder goes to the right side of the rear axle cylinder. While the front wheels turn to the left, the oil flow between the axle cylinders occurs in the opposite direction. The card seen in Fig.4. was designed in the EAGLE program and the PCB drawing was made in the same program. Esp-12 Wi-Fi module has been used to communicate with the card remotely. As a processor, the ATmega32U4 has been preferred because it provides sufficient input and output, which is affordable, and easily accessible. The C computer language was used to program the processor and it was compiled using the Arduino interface. The codes were tried to be written as simple,

IV. MEASURED RESULTS

A. Electrical Tests of 4WS ECU

Measurements are performed at laboratories by using multimeter. The designed card has 12V voltage, 11A current and 132W power dissipation. The measured results of the 4WS ECU are shown in Table 1.

TABLE I
MEASURED RESULTS OF THE 4WS ECU

Rating	Measured
Voltage	12V
Curent	11A
Power	132W

The operating temperature range is -15 degrees C to 85 degrees C. The operating temperature is shown in Table 2.

TABLE II
OPERATING TEMPERATURE OF THE 4WS ECU

Rating	Max.	Min.
Operating Temperature	-15 °C	85 °C

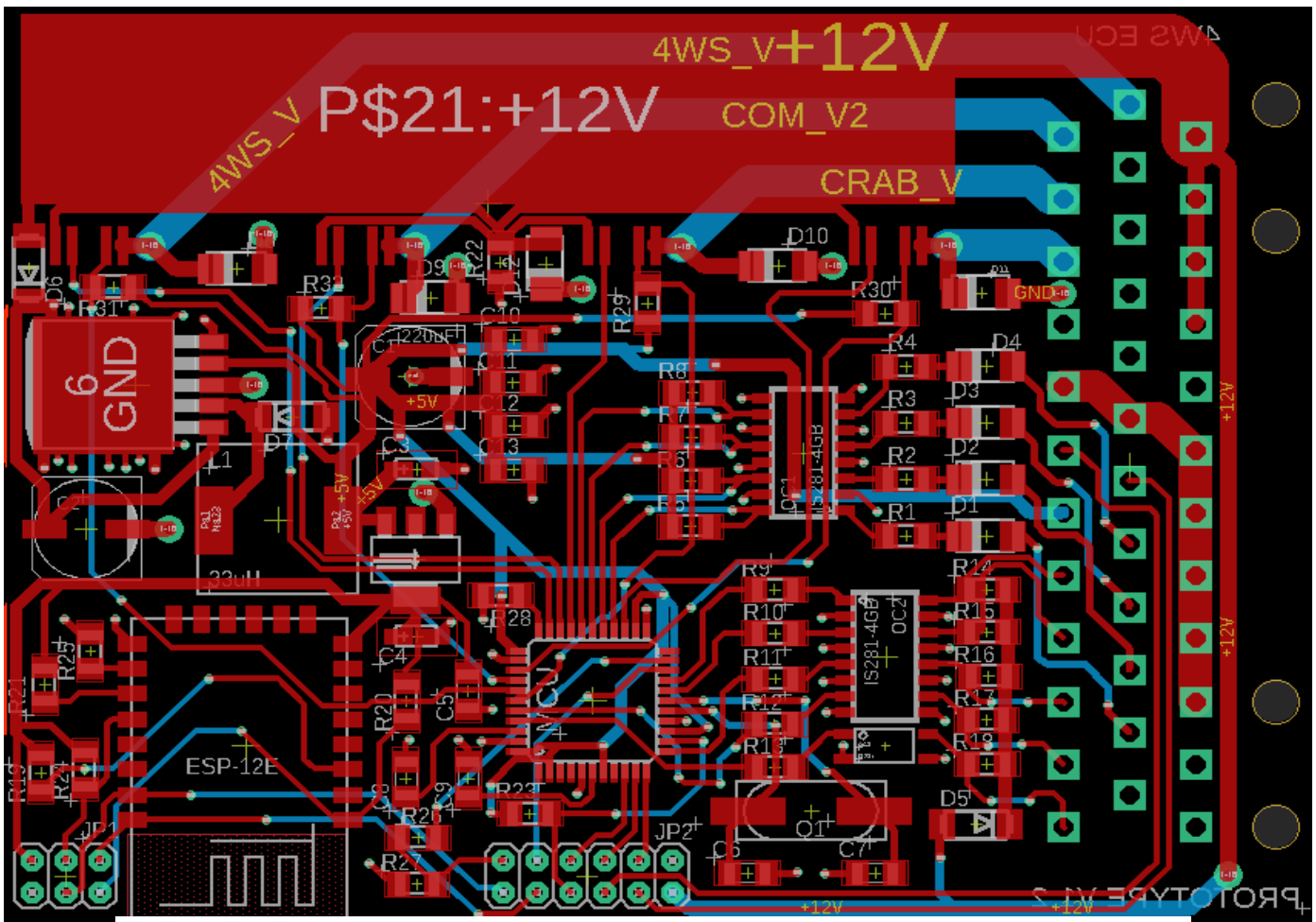


Fig.4. 4WS eagle schematic



Fig.5. Work machine with 4WS ECU

4ws card was attached to the work machine with the help of connector to obtain electrical test results. The set up can be seen in Fig.5.

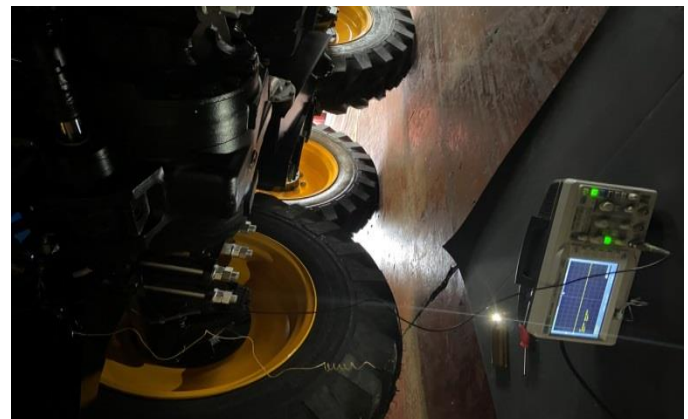


Fig.6. Measurement setup

The measurements are performed at factory by using oscilloscope. First, the signal input and GND ports of the oscilloscope are connected to the 4WS ECU. The set up can be seen in Fig.6.

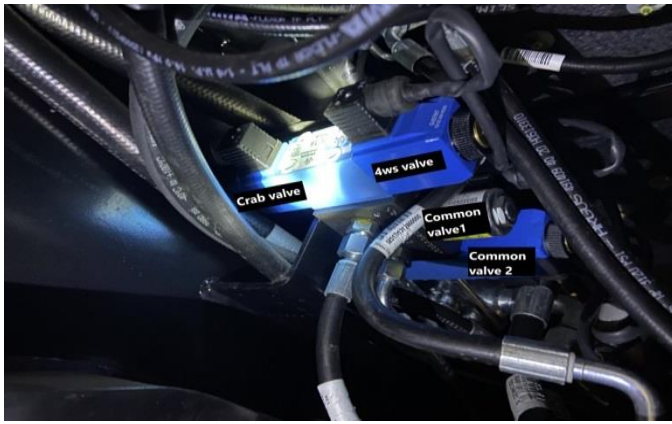


Fig.7. Condition of valves in standard rotation mode

If standard rotation mode is selected when the machine is activated, all solenoids are in the off state. The set up can be seen in Fig.7.

The solenoids output voltages were measured as 0V, the screen shot can be seen in Fig.8.

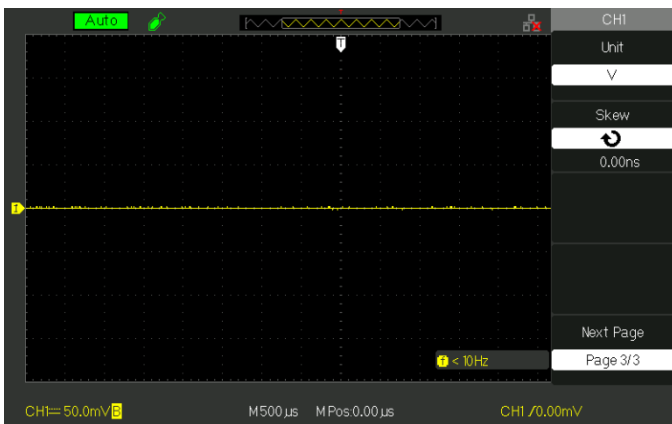


Fig.8. Output Voltages of the all solenoid valves in the standard rotation mode

If circle rotation mode is selected when the machine is activated, 4ws, common valve 1 and common valve 2 solenoid valves are energized (ON). The set up can be seen in Fig.9.

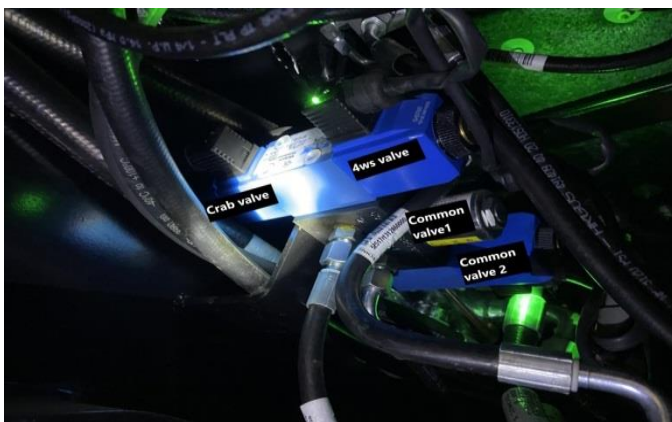


Fig.9. Condition of valves in circle rotation mode

The 4ws valve, common valve 1 and common valve 2 output voltages were measured as 12V when crab valve output voltage was measured as 0V in circle rotation mode, the screen shots can be seen in Fig.10.and Fig.11.

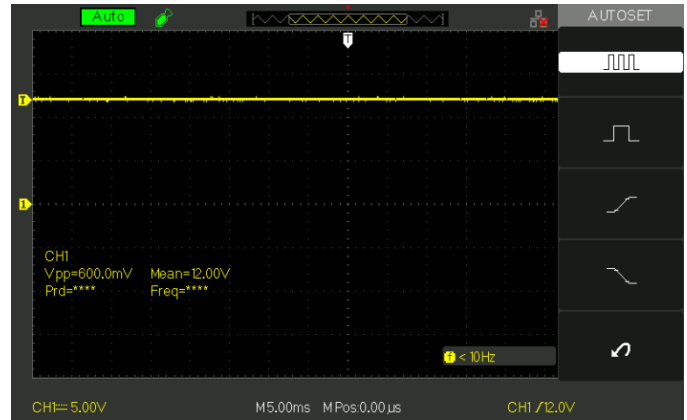


Fig.10. Output voltages of the 4ws, common valve 1 and common valve 2 solenoid valves in the circle rotation mode

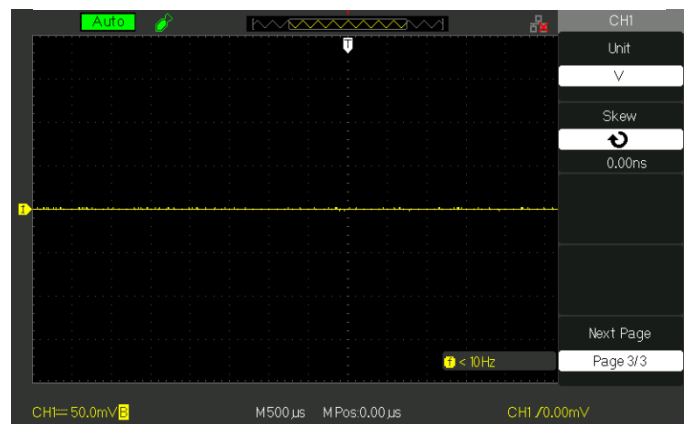


Fig.11. Output voltages of the crab solenoid valve in the circle rotation mode

If crab rotation mode is selected when the machine is activated, crab, common valve 1 and common valve 2 solenoid valves are energized (ON). The set up can be seen in Fig.12.



Fig.12. Condition of valves in crab rotation mode

The crab valve, common valve 1 and common valve 2 output voltages were measured as 12V when 4ws valve output voltage was measured as 0V in the crab rotation mode, the screen shots can be seen in Fig.13.and Fig.14.

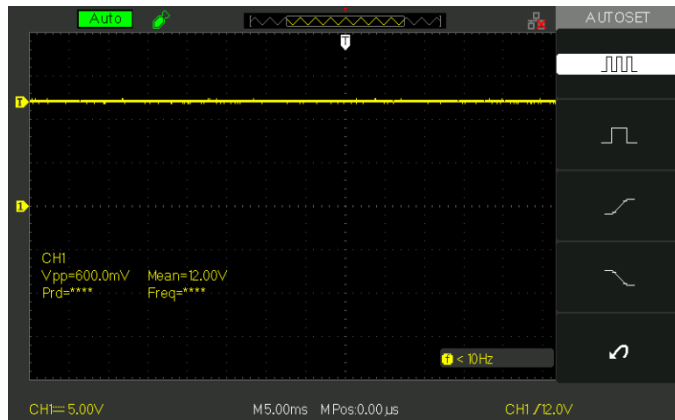


Fig.13. Output voltages of the crab valve, common valve 1 and common valve 2 in the crab rotation mode

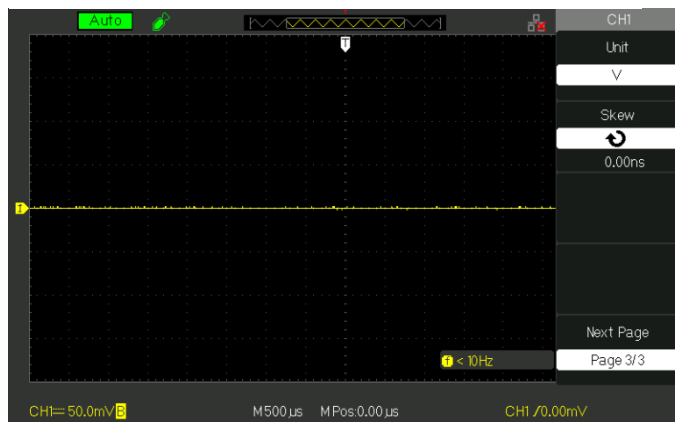


Fig.14. Output voltages of the 4ws valve in the crab rotation mode

B. Experimental Results

In order to test the 4WS card, work machines with and without the 4WS card installed, closed the one-kilometer channel with the same driver (loading, unloading and pushing the sand into the channel). As a result of these same procedures, the required labor hours, and the required fuels were calculated and the required maintenance periods in the annual period are shown in Table III.

TABLE III
OPERATING TEMPERATURE OF THE 4WS ECU

Type of Vehicle	Process completion time	Fuel consumption amount	Number of annual maintenance
With 4WS	3 hours	24 liters	4 times
Without 4WS	5 hours	35 liters	2 times

According to Table 3, fuel consumption and completion time are decreasing since maintenance time of the valves is increasing.

V. CONCLUSIONS AND RECOMMENDATIONS

Throughout the study, the requirements of the 4WS system were decided initially. Turn control valve and multipurpose turn control valve were selected according to these requirements. After the pre-feasibility and preliminary computation are finished, the 4WS system has been designed. Mechanical and electronic parts of the system have been described. The system was first designed with EAGLE. After the modeling is completed, the assembly of the prototype electronic card has been completed and the accuracy of the parts has been tested. Then, the software of the vehicle was completed to perform field tests and the conclusions on the field were observed. It was observed that the system was working appropriately by drivers. During the study, activities concerned to the 4WS system are included. 4WS system has been approved by testing the studies in the field as an academic and a real system.

Figure 15 shows the prototype and first version V1.1 of the card.

Benefits of the system:

- Car more efficient and stable on cornering.
- Improved steering responsiveness and precision.
- Notable improvement in rapid, easier, safer lane changing maneuvers.
- Smaller turning radius and tight space maneuverability at low speed.
- Relative Wheel Angles and their Control.
- Risk of hitting an obstacle is greatly reduced.

The system, designed to increase the safety and comfort of vehicles, is currently being actively used and tested. Security vulnerabilities related to rotations are found in the machines, and efforts are made to solve the problems. With this project, the fact that the software and hardware are under our own control, with the great advantages it provides in terms of cost, and changes and developments have become easier. Today, new approaches are emerging in line with the increasing demands for safety in vehicles and transport systems. Among these studies, autonomous vehicles and their applications come to the fore. Autonomous vehicle applications, which gain more importance and become widespread day by day, are considered as the vehicles of the future. For this reason, in this research study, general information about wheel steering applications, which has an important role in vehicles, has been given and an electronic board design has been made. Although these systems exist to support the driver for the time being, it is predicted that the driver will soon support these systems, and even the driver will be able to travel and use them without looking at the road in the vehicle. With the card designed in this study, it is thought that future studies will be carried out on topics such as artificial intelligence and data mining, which are popular today. Starting with the main major problems, modifications and improvements are made on the board to find the best suitable solution. Last word, the card developed in this study was being imported. With the use of the card in vehicles, it will reduce imports and contribute to the country's economy. The card was developed within the scope of university and industry cooperation.

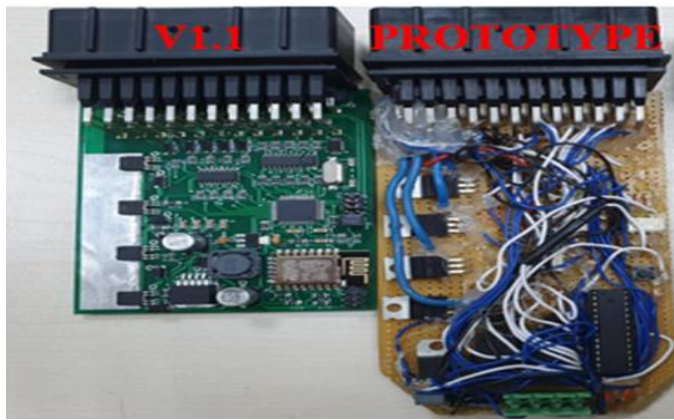


Fig.15. 4WS ECU V1.1. and prototype



ERGUN ERÇELEBİ received B.S. degree in Electrical and Electronics Engineering from METU, Gaziantep, Turkey in 1990 and M.S. and Ph.D. degrees in Electrical and Gaziantep in 1992 and 1999 respectively. Since 2011, he has been a Professor with the Electrical and Electronics Engineering Department at University of Gaziantep in Gaziantep, Turkey. Since 2014, he has been the head of the Electrical and Electronics Engineering Department. His research interests include embedded systems, signal processing and artificial neural networks.

REFERENCES

- [1] A.A.Sarioğlu (2002). Taşıtlar Yanal Kayma ve Kontrolü, (Master dissertation, Institute of Science, Istanbul University).
- [2] Lajqi, S. (2013). Suspension and Steering System Development of a Four Wheel Drive and Four Wheel Steered Terrain Vehicle (Doctoral dissertation, Univerza v Mariboru (Slovenia)).
- [3] Xu, F. X., Liu, X. H., Chen, W., Zhou, C., & Cao, B. W. (2019). Improving handling stability performance of four-wheel steering vehicle based on the H_2/H_∞ robust control. *Applied Sciences*, 9(5), 857.
- [4] Pehan, S., Lajqi, S., Pšeničnik, J., & Flašker, J. (2011). Modeling and simulation of off road vehicle with four wheel steering. In *IRMES Conference Proceedings* (pp. 77-83).
- [5] Li, B., & Yu, F. (2009, June). Optimal model following control of four-wheel active steering vehicle. In *2009 International Conference on Information and Automation* (pp. 881-886). IEEE.
- [6] Barec, P., Maly, M., & Vozenilek, R. (2004, October). Control System of Vehicle Model with Four Wheels Steering (4WS). In *International Scientific Meeting Motor Vehicles and Engines* (pp. 1-7).
- [7] Park, K., Joa, E., Yi, K., & Yoon, Y. (2020). Rear-Wheel Steering Control for Enhanced Steady-State and Transient Vehicle Handling Characteristics. *IEEE Access*, 8, 149282-149300.
- [8] Tim Gilles (2014), "Automotive Service Inspection, Maintenance, Repair".
- [9] Jack Erjavec (2009), "Automotive Technology—A System Approach", Cengage Learning, 5th Edition.
- [10] <https://www.te.com/usa-en/product-346350-5.html>

BIOGRAPHIES



AYNUR TUBA ÇAKMAK received B.Sc. degree in Electrical and Electronics Engineering from Gaziantep University, Turkey, in 2018. She is currently a M.Sc. student at Department of Electrical and Electronics Engineering at Gaziantep University. Her research interests include circuit design, embedded system, c

software.

Development of a Counterfeit Vehicle License Plate Detection System by Using Deep Learning

Burak Ağgöl and Gökhan Erdemir


Abstract— In this study, a deep learning-based counterfeit plate detection system that compares and detects vehicles with the make, model, color, and license plate is designed. As known that the relevant government institutions are responsible for keeping all detailed information about all motor vehicles in their database. All registration details are stored in the database. It is possible to find unregistered vehicles by comparing database records with detected details. In general, vehicles with counterfeit license plates are used in illegal actions. Therefore, it is of great importance to detect them. Generally, license plate recognition systems successfully detect counterfeit license plates that are randomly generated. Security units typically use such systems at toll roads, bridge crossings, parking lot entrances and exits, sites, customs gates, etc. This kind of system only checks the plate is exists or not in the database. But it is unsuccessful if the vehicle uses existing plate numbers such as stolen ones. In this study, the developed system can detect not only vehicles' plate numbers but also make, model, year, and color information by using deep learning. Thus, the system can also detect randomly generated plates and stolen plates that belong to another vehicle.

Index Terms— deep learning, convolutional neural networks (CNN), counterfeit plate.


I. INTRODUCTION

DEEP LEARNING is a machine learning method that refers to deep neural networks. It is a system that allows us to train artificial intelligence to predict the outputs of a given dataset using artificial neural networks (ANN). Deep learning is a more advanced form of machine learning and ANN. The main difference is that it creates new features by learning from the data itself. While certain parts need to be defined and created correctly by users in machine learning, in deep learning, the system is expected to produce and label features and eventually produce an output using ANN [1]. In this study, a system that recognizes specific vehicle brand models, license plate information, and color information has been created. The system's primary purpose is to find the make and

BURAK AĞGÖL, is with Computer Technologies Department, Istanbul Ayyansaray University, Istanbul, Turkey, (burakaggul@ayvansaray.edu.tr).

 <https://orcid.org/0000-0002-9183-1568>

GÖKHAN ERDEMİR, is with Department of Electrical and Electronics Engineering, Istanbul Sabahattin Zaim University, Istanbul, Turkey, (gokhan.erdemir@izu.edu.tr).

 <https://orcid.org/0000-0003-4095-6333>

Manuscript received March 25, 2022; accepted May 25, 2022.

DOI: [10.17694/bajece.1093158](https://doi.org/10.17694/bajece.1093158)

model information of the vehicle by using deep learning methods. The working principle of such systems, which can be added in positive or negative cases, is based on certain features of the vehicles. The vehicle's model, year, color and license plate, etc., distinguishing features are used. Today, it is seen that deep learning methods are used to get fast and accurate results from such applications. There are about 20 classes in deep learning methods. But these classes are methods that are derived from six categories. If it is necessary to give an example about derivation, it can be said to create a more complex structure by increasing the number of layers and recognizing this as a new approach. These six methods are (a) Convolutional Neural Networks (CNN), (b) Recurrent Neural Network (RNN), (c) Long/Short Term Memory (LSTM), (d) Restricted Boltzmann Machines (RBM), (e) Deep Belief Networks (DBN), and (f) Denoising Autoencoders (DAE). CNN is an algorithm that allows us to analyze various objects in data. CNN, a forward-looking neural network, is a method developed by considering the visual centers of animals as output. By convolutional processes, it can be regarded as a neuron's response to stimuli from its stimulus field [2-5]. RNN, with a simple example to understand the structure of recurrent neural networks more efficiently, puts our thoughts based on the knowledge we have learned before. In other words, when we read a sentence, we make sense of each word by considering the previous word. Traditional neural networks don't work that way. So this is a considerable shortcoming. RNN, on the other hand, has a structure that deals with such problems [1, 6]. LSTM networks are different from RNN networks. It has emerged as a structure that meets the need to estimate context gaps, known as the disadvantage of RNNs. [1, 7]. RBM is a neural network that can learn probabilistic distributions of Boltzmann machine input data set, which performs classification, regression, and feature learning [1]. DBN is created by stacks of Restricted Boltzmann Machines (RBM). RBMs are realized by training and learning, respectively. But horizontally, there is no connection between the layers [8, 9]. Autoencoders are a type of ANN commonly used for feature selection and extraction.

According to the structure of networks, the hidden layer can have more layers than the input layer [9, 10]. There is a risk for an identity function called the null function. It means that output equals the input, somehow. This situation makes Autoencoder useless. DAE can solve the problem by randomly turning some of the input values to zero. In general, the ratio of the input values which turn to zero should be set between

30-50% depending on the data type [9, 10].

This study used CNN to detect counterfeit vehicle plates by considering vehicle registration information as the make, model, color, etc. A vehicle image dataset that has been generated by data augmentation was used for the training process [11]. In section II, CNN, one of the typical deep learning methods, is mentioned. This section also presents database structure, data augmentation, color, and license plate recognition steps. Section III presents the application steps of the system and its performance. The last section discusses the accuracy of the applied methods and the results we obtained.

II. MATERIALS AND METHODS

We used the Python programming language and its deep learning libraries in this study. The system's primary purpose is that a make, model, etc., information of the vehicle as output is produced based on the trained data set results on the vehicle image given as input. The flow chart of the developed system is shown in the Figure 1.

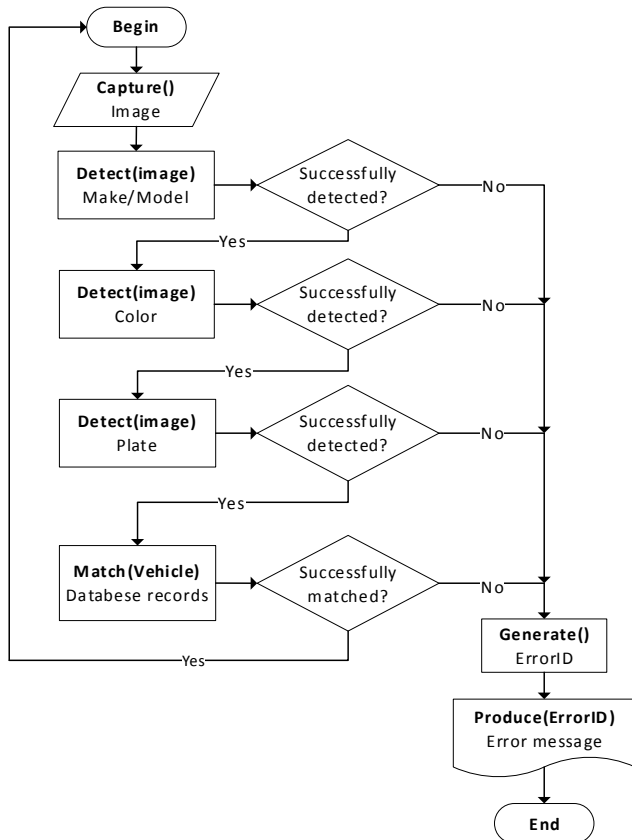


Fig.1. The flow chart of the system.

The first step of the system is to detect the vehicle's make and model on captured images from the video stream by using CNN. In this step, the dataset described in [11] was used for training. After detecting the make and the model of the vehicle, the vehicle color is recognized by using morphological operations on the image. And then, the vehicle plate number is recognized on the image by using pattern and character recognition techniques. The obtained information

about the detected vehicle such as the make, model, year, color, and plate number is compared with database records as a final step. In this study, a virtual temporary database was generated to compare detected and actual information about the vehicle. If any errors occur during each detection step, an individual error code is generated by the system. The process is terminated by a warning message which includes the error code to the user. If an unregistered or suspicious vehicle is detected, such as make and model do not match with records, the system marks the detected vehicle. It produces a warning message for the user.

A. Dataset

The dataset's content consists of 10000 images for training, 1000 for testing, and 100 images for experimental studies [11]. All images in the dataset are in RGB format with 600x450 pixel resolution [11]. Figure 2 shows some examples in the dataset [11]. The data set was created as four classes based on the front and rear views of a single model of two brands. In the dataset [11], the Honda Civic and Ford Focus models were used. This study used the same brands and models denoted as HC and FF on the system, respectively. It is possible to increase brands and model variations by following the same steps [11].

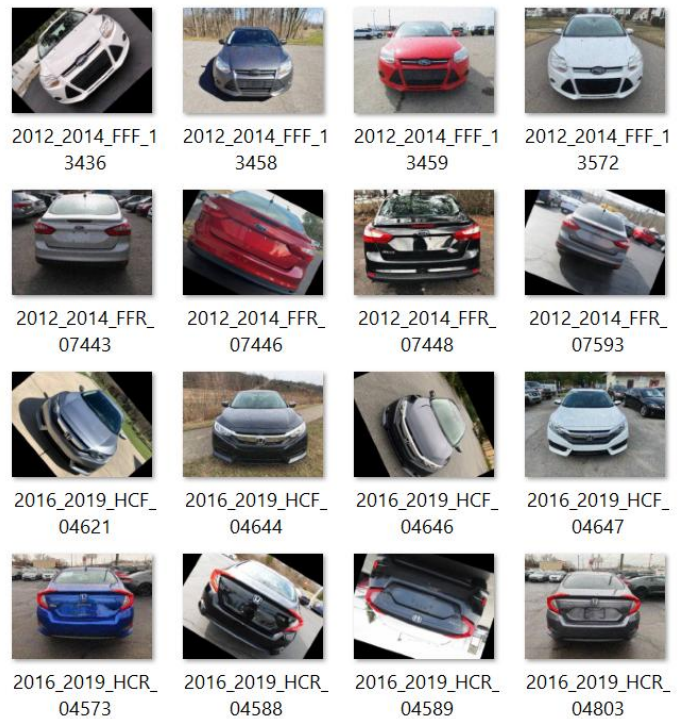


Fig.2. Some examples in the dataset

B. Data Augmentation

The essential requirement for obtaining high performance and low error results from deep learning algorithms is the use of large amounts of data in the model's training process. In case of insufficient data, generating synthetic data using data augmentation methods is one of the most common approaches.

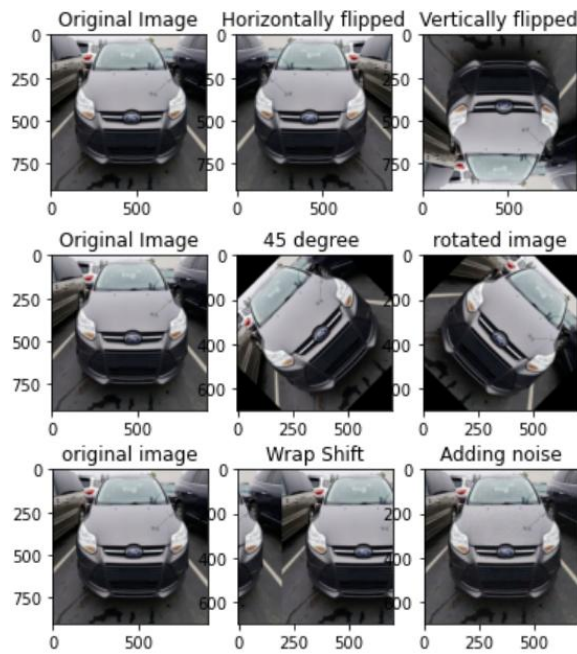


Fig.3. Samples for data augmentation

This study used data augmentation methods such as rotating, vertically shifting, horizontally symmetrical, vertically symmetrical, zooming, cutting, and adding noise to enrich the dataset [11]. Samples after data augmentation can be seen in Figure 3.

C. Convolutional Neural Networks (CNN)

In this study, we used CNN for training the model. To be explained most simply, this method is a deep learning algorithm that allows us to distinguish various objects or objects on an image from each other. Essentially, CNN consists of one or more convolutional layers, a subsampling layer, and one or more connected layers, a standard multilayer neural network. CNN solves the classification result using a typical neural network. But it also needs other layers to detect some information and features. The Convolutional Layer (CL) is the first layer of CNN. It detects low and high-level features on the image by applying some filters. In this layer, the image comes as input and is passed through a filter. It creates an attribute map due to the values formed after filtering. This stage can be seen in Figure 4 step by step. First, the filter is positioned in the upper left corner of the image. Then the indices of these two matrices are multiplied with each other, and all the results are added. After that, the filter is shifted to the right by 1 pixel, and the multiplication and addition processes are repeated. The same operations are applied to all pixels. The feature map is ready as a result of these operations. If the colored images use for classification, these processes are repeated for the RGB layers. Features are detected by moving the filter over the image and using matrix multiplication. More than one convolutional layer can be used on an image. The nonlinear layer, called the activation layer usually comes after the convolutional layers. The linearity is the main problem for this layer. While nonlinear functions such as sigmoid and then were used frequently in the past, today, the rectified linear activation function (ReLU) is commonly used that gives the

best results in training the neural network. ReLU is usually used in the hidden layer, not the output. According to the problem, sigmoid or Softmax is preferred for the output [12]. The ReLU activation function is used to determine the best results. Pooling is generally used after the sequential convolutional layers. There are no learned parameters in this layer of the network. The goal is to minimize the computational complexity by keeping the number of channels of the input matrix constant and reducing the height and width information.

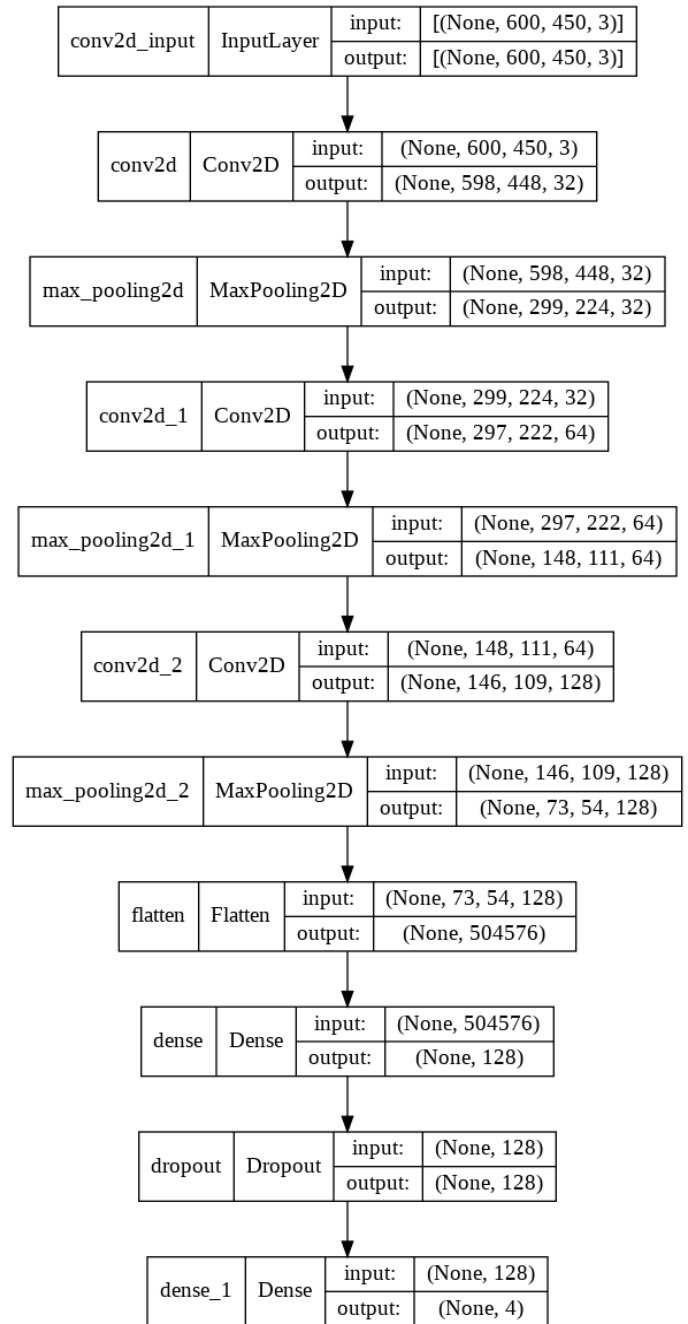


Fig.4. Designed CNN model's structure

The flattening layer is responsible for preparing data for the last layer, the fully connected layer. This layer contains the data converted into a one-dimensional array from the

convolutional and the pooling layers. The fully connected layer's inputs are matrices created by the flattening layer. It is known as the result layer. It takes the value from the flattening output of the ANN as the input layer. Then, this layer produces the output value according to the activation function of the input values and the values it receives from the hidden layers. As a result of this process, the result estimation is produced with the Softmax activation function that has been determined. Another parameter used in the algorithm is Dropout, which is used in the fully connected layer. This parameter allows some neurons to be neglected to prevent overfitting during training [13].

Activation functions are used to introduce nonlinear features in ANN. In equation (1) for a simple neural network, (x) is used for the inputs, (w) for the weights, and (b) for the deviation/trend value. Output (y) is obtained by adding the activation function (f) to the output to form the result.

$$y = f\left(\sum_i^n (w_i x_i + b)\right) \quad (1)$$

ANN without an activation function acts as a linear function. Therefore, the neural network remains with limited learning power consisting of single-order polynomials. The main goal here is to learn what is desired in non-linear situations in the artificial neural network. The main reason is that actual data such as images, video, text, and audio will be given to our neural network to learn. In this way, our neural network performs faster and more accurately learning. On the other hand, the backpropagation algorithm is used to calculate the weights. The most essential strategy is to minimize the error rate. This is a step in the form of multiplying the inputs and the weights as shown in equation (1), sum with the bias value that shifts the activation function to the right or left, and applying the activation [14, 15].

Our model sequentially 32, 64, 128 units (128 outputs) with ReLU activated layers. For this reason, the ReLU function with 128 outputs are used while creating the fully connected layer. As shown in equations (2) and (3), it takes values in the range of ReLU [0, +∞).

$$f_1(x) = \begin{cases} 0, & x < 0 \\ x, & x \geq 0 \end{cases} \quad (2)$$

$$f_2(x) = \begin{cases} 0, & x < 0 \\ 1, & x \geq 0 \end{cases} \quad (3)$$

In general, in ANN models, low computational costs and short processing time are desired via aimed that more efficient results with low processing load. The ReLU is exactly at this point where it takes the value 0 for values on the negative axis [16-18] and decreases computational costs.

Four classes exist in the proposed model. Therefore, the model must produce four output values. Adding a new layer can suffice to solve this problem. However, before adding the new layer, the dropout value is set as 0.5 to reduce the weights

by half and prevent over-learning. Then, a new layer with four outputs with a softmax activation function adds to the model.

The softmax function structure is similar to the sigmoid function. It is used in the output layer of the model when there are more than two classification problems [12]. As shown in Equations (4) and (5), it provides a probability value in the range of 0-1 to determine whether the input belongs to a specific class or not [17, 19].

$$f_i(\vec{x}) = \frac{e^{x_i}}{\sum_{j=1}^j e^{x_j}}, i = 1, 2, \dots, j \quad (4)$$

$$\frac{\partial f_i(\vec{x})}{\partial x_j} = f_i(\vec{x})(\delta_{ij} - f_i(\vec{x})) \quad (5)$$

D. Color and License Plate Recognition

Vehicle color detection is a subsystem that estimates a vehicle's color by performing morphological operations on the image [20]. For this process, RGB values of all pixels on the image cropped from certain regions are gathered, as shown in Figure 5 on the picture of the vehicle. The most common RGB value is estimated as the color of the vehicle. The accuracy in the color estimation of the vehicle is directly related to the resolution quality of the image [20].

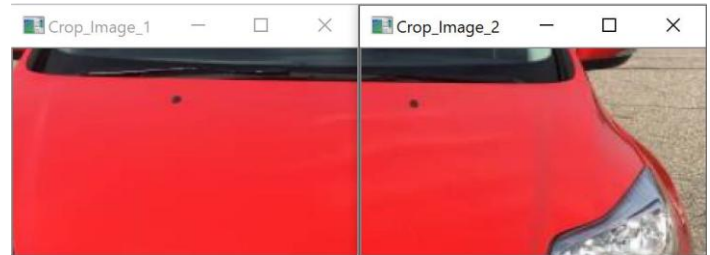


Fig.5. Cropped image to estimate vehicle's color.

License plate recognition systems are generally used for speed violations, light violations, toll highway crossings, customs, etc. The picture is taken with a camera from the front or rear. And then, the plate number area is detected by applying morphological processes [23]. The color image is converted to be entirely gray. Edge vertices on the image are determined by using an edge detection algorithm on the resulting image. The license plate number is converted from the image to the text using pattern and character recognition algorithms on the detected area [20-23].

III. CASE STUDIES

Deep learning, image processing, and mathematics libraries of the Python are used during coding. On the other hand, TensorFlow, is an open-source deep-learning library, was used for creating the deep learning model. Its flexible structure allows you to deploy calculations using one or more CPUs, GPUs, with a single API, regardless of platform. The other essential library in this study is Keras, an open-source neural network library. Numpy library, a basic math library used for scientific calculations, was also used. PIL is a graphics

processing library used to perform any operation on an image. Imutils library was used to flip, rotate, resize, contour extraction, edge detection, etc., on an image. PyTesseract, a character recognition tool that allows us to perform morphological operations, was used for the plate number recognition process. Besides, synthetic images were used to perform segmentation, geometric transformations, color space manipulation, analysis, filtering, morphology, feature detection on an image.

TABLE I
COMPARISON OF ACTIVATION FUNCTIONS

Activation functions	Training loss	Training accuracy	Test loss	Test accuracy	Epoch
sigmoid	1,3863	0,2498	1,3863	0,2500	10
tanh	1,3965	0,2498	1,3966	0,2500	10
ReLU	0,0939	0,9656	0,1360	0,9520	10
LeakyReLU	0,0673	0,9767	0,1054	0,9640	10
swish	1,3862	0,2498	1,3862	0,2500	10

While creating the proposed deep learning model, the model was compared with five different activation functions. The original dimensions of 600x450 pixels were entered in the size of the images in the database used for training and testing. It was performed on the Google Colab server, which provides approximately 12 hours of limited-time support by giving ten epochs and batch_size 32 values. In Table I, it can be seen that the ReLU and LeakyReLU activation functions gave the best results. In addition, the train and test values of the functions were quite close to each other. This is because the LeakyReLU function is derived from the ReLU function. Two activation functions were used to determine our model with an epoch value of 40. Table II and Table III show that the ReLU function is more successful with a slight value difference.

TABLE II
OPTIMIZED VALUES OF ReLU

Batch Size = 32		ReLU		
		adam	rmsprop	sgd
1 Epoch	loss	1,009	1,715	1,321
	accuracy	0.608	0.559	0.351
	val_loss	0.621	0.642	1,165
	val_accuracy	0.741	0.727	0.499
10 Epoch	loss	0.138	0.268	0.461
	accuracy	0.949	0.922	0.808
	val_loss	0.111	0.167	0.436
	val_accuracy	0.966	0.943	0.827
20 Epoch	loss	0.074	0.302	0.272
	accuracy	0.974	0.927	0.898
	val_loss	0.080	0.175	0.235
	val_accuracy	0.985	0.958	0.919
30 Epoch	loss	0.059	0.378	0.186
	accuracy	0.981	0.910	0.928
	val_loss	0.038	0.287	0.183
	val_accuracy	0.988	0.949	0.930
40 Epoch	loss	0.048	0.423	0.138
	accuracy	0.985	0.901	0.945
	val_loss	0.067	0.324	0.129
	val_accuracy	0.987	0.955	0.962
Training and Test Results	Training Loss	0.013	0.152	0.067
	Training Acc.	0.996	0.958	0.975
	Test Loss	0.067	0.324	0.129
	Test Acc.	0.987	0.955	0.962

TABLE III
OPTIMIZED VALUES OF LeakyReLU

Batch Size = 32		LeakyReLU		
		adam	rmsprop	sgd
1 Epoch	loss	1,789	1,356	1,321
	accuracy	0.476	0.326	0.351
	val_loss	0.694	12.57	1,165
	val_accuracy	0.733	0.453	0.499
10 Epoch	loss	0.320	0.398	0.461
	accuracy	0.910	0.838	0.808
	val_loss	0.295	0.378	0.436
	val_accuracy	0.931	0.853	0.827
20 Epoch	loss	0.411	0.235	0.272
	accuracy	0.907	0.907	0.898
	val_loss	0.320	0.201	0.235
	val_accuracy	0.964	0.932	0.919
30 Epoch	loss	0.459	0.159	0.186
	accuracy	0.892	0.937	0.928
	val_loss	0.377	0.152	0.183
	val_accuracy	0.957	0.941	0.930
40 Epoch	loss	1,068	0.121	0.138
	accuracy	0.882	0.953	0.945
	val_loss	0.643	0.103	0.129
	val_accuracy	0.936	0.963	0.962
Training and Test Results	Training Loss	0.206	0.051	0.067
	Training Acc.	0.936	0.983	0.975
	Test Loss	0.643	0.103	0.129
	Test Acc.	0.936	0.963	0.962

The vehicle information from which the brand/model, license plate, and color information was obtained were compared with the records in the database. As shown in Table IV, it is defined within the records that when the vehicle's make/model, license plate, and color information match, it sends a successful message to the screen. If any information does not match, it gives a failed message. While the resolution values for the images were minimum of 500x500 pixels, successful results were obtained in 14 images with an accuracy of 87.5% for 16 test images and a failure rate of 12.5% in 2 images due to the low resolution of the test. When testing images with low pixel resolution, it was observed that they failed in color detection, plate region detection, or character reading.

TABLE IV
TEST RESULTS FOR 100 RANDOMLY CHOSEN IMAGES

Make / Model	Images number	Accurate Detection	Accuracy rate
2012_2014_FFF	25	25	100%
2012_2014_FFR	25	25	100%
2016_2019_HCF	25	25	100%
2016_2019_HCR	25	25	100%

IV. CONCLUSION

According to training and test results which were shown in Table II and III, it can be seen that the developed system almost has 98.7% accuracy on the test data to detect a vehicle's brand and model. The color recognition subsystem was designed as the sum and average of the pixel numbers on the image pieces taken by giving specific coordinates on the vehicle image without being subject to learning, training, or testing. The most significant handicap for this process, a simple and straightforward method, is the reflection of light in sunny weather. It is thought that the system will work with the same performance when a larger and more complex data set is used. As a result of this study, a make/model, year, and color recognition system has been developed that can be integrated into the existing license plate recognition systems.

REFERENCES

- [1] A. Şeker, B. Diri, and H. Balık, "Derin Öğrenme Yöntemleri ve Uygulamaları Hakkında Bir İnceleme", *Gazi Journal of Engineering Sciences*, vol. 3, no. 3, 2017, pp. 47–64.
- [2] K. Fukushima, "Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position", *Biological Cybernetics*, vol. 36, no. 4, 1980, pp. 193–202.
- [3] Z. Yumeng, C. Peng, F. Liuping and C. Fangfang, "Research on Pseudo-Random Noise Information Identification Technology of Printed Anti-Counterfeiting Image Based on Deep Learning", 2020 5th International Conference on Computer and Communication Systems (ICCCS), 2020, pp. 206-209.
- [4] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-Based Learning Applied to Document Recognition", *Proc. of the IEEE*, vol. 86, no. 11, 1998, pp. 2278-2324.
- [5] Y. Lecun, L.D. Jackel, B. Boser, J.S. Denker, H.P. Graf, I. Guyon, D. Henderson, R. E. Howard and W. Hubbard, "Handwritten digit recognition: applications of neural network chips and automatic learning", *IEEE Communications Magazine*, vol. 27, no. 11, 1989, pp. 41-46.
- [6] J. L. Elman, "Finding structure in time", *Cognitive Science*, vol. 14, no. 2, 1990, pp. 179–211.
- [7] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory", *Neural Computation*, vol. 9, no. 8, 1997, pp. 1735–1780.
- [8] G. E. Hinton, "Reducing the Dimensionality of Data with Neural Networks", *International Encyclopedia of Education*, vol. 313, no. July, 2006, pp. 468–474.
- [9] L. Tang, Z. Gao and L. Huang, "Plate Recognition Based on Deep Learning", 2018 12th IEEE International Conference on Anti-counterfeiting Security and Identification (ASID), 2018, pp. 116-120.
- [10] Ş. Abdulkadir and A. G. Yüksek, "Stacked Autoencoder Method for Fabric Defect Detection", *Cumhuriyet Science Journal*, vol. 38, no. 2, 2017, pp. 342–342.
- [11] G. Erdemir and B. Ağgöl, "Data Augmentation for a Learning-Based Vehicle Make-Model and License Plate Matching System", *European Journal of Technic*, vol. 10, no. 2, 2020, pp. 331–339.
- [12] C. Bircanoğlu and N. Arica, "A comparison of activation functions in artificial neural networks", 2018 26th Signal Processing and Communications Applications Conference (SIU), 2018, pp. 1-4.
- [13] N. Srivastava and G. Hinton, A. Krizhevsky, I. Sutskever and R. Salakhutdinov, "Dropout: A Simple Way to Prevent Neural Networks from Overfitting", *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 15, 2018, pp. 7642–7651.
- [14] M. M. Lau and K. H. Lim, "Review of adaptive activation function in deep neural network", 2018 IEEE EMBS Conference on Biomedical Engineering and Sciences, IECBES 2018 - Proceedings, 2019, pp. 686–690.
- [15] H. H. Tan and K. H. Lim, "Vanishing Gradient Mitigation with Deep Learning Neural Network Optimization", 2019 7th International Conference on Smart Computing and Communications (ICSCC), 2019, pp. 7–10.

- [16] M. Kaloiev and G. Krastev, "Comparative Analysis of Activation Functions Used in the Hidden Layers of Deep Neural Networks", 2021 3rd International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA), 2021, pp. 1-5.
- [17] O. Sharma, "A New Activation Function for Deep Neural Network", *Proceedings of the International Conference on Machine Learning, Big Data, Cloud and Parallel Computing: Trends, Perspectives and Prospects, COMITCon*, 2019, pp. 84–86.
- [18] H. Chung, S. J. Lee, and J. G. Park, "Deep neural network using trainable activation functions", *Proceedings of the International Joint Conference on Neural Networks*, vol. 2016-October, no. 1, pp. 348–352.
- [19] M. A. Mercioni and S. Holban, "The Most Used Activation Functions: Classic Versus Current", 2020 International Conference on Development and Application Systems (DAS), 2020, pp. 141-145.
- [20] B. Ağgöl and G. Erdemir, "Açık Kaynak Kodlu Taşıt Renk Tespit Yazılımı Geliştirilmesi", *İstanbul Sabahattin Zaim Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, vol. 3, no. 1, Mar. 2021, pp. 47-50.
- [21] O. Kaplan, Ş. Sağiroğlu, Ö.F. Çolakoğlu, "Erciyes Üniversitesi Bilgisayar Mühendisliği Bölümü Araç Tanıma Sistemi", 2002, pp. 2–6
- [22] Ş. Sağiroğlu and E. Beşdok, "A Novel Approach for Image Denoising Based on Artificial Neural Networks", vol. 15, no. 2, 2012, pp. 71–86.
- [23] O. Bingöl and Ö. Kuşçu, "Bilgisayar Tabanlı Araç Plaka Tanıma Sistemi", *Bilişim Teknolojileri Dergisi*, vol. 1, no. 3, 2008, p. 1-5.

BIOGRAPHIES



BURAK AĞGÖL received his undergraduate degree from Istanbul Sabahattin Zaim University, Faculty of Engineering and Natural Sciences, Department of Computer Engineering, in 2015. He graduated from Istanbul Sabahattin Zaim University Graduate Education Institute, Department of Computer Science and Engineering in 2020. Now, he is a Ph.D. student in the same department. He works as a lecturer in Istanbul Ayyansaray University Plato Vocational School Computer Technologies Department. His areas of interest are artificial intelligence, deep learning, machine learning, and image processing.



GÖKHAN ERDEMİR received his B.Sc., M.Sc. and Ph.D. degrees from Marmara University, Turkey, respectively. During his Ph.D., he worked as a research scholar at Michigan State University, Department of Electrical and Computer in East Lansing MI, USA. Now, he is an assistant professor at Istanbul Sabahattin Zaim University, Department of Electrical and Electronics Engineering. His research topics include robotics, mobile robotics, and control systems.

Drowsiness Detection System Based on Machine Learning Using Eye State

Emine Merve Öztürk, Ayhan Küçükmanisa and Oğuzhan Urhan


Abstract—Drowsiness is one of the major causes of driver-induced traffic accidents. The interactive systems developed to reduce road accidents by alerting drivers is called as Advanced Driver Assistance Systems (ADAS). The most important ADAS are Lane Departure Warning System, Front Collision Warning System and Driver Drowsiness Systems. In this study, an ADAS system based on eye state detection is presented to detect driver drowsiness. First, Viola-Jones algorithm approach is used to detect the face and eye areas in the proposed method. The detected eye region is classified as closed or open by making use of a machine learning method. Finally, the eye conditions are analyzed at time domain with PERcentage of eyelid CLOSure (PERCLOS) metric and drowsiness conditions are determined by Support Vector Machine (SVM), kNN and decision tree classifiers. The proposed methods tested on 7 real people and drowsiness states are detected at 99.77%, 94.35%, and 96.62% accuracy, respectively.

Index Terms— Driver Drowsiness, ADAS, Viola-Jones, PERCLOS, Machine learning.


I. INTRODUCTION

DRIVERS' DROWSINESS is one of the major causes of road traffic accidents. According to many surveys 25-30% of road accidents are caused by drowsiness of the driver, and as a result, many lives are lost, many properties are damaged, and these numbers increase each day [1]. Drowsiness (also means sleepiness) is a situation where a person feels a need to fall asleep and is a situation of sleep-wake cycle [2].


EMİNE MERVE ÖZTÜRK, is with Department of Electronics and Communication Engineering of University of Kocaeli, Kocaeli, Turkey, (e-mail: e.merve.ozturkk@gmail.com).

 <https://orcid.org/0000-0002-1541-7031>

AYHAN KÜÇÜKMANİSA, is with Department of Electronics and Communication Engineering of University of Kocaeli, Kocaeli, Turkey, (e-mail: ayhan.kucukmanisa@kocaeli.edu.tr).

 <https://orcid.org/0000-0002-1886-1250>

OĞUZHAN URHAN, is with Department of Electronics and Communication Engineering of University of Kocaeli, Kocaeli, Turkey, (e-mail: urhano@kocaeli.edu.tr).

 <https://orcid.org/0000-0002-0352-1560>

Manuscript received November 25, 2021; accepted June 13, 2022.

DOI: [10.17694/bajece.1028110](https://doi.org/10.17694/bajece.1028110)

Recent survey by the National Highway Traffic Safety Administration (NHTSA) estimates that 56,000 of road accidents are caused by drowsy drivers annually in the U.S.A. resulting in 40,000 injuries and 1,550 fatalities [3]. It requires considerable amount of effort to develop an effective system that can detect drowsiness and take appropriate measures against road accidents. Some progress has been made in the construction of smart vehicles to prevent such accidents [4]. With increasing interest in intelligent vehicles, the development of robust and practical fatigue and drowsiness detection systems has gained utmost importance.

ADAS is part of the active safety systems that is designed to alert the drivers to help them avoid traffic accidents. The main objective is to contribute the reduction of traffic accidents by using newly developed technologies; that is, incorporating new systems for increasing vehicle security, and at the same time, decreasing the dangerous situations that may arise during driving due to human errors [5]. Many surveys show that ADAS can prevent up to 40% of road accidents depending on the ADAS type and the type of accident scenario [6].

Techniques used to detect driver drowsiness can generally be divided into three main categories. The first category includes methods based on evaluation of biomedical signals such as brain, muscle, and cardiovascular activity. Generally, these methods require electrodes that are attached to the body of the driver, which is mostly considered uncomfortable to the driver. The method belonging to the second category mainly evaluates driver performance by observing changes in vehicle side position, speed, steering wheel, and other Control Area Network (CAN) bus signals. The advantage of these approaches is that the signal is meaningful, and the acquisition of the signal is very easy. The third category approaches address the problem of finding drowsy drivers using computer vision techniques applied to the human face [7-9]. This category includes methods based on driver visual analysis using image processing techniques. These approaches are effective because drowsiness, driver's facial appearance and head/eye activity are taken into consideration in this case.

In this paper, Viola-Jones [10] algorithm is used for eye pair and face detection. The next step is an advanced and efficient approach of calculating PERcentage of eyelid CLOSure (PERCLOS) of the driver. PERCLOS means the proportion of time that the subject's eyes are closed over a specific period of time [11].

The second section of this paper compares some works related to this study. In the third section, methods of face detection, eye pair detection and drowsiness detection of the driver is presented. Experimental results are presented in the fourth section. The last section concludes the paper.

II. RELATED WORK

Various studies have been done on driver drowsiness detection based on eye state. The performance of these studies varies based on methods used for detection of the eye pair.

In [12], iris-sclera pattern analysis method is used to detect the open eye. In the next stage, PERcentage of eyelid CLOSure (PERCLOS) metrics is utilized to determine the drowsiness state of the driver. The entire system is designed to be independent of any specific data sets for face or eye detection. The proposed system is evaluated using real-life images and videos. At the end of the experiments, an open-eye detection accuracy of 93% is achieved.

In another paper [13], eye tracking based approach for drowsiness detection is used. In this work, detected eye regions from Viola-Jones algorithm feed to PERCLOS method for successful detection of drowsiness of a driver. The method achieves an accuracy of 95%.

A binary SVM classifier with a linear kernel is used as classifier in [14]. Drowsiness detection is performed under different lighting conditions, and it is shown that this approach performs well in the challenging lighting conditions. This system achieves an overall accuracy of 94.58% in four test cases which is an impressive result.

In [15], a drowsiness detection method based on face and eye detection within a video input is proposed. PERCLOS is calculated based on the area of iris region. The experimental results suggest that this method can potentially detect drowsiness based on PERCLOS as it is found that when the driver is drowsy the PERCLOS value is lower (compared to PERCLOS value when the driver is alert). Better results are obtained when both iris regions (from left and right eyes) are used to measure PERCLOS values.

The proposed method in [16] benefits from face landmarks to estimate the user’s eye aspect ratio, subsequently applying an optimized-SVM to classify the state feature. Then, a decision rule is adopted to determine whether the driver is drowsy or not.

III. PROPOSED METHOD

The proposed system is discussed in this section, where the driver drowsiness is detected. The flowchart of the proposed system is given in Figure 1.

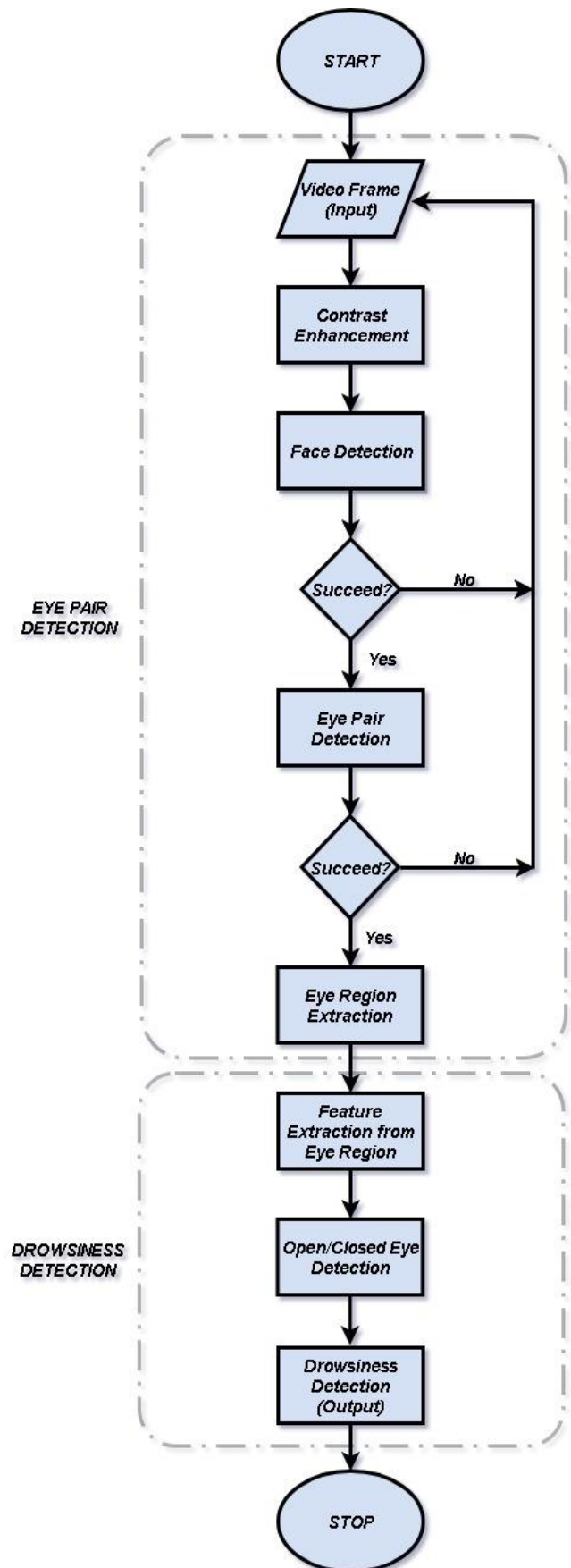


Fig.1. Flowchart of the proposed system

A. Pre-processing

In this paper the input videos are captured from 7 real people. The original-colored frames are shown in Figure 2. The lighting condition varies due to the outdoor environment. That is why the frames within the video needs to be pre-processed in order to enhance the images. Histogram stretching is applied to improve the contrast of low contrast images. Sample contrast-enhanced frames are shown Figure 3.

B. Face and Eye Detection

Viola-Jones algorithm is used for both face and eye pair detection since its efficiency in detection performance and relatively low computational load. Viola-Jones algorithm requires full view frontal faces to be able to operate. There are some operations which enable it to run in real-time that are integral image, Haar-like features [10], AdaBoost [17] and the cascade classifier. In Fig. 4, detected faces in the frames are shown and then in Figure 5 the detected eye region can be seen.

The Viola-Jones face detection method uses combinations of simple Haar-Like features to classify faces. Haar-like features are rectangular digital image features that get their name from their similarity to Haar wavelets. The value of a two-rectangle feature is the difference between the sum of the pixels within two rectangular regions. The regions have the same size and shape. A three-rectangle feature computes the sum within two outside rectangles subtracted from the sum in a center rectangle. Finally, a four-rectangle feature computes the difference between diagonal pairs of rectangles.

Integral image computation is a step of Viola-Jones method where input face image is converted into an integral image. This integral image is used for quick feature detection. The meaning of integral image is the outline of the pixel values in the original images. The integral image at location (x, y) contains the sum of the pixels above and to the left of (x, y) inclusive. The integral image computation can be performed in the equation given below:

$$ii(x, y) = \sum_{x' < x, y' < y} i(x', y') \quad (1)$$

where each location of x and y in the integral image is the sum of pixel values in above and left location of x and y .

AdaBoost is a machine learning boosting method capable of finding a highly accurate hypothesis by combining many weak hypotheses each with average accuracy. The AdaBoost method is generally viewed as the first step straight into more practical boosting methods helps to select small features from the face that facilitates fast and easy computation. Unlike other methods, AdaBoost algorithm gives desired region of the object by discarding unnecessary background.

The Viola-Jones algorithm eliminates candidates quickly using a cascade of stages of cascade classifier. The cascade

eliminates candidates by making stricter requirements in each stage with later stages being much more difficult for a candidate to pass. Candidates exit the cascade if they pass all stages or fail any stage. A face is detected if a candidate passes all stages. The cascade classifier consists of levels each containing a weak classifier. The responsibility of each level is to evaluate if a given sub-window is actually non-face or maybe a face. Typically, early levels are passed more frequently with later levels being more demanding.

Feature extraction is the process of reducing the size of data by identifying the distinguishing features of the problem. Feature can be shape, color, pattern, reflection, edge properties, etc. The performance will also increase as the selected features can be effectively separated from the others. In this study, Histogram Oriented Gradient (HOG) is used for feature extraction. The method is to obtain feature data by dividing a filter placed on the frames into overlapping blocks and summing the amplitudes of the pixels within each block in an orientation histogram and all the histogram entries are used as the feature vector to describe the object.

C. Eye State Detection

The blinking of the eyes is detected using Support Vector Machines (SVM) in the proposed method. SVM classifier is one of the most powerful classification techniques used for eye state detection. By distinguishing the features of closed and open eye pairs from each other in the most appropriate way, the detection of eye state is provided. In Figure 6, some detected eye states in the example frames are shown.

D. Driver Drowsiness Detection

Driver drowsiness detection is decided using the PERCLOS metric on three levels as shown in Table 1. The SVM, kNN, decision tree classifiers are used for decision of the drowsiness state.

PERCLOS is the rate of eye closure calculated at specific time intervals. That is, the ratio of the number of closed eyes in the number of frames over the selected period to the total number of frames in the period. PERCLOS (2) is calculated as: N_t is the number of frames in a period; N_a is the number of frames in which the eye is open.

$$\text{PERCLOS} = \frac{N_t - N_a}{N_t} \times 100 \% \quad (2)$$

Finally, SVM, kNN, decision trees and machine learning methods are used to determine driver drowsiness.

TABLE I
DRIVER DROWSINESS DETECTION

PERCLOS	Levels	Explanation
0.0-0.15	No warning	Awake and vigilant
0.15-0.30	Warning	Distracted
0.30 and above	Danger	Drowsy



Fig. 2. Sample colored frames from each video



Fig. 3. Sample contrast-enhanced grayscale frames from each video

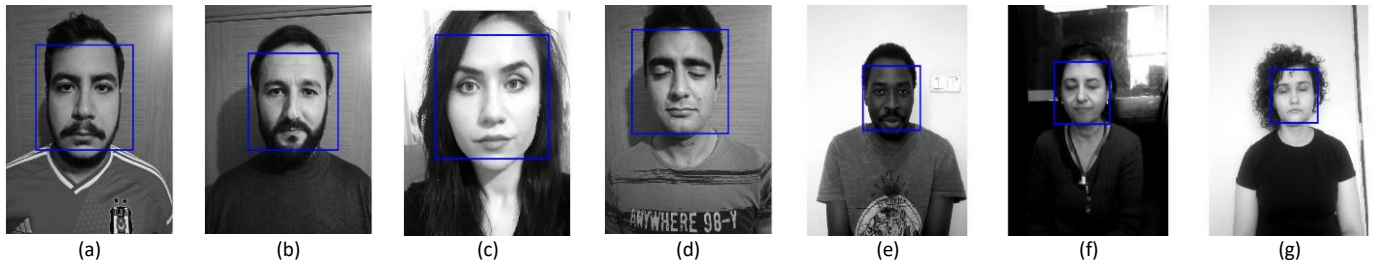


Fig. 4. Detected faces from each video

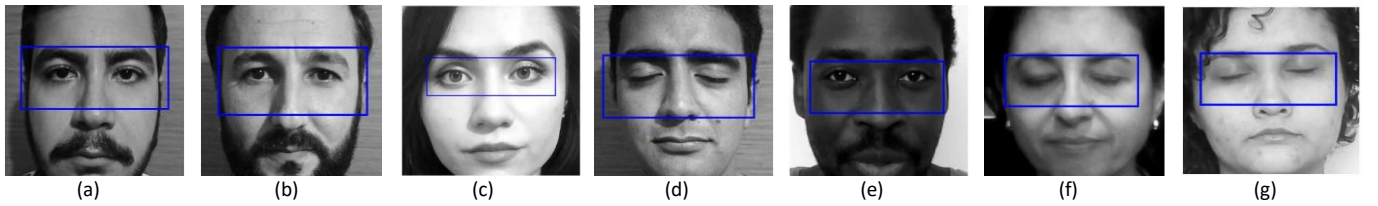


Fig. 5. Detected eyes from each video

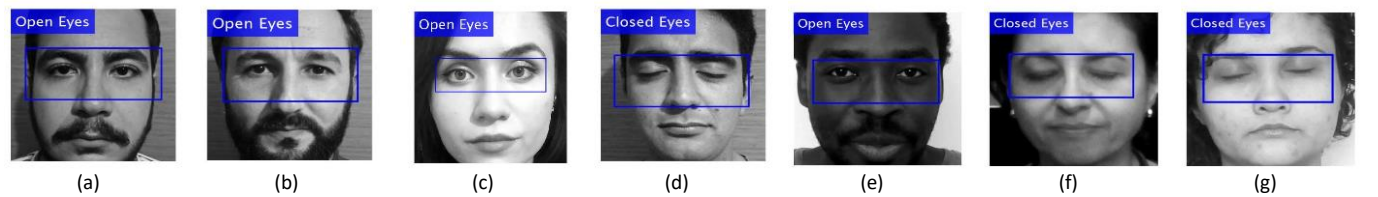


Fig. 6. Eye state detection from each video sample

IV. EXPERIMENTAL RESULTS

The dataset of the study consists of 7 videos with a total of 19,156 frames. The proposed approach is executed on an Intel Core i5 processor with 2.50 GHz processor and 6 GB RAM memory with 1.5fps processing speed in MATLAB implementation.

The performance of the experiment is assessed with accuracy, precision and recall criteria. Related equations for these criteria are given in (3), (4) and (5). In these equations, TN is the number of correct estimates for which a sample is negative; FP, the number of false estimates that a sample is

positive; FN indicates that an estimate is negative, and TP is the number of accurate estimates that a case is positive [18].

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

$$\text{Precision} = \frac{TP}{TP + FN} \tag{4}$$

$$\text{Recall} = \frac{TP}{TP + FP} \tag{5}$$

For the proposed method, the equations in (3), (4) and (5) are computed by frame and video-based performances. Frame-based performance is the performance of eye condition detection and is shown in Table 3. In Table 3, closed eyes are represented by 0 and open eyes by 1. Video-based performance is the result of driver fatigue detection. Video-based performance results are the results obtained by classifying the PERCLOS values calculated from 10 to 100 window sizes (ws) in each video. ws is also N_t in (2). Video based average performance of SVM, kNN, decision tree classifiers is given in Table 4.

Figure 7 shows the eye state examination of Figure 6(e). The vertical axis represents the closed and open eyes detection results for each frame in the horizontal axis with the numbers 0 and 1, respectively. With the value ws is 100, the PERCLOS value is calculated as 0.15 in the time period shown in the graph. Thus, it is concluded that the driver is distracted.

The drowsiness results according to Table 1, classified by SVM, kNN, decision tree classifiers respectively, are shown in Figure 8 for Figure 6 (e). The eye state classification results 0, 1, 2 in the graphs represent that the driver according to Table 1 is awake and vigilant, distracted, drowsy, respectively.

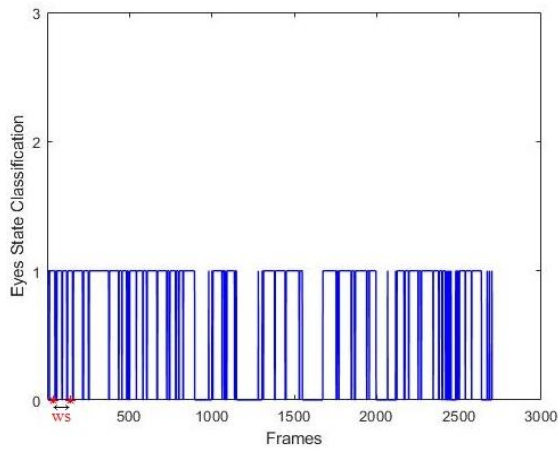


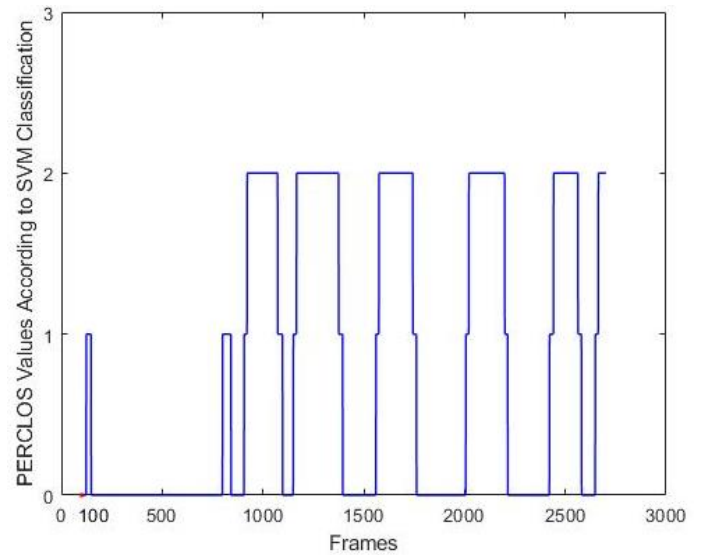
Fig. 7. Temporal representation of the eye state from Figure 6(e)

TABLE II
FRAME BASED PERFORMANCE EVALUATION

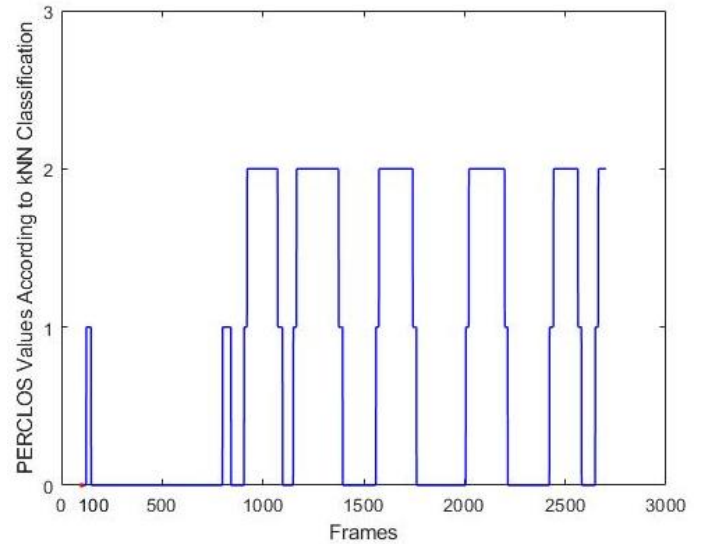
	V1	V2	V3	V4	V5	V6	V7	All
ACC	0.98	0.91	0.97	0.89	0.98	0.92	0.99	0.98
PRE 0	0.98	0.92	0.97	0.91	0.99	0.88	0.99	0.98
PRE 1	0.98	0.91	0.96	0.89	0.97	0.94	0.99	0.98
REC 0	0.93	0.79	0.95	0.74	0.94	0.82	0.99	0.92
REC 1	0.99	0.97	0.98	0.96	0.99	0.96	0.99	0.99

TABLE III
VIDEO BASED PERFORMANCE EVALUATION

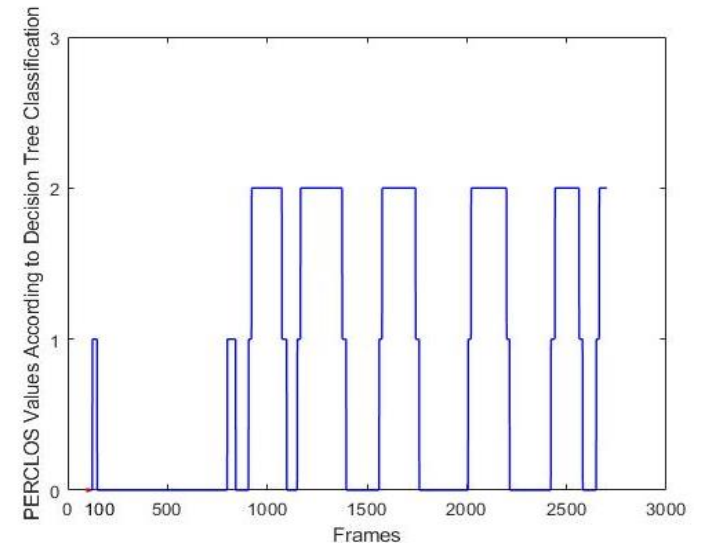
Window Size	SVM	kNN	Decision Tree
10	1.000	0.976	0.973
20	1.000	0.945	0.959
30	0.997	0.914	0.950
40	0.998	0.912	0.955
50	0.998	0.912	0.944
60	0.994	0.903	0.949
70	0.996	0.914	0.948
80	0.995	0.928	0.951
90	0.997	0.939	0.955
100	0.998	0.944	0.959



(a)



(b)



(c)

Fig. 8. Temporal representation of PERCLOS using classifiers from Figure 6(e) (a) SVM (b) kNN (c) Decision Tree

V. CONCLUSION

In this study, driver drowsiness is determined using the physical changes of the driver. Forming data set is the most important step for the proposed method. In order to make the algorithm more robust, a wider dataset has been created. A total of 7 different people's videos are used to form the dataset which includes 18.125 frames in total. Then Viola Jones algorithm is used to detect the face and eye pairs. The drowsiness of the driver is determined by taking into account the closed and open eyes detected by the PERCLOS approach. In addition, SVM, kNN and decision tree classification methods are used to for eye state and driver drowsiness detection. In this study, it has been observed that illumination intensity changes, head movements, head rotation, iris movements and body shakes can affect the detection of eye condition.

In the future, individuals wearing eyeglasses will be included in the study and dark environment images will be used by using NIR camera. The study will also be strengthened with hybrid methods and finally the output will be implemented on an embedded system.

ACKNOWLEDGMENT

This work is supported by Kocaeli University Scientific Research Projects Coordination Unit under Grant No. 2019/021.

REFERENCES

- [1] V. Vibin, S. Amritha, K. Sreeram and K. P. Remya. "Ear based driver drowsiness detection system", IOSR Journal of Engineering, 2018.
- [2] J. A. Ojo, L. T. Omilude, and I. A. Adeyemo. "Fatigue detection in drivers using eye-blink and yawning analysis", International Journal of Computer Trends and Technology, vol. 50, no 2, 2017.
- [3] S. Sooksatra, T. Kondo, P. Bunnun and A. Yoshitaka, 2018, "A drowsiness detection method based on displacement and gradient vectors", Songklanakarin J. Sci. Tech. vol. 40 no. 3, 2018, pp. 602-608.
- [4] C. In-Ho and K. Yong-Guk, "Head pose and gaze direction tracking for detecting a drowsy driver", Appl. Math. Inf. Sci. vol. 9, No. 2L, 2015, pp. 505-512.
- [5] M. J. Flores and J. M. Armingol, "Real-time warning for driver drowsiness detection using visual information", Journal of Intelligent and Robotic Systems vol. 59, no. 2, 2010, pp:103-125.
- [6] O. Gietelink and J. Ploeg, "Development of advanced driver assistance systems with vehicle hardware-in-the-loop simulations", Vehicle System Dynamics, vol. 44, no. 7, 2006, pp. 569-590.
- [7] S. S. Nagargoje and D. S. Shilvant, "Drowsiness detection system for car assisted driver using image processing", International Journal of Electrical and Electronics Research, Vol. 3, no. 4, 2015, pp: 175-179.
- [8] T. K. Dange and T. S. Yengatiwar, "A review method on drowsiness detection system", International Journal of Engineering Research & Technology, vol. 2, issue 1, 2013.
- [9] I. G. Daza, N. Hernandez, L. M. Bergasa, I. Parra, I., J. J. Yebes, M. Gavilan, R. Quintero, D. F. Llorca, M. A. Sotelo, "Drowsiness monitoring based on driver and driving data fusion", 14th International IEEE Conference on Intelligent Transportation Systems, Washington, DC, USA, 2011.
- [10] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of Simple features", Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001.
- [11] D. Sanka, F. Dileepa, J. Sadari, W. Sandareka, and D. Chathura, Efficient PERCLOS and gaze measurement methodologies to estimate driver attention in real time, 2014 Fifth International Conference on Intelligent Systems, Modelling and Simulation, 2014.
- [12] A. D. Panicker and M. S. Nair, "Open-eye detection using iris-sclera pattern analysis for driver drowsiness detection", Sadhana, vol. 42, no. 11, 2017, pp. 1835-1849.

- [13] T. P. Nguyen, M. T. Chew and S. N. Demidenko, "Eye tracking system to detect driver drowsiness", 6th International Conference on Automation, Robotics and Applications, Queenstown, New Zealand, 2015.
- [14] B. N. Manu, "Facial features monitoring for real time drowsiness detection", 12th International Conference on Innovations in Information Technology (IIT), 2016.
- [15] S. Junaedi and H. Akbar, H. "Driver drowsiness detection based on face feature and PERCLOS", International Conference on Computation in Science and Engineering, 2018.
- [16] B. S. M. Caio, C. M. Márcio, M. M. de S. João, M. do N. Lucas, B. M. July, D. L. Isis and L. D. Enrique, "Real-time SVM classification for drowsiness detection using eye aspect ratio", Probabilistic Safety Assessment and Management PSAM 14, 2018.
- [17] Y. Freund and R. E. Schapire, "A Short Introduction to Boosting", Journal of Japanese Society for Artificial Intelligence, vol. 14, no. 5, 771-780, 1999.
- [18] V. Sofia, R. Brian, R. Anca, van der K. Esther, "Confusion matrix-based feature selection, Proceedings of the 22nd Midwest Artificial Intelligence and Cognitive Science Conference", Cincinnati, Ohio, USA, 2011.

BIOGRAPHIES

EMİNE MERVE OZTURK received B.S. (2020) degree in electronics and communication engineering from the University of Kocaeli. Her interests include image and video processing.



AYHAN KÜÇÜKMANİSA received the B.Sc., M.Sc., and Ph.D. degrees in electronics and telecommunication engineering from the University of Kocaeli, Izmit, Turkey, in 2010, 2012, and 2018, respectively. Since 2013, he has been with the Department of Electronics and Telecommunications Engineering, University of Kocaeli, where he is currently Research Assistant. His research interests include digital signal, image/video processing, machine learning, deep learning, embedded systems and intelligent transport systems.



OĞUZHAN URHAN (S'02-M'06-SM'19) received his B.Sc., M.Sc., and Ph.D. degrees in Electronics and Telecommunication engineering from the University of Kocaeli, Kocaeli, Turkey, in 2001, 2003, and 2006, respectively. Since 2001, he has been with the Department of Electronics and Telecommunications Engineering, University of Kocaeli, Turkey, where he has been full professor since 2015. He was a visiting professor at Chung Ang University, South Korea, from 2006 to 2007. He is the director of Kocaeli University Laboratory of Embedded and Vision Systems (KULE). His research interests include digital image/video processing, embedded system and machine learning applications.



Proactive Metering of Mobile Internet User Experience

Mete Uzun and Osman Abul


Abstract—Having 67% worldwide share, mobile internet is very important for Internet Service Providers (ISPs). Since mobile Internet access is a collective service, Key Performance Indicators (KPIs) measuring quality of data traffic on select network segments/servers may not correctly indicate true user experience. For this reason, mobile ISPs are investing in sophisticated high-end commercial speed analysis systems which typically collect and analyze network traffic data from key network segments/servers. Unfortunately, their utility is quite limited as long as the proactive network intervention is considered. In this work, we develop a MapReduce based network speed analysis system which measures end-to-end network speed to quantify true user experience across multiple geographic regions and service categories. Also functioning as an online decision support system, it enables network administrators with timely ISP network intervention right before potential arrival of mass number of user complaints. The system has been tested with a leading mobile ISP in Turkey. The results confirm its effectiveness.

Index Terms— Big data, Hadoop, Internet service providers, MapReduce, Mobile internet, Network speed analysis, User experience.


I. INTRODUCTION

PACKET-SWITCHING mobile networks are increasingly carrying Internet traffic for multimedia, social media, web browsing, games, VoIP, IPTV and many other applications. As a result, the mobile Internet traffic is growing very fast. To this end, monthly global mobile data traffic is estimated to surpass 77 Exabyte with a share of 20% of total Internet traffic by the end of 2022, according to the Visual Networking Index of the Global Mobile Data Traffic Forecast by Cisco [1]. This is indeed why the versatile data traffic analysis and mobile network management with optimizations

METE UZUN, is with Department of Computer Engineering TOBB University of Economics and Technology, Ankara, Turkey, (e-mail: meteuzn@gmail.com).

 <https://orcid.org/0000-0002-6329-2835>

OSMAN ABUL, is with Department of Computer Engineering TOBB University of Economics and Technology, Ankara, Turkey, (e-mail: osmanabul@etu.edu.tr).

 <https://orcid.org/0000-0002-9284-6112>

Manuscript received March 17, 2022; accepted June 22, 2022.

DOI: [10.17694/bajece.1089321](https://doi.org/10.17694/bajece.1089321)

are crucial business issues for mobile Internet Service Providers (ISPs) [2].

Internet speed is very important for mobile users. ISPs employ several Key Performance Indicators (KPIs) measuring various aspects of data traffic flow [3]. KPIs are typically measured on specific core network nodes/segments by probing/tapping approach [4]. Unfortunately, due to multiple hops with distinct nodes/segment features, measuring network performance on a specific probe gives little information as far as the users' overall network speed experience is concerned. Clearly, KPIs depending on the tapping approach may fail in correctly characterizing the network quality and the true user experience. This suggests that for a proactive solution we need to measure the network quality and the true user experience with a holistic end-to-end analysis [5].

We live in the big data era in which many advanced technology systems produce high volume data. The big data harbors big opportunities to improve Internet services, to meet user needs and to create business value. Performance measurements on very large number of connection streams uncover unknown correlations of network problems, network capacity on busy hours, user experiences of QoS, hidden patterns of usage, customer preferences, market trends and other useful business and network information [14]. The analytical findings from big data can lead to more effective network operations, improved QoS efficiency, prescribed competitive intelligence, new revenue opportunities, better customer service, and other benefits.

Characterization of the user experience and detection of network problems before customer complaints are very important feedbacks for ISPs and the telecommunication sector. Network capacity on busy hours, real user experience, Quality of Service (QoS) and correlations of network problems can be gathered from analyzing connection streams and performance indicators of network centric measurement on the content servers. Another example is that, customer Internet usage trend is a very fundamental information for determining sales strategy. It is very important that the market share, the market value and the market intelligence can be easily estimated for businesses striving to increase operational efficiency and gain competitive advantage. Moreover, competitive intelligence and competitor analysis can be compiled by analyzing patterns of customer usage and mobile network data segmentation on the content servers. In this way, we can translate network capabilities to user experience, QoS and competitive intelligence using big data analysis on the Content Delivery Network [6].

Our system, called Proactive ISP Speed Meter, is able to process large speed test results in online parallel/distributed manner with MapReduce. Then using the big data analytics, the network speed maps and the service usage rates are used to characterize the true user experience: (i) for each geographic region served, and (ii) for each content delivery service category, separately. Using the relevant extracted information from end-to-end roundtrip speed test results, we can proactively translate them into Quality of Experience (QoE) per geographic region/service category. We show that our system can proactively detect potential massive user complaints before they emerge. Moreover, it provides mobile ISPs with competitive intelligence that may be exploited as a part of the decision support in future infrastructure investments.

The paper is organized as follows. We present the background and related work in the next section. Section III introduces our methodology and two main deployments (Geographic Region Based and Content Delivery Service Based) of Proactive ISP Speed Meter. An experimental evaluation of the system on live mobile network of a leading ISP provider of Turkey is provided in Section IV. Finally, Section V concludes. A preliminary version of our work, which covers only the Geographic Region Based Deployment, appears in [5].

II. BACKGROUND AND RELATED WORK

Throughput, Round Trip Time (RTT) and Latency are the main metrics that define the performance of data transmission speed. Throughput is defined as the amount of successfully transferred data in a unit time. We consider downlink and uplink speed separately. RTT measures the elapsed time between a request and its response. The smaller RTT, the better broadband speed the clients will experience [8]. Latency is the time delay between the release and the arrival of data transmission. Latency especially degrades the user experience on cloud-based applications due to much longer response times [9]. Big latency causes big RTTs. RTT, throughput and latency are dependable KPIs for user experience metering of network speed [10]. Hence, our system mainly depends on those KPIs for user experience analysis.

We can roughly categorize the Internet experience KPIs into three groups: (i) Network centric, (ii) User centric and (iii) Competitive intelligence centric (Table 1) [26]. Network centric KPIs include throughput, RTT and latency and serve as the basic QoS metrics. User centric KPIs, including page load time and TCP connection time, are mainly functions of network centric KPIs and serve as QoE. Competitive intelligence centric KPIs like number of visits/downloads, on the other hand, are useful to identify highly demanded services. Hence, they provide mobile ISPs with strategic decision making in future infrastructure investments.

Some highly consulted KPIs, which are commonly employed by most network quality evaluation systems, may not correctly surface speed bottlenecks or true user experience. Being a reliable data delivery protocol, Transmission Control Protocol (TCP) can detect errors in packets and recover from

damaged, lost, duplicated or out-of-order delivered packets [7]. Therefore, the errors will affect Data Traffic Availability and Bad Session KPIs, but may not affect customer perception of network speed. As a result, very high service up percentages reported by Data Traffic Availability and Bad Session KPIs, measured at the nodes of packet core network (Gateway GPRS Support Node-GGSN, Packet Data Network Gateway-PGW) or at data optimization servers, does not necessarily mean that high speed Internet service is provided to mobile users. For such reasons, our system avoids the use of such KPIs.

TABLE I
KEY PERFORMANCE INDICATORS

KPI	Measurement
Throughput	Network Centric
Latency	Network Centric
RTT	Network Centric
Dropped Frames	Network Centric
Page Load Time	User Centric
TCP Connection Time	User Centric
Visits, Plays, Viewers Rate	Competitive Intelligence Centric
Content Category Rate	Competitive Intelligence Centric

There are web services and standalone tools that provide analysis of Internet access/speed metrics for individual users [11]. For instance, Speedtest [12] is one of the most widely used bandwidth testing analysis tool. However, such tools are user-centric and cannot automatically characterize service quality of the ISP network. As a result, such tools are not appropriate for assisting in large-scale bottleneck detection and network management, as a part of decision support. Our big data analytics based distributed system, on the other hand, automatically measures high volume of throughput, RTT and latency KPIs periodically and proactively characterizes network speed quality of the whole network.

Because of the explosive growth of Internet connections, Content Delivery Network (CDN) produce gigantic number of server-side connection streams [13]. Internet content delivery servers are harnessing wealth and volume of information of the Internet activities, which may include web server logs, click stream data, social media, online shopping, video watching, network activities, mobile-phone call details and sensors connected to Internet of Things. After all of the Internet activities of users, highly accessed data are accumulated on the content servers for fast re-accesses. Our system measures competitive intelligence centric KPIs per service category to identify highly accessed content servers.

Utilizing a highly effective and efficient MapReduce programming model [15, 16], Hadoop [17] is typically used to process big data [14]. Hadoop provides opportunities for big data analytics [18] for mobile network service providers [19].

III. METHODOLOGY

We developed a network speed analysis system, called Proactive ISP Speed Meter, for ISP wide Internet speed performance analysis. Our focus is on mobile ISPs which typically have radio access and core networks for providing Internet access to mobile users (Fig. 1).

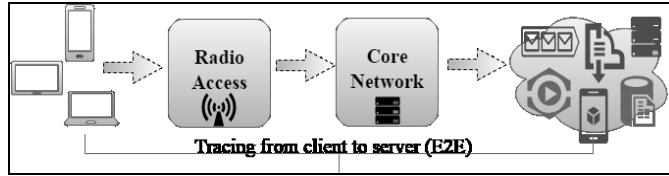


Fig. 1. Typical Mobile ISP infrastructure for Internet connection. Mobile devices access to ISP servers through radio access and core network facilities.

Proactive ISP Speed Meter has two deployments: (i) geographic region based, and (ii) content delivery service based. The former deployment (Fig. 2) periodically samples the Internet access speed from each geographic region while the latter deployment (Fig. 4) analyzes the Internet access speed per content delivery service category. This way the ISPs are informed with how to gauge user experience per geographic region and per content delivery service category, and also to be able to react differently for various occasions of bottlenecks.

Proactive ISP Speed Meter’s end-to-end speed analysis involves speed tests on the communication channel between mobile devices and servers on the network through the radio access and the core network interfaces [20]. Hence, the analysis results correctly indicate quality of network between clients and the server as the analysis does not depend on any specific segment of the radio access or the core network. Since the field tests can be done automatically without any technician interference, the network capacity planning can be efficiently conducted at any time.

A. Geographic Region Based Deployment

The system consists of three main components as shown in Fig. 2: (i) a client-side distributed speed test tool, called Speed Meter, running on smart devices as a daemon service using the ISP’s address range, (ii) a speed test data analysis system, called “Speed Analyzer”, running on Hadoop, and (iii) a dashboard application presenting analysis results for decision making.

1) Speed Meter with End-to-End Packet Analysis

The Speed Meter, implemented in Python and deployed as a daemon service on Linux, runs on geographically distributed mobile clients. It captures live data packets from the network and dumps them in pcap format using Dumpcap [21]. As a part of the test process it establishes a communication link with a server on the Internet at certain intervals for downloading and uploading various test files.

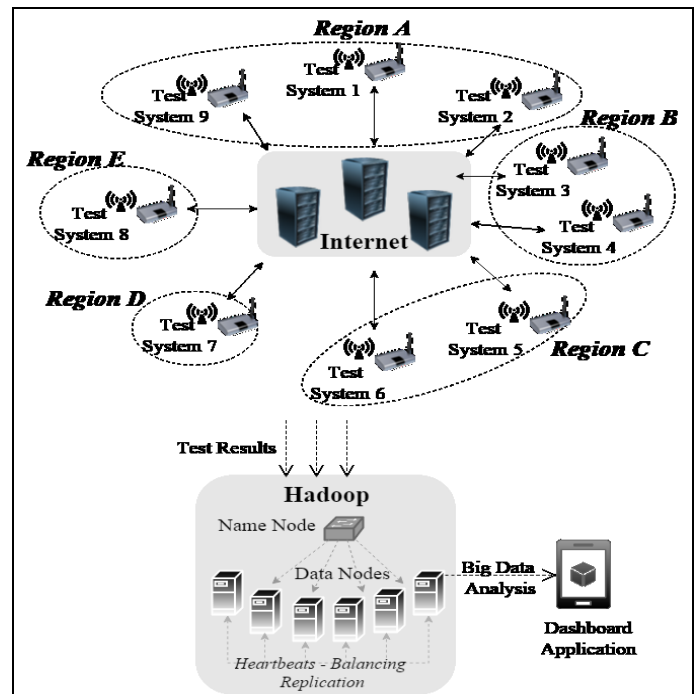


Fig. 2. Proactive ISP Speed Meter System (Geographic Region based deployment). It has three components of (i) the Speed Meter (running on clients), (ii) A Hadoop based speed analysis tool, and (iii) a dashboard application.

Tcptrace [22, 23] is used to analyze data captured by Dumpcap. The tool can produce several different types of output containing information on each connection seen, such as elapsed time, segments sent and received, retransmissions, round trip times, window advertisements, throughput, and more. The Speed Meter computes the download and upload speeds, RTT, average number of retransmissions grouped for each server name. Finally, the Speed Meter test results are pooled on a common area for further aggregate analysis. A sample output of which is shown in Table 2.

TABLE II
A SAMPLE SPEED METER RESULTS DATA

Sample Test Result					
System ID	Download (Mbit/s)	Upload (Mbit/s)	RTT (ms) (a2b b2a)	Retrans mit (a2b b2a)	Server
AnYen1	04.74	2.92	230.38 0.86	14 0	Vodafone
AnYen1	09.20	2.94	223.11 0.11	09 3	Avea
AnYen1	08.70	3.03	265.24 0.07	08 1	Doruknet
AnCan79	10.09	3.01	169.51 0.05	10 0	Vodafone
AnGol35	07.20	3.10	228.84 0.09	12 0	Vodafone
AnCan94	11.27	2.92	175.04 5.74	01 0	Avea

Not only the aggregate test results but individual test results collected by the Speed Meter is informative too. For instance, the hypothetical sample results given in Table 2 show horrendous user experience, e.g., file downloads taking far longer to complete. We outline the speed test results for each geographic region separately.

The Speed Meter algorithm is shown in Algorithm 1. Every 10 seconds (a configurable value) it uploads and downloads a packet and these communications are recorded by Dumpcap. The recorded communications are then analyzed by Tcptrace to give download/upload speeds and RTT for unlost communications. The speed test results are then sent to Hadoop for ISP wide analysis.

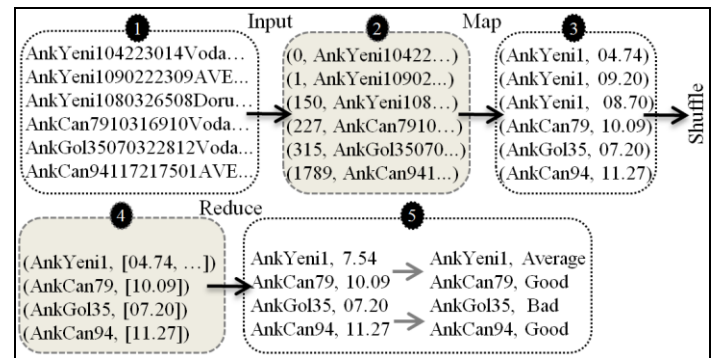


Fig. 3. Map and Reduce Data Flow. It shows an example run for download speed given in Table 2.

Algorithm 1 The Speed Meter Algorithm

```

1: timer = 10sn. start
2: while timer do
3:   while call("dumpcap") do
4:     runSpeedTest()
5:   end while
6:   Determine: results ← call("tcptrace")
7:   if results != NULL then
8:     download ← calculateDownloadSpeed()
9:     upload ← calculateUploadSpeed()
10:    rtt ← calculateRtt()
11:   end if
12:   print download, upload, rtt
13: end while

```

Algorithm 1. The Speed Meter Algorithm. Each client runs this algorithm and emits the performance results to network speed logs, which is analyzed afterwards by speed data analysis component on Hadoop.

2) Performance Analytics with MapReduce

Coming from various regional Speed Meters, the speed test results as large data sets are processed and analyzed in parallel and distributed fashion. The Map and Reduce jobs are implemented in Java.

MapReduce works in the map and the reduce phases. Key-value pairs act as input and output at each phase. The input to the map function are raw test results as shown in Table 2. The map function (Algorithm 2) extracts system id with respective KPIs and emits them as its output. Then in the shuffle operation, the key-value pairs are grouped and sorted by key. After shuffling, the reduce function (Algorithm 3) averages the performance metrics to obtain the network speed. The speed labeling, indicated as “Good”, “Average” and “Bad” tickets, qualifies the user experience for each test region. We use two thresholds, the value for which are experimentally selected, for ticket labeling. The data flow is illustrated in Fig. 3, with an example download speed obtained from Table 2.

Algorithm 2 Map(docId a, doc d)

```

1: for term : docd do
2:   systemId ← getSystemId(term)
3:   download ← getDownload(term)
4:   upload ← getUpload(term)
5:   rtt ← getRtt(term)
6:   Emit(systemId, AssociativeArray[download, upload, rtt])
7: end for

```

Algorithm 2. The Mapper algorithm.

Algorithm 3 Reduce(systemId s, values[d1,d2...] d)

```

1: for value : d do
2:   downloadTotal ← downloadTotal + value[0]
3:   uploadTotal ← uploadTotal + value[1]
4:   rttTotal ← rttTotal + value[2]
5:   counter ++
6: end for
7: ticketDownload ← label(totalDownload/counter)
8: ticketUpload ← label(totalUpload/counter)
9: ticketRtt ← label(totalRtt/counter)
10: Emit(s, AssociativeArray[ticketDownload, ticketUpload, ticketRtt])

```

Algorithm 3. The Reducer algorithm.

3) Dashboard Application

The Dashboard application contains a geographic region map for visualization of the respective tickets. The tickets are as follows: green colored “Good” ticket, red colored “Bad” ticket and yellow colored “Average” ticket. We would like to note that the visualization is online to allow enough time for the decision makers to react in a timely manner.

B. Content Delivery Service Based Deployment

With the ever-increasing mass access to certain content (e.g., popular videos, recent albums, flash news etc.) from the Internet sources, most ISPs either mirror such contents or cache them for fast accesses for their subscribers. To do so,

they run specialized content delivery servers and check their performance frequently. For the sake of the overall QoE, any bottleneck due to content delivery servers needs to be identified and resolved instantly. Unfortunately, Geographic Region Based Deployment cannot directly provide content delivery server based user experience. For this reason, our system has another deployment named Content Delivery Service Based Deployment.

The Content Delivery Service Based Deployment (Fig. 4) consists of three main components. First component is the logs collector interface running as a daemon service. The logs collector warehouses performance indicators of network centric measurement and connection stream data (e.g., media & web QoE metrics) on the content delivery servers. The Hadoop-based logs analyzer is the second component. The analyzer, implemented in Java, derives high-quality structured data from unstructured text using entity extraction. The third component, implemented in C#, is a GUI dashboard application presenting analysis results for proactive decisions. The dashboard application constantly monitors the network status and provides users with decision making for operational excellence through competitive intelligence reports.

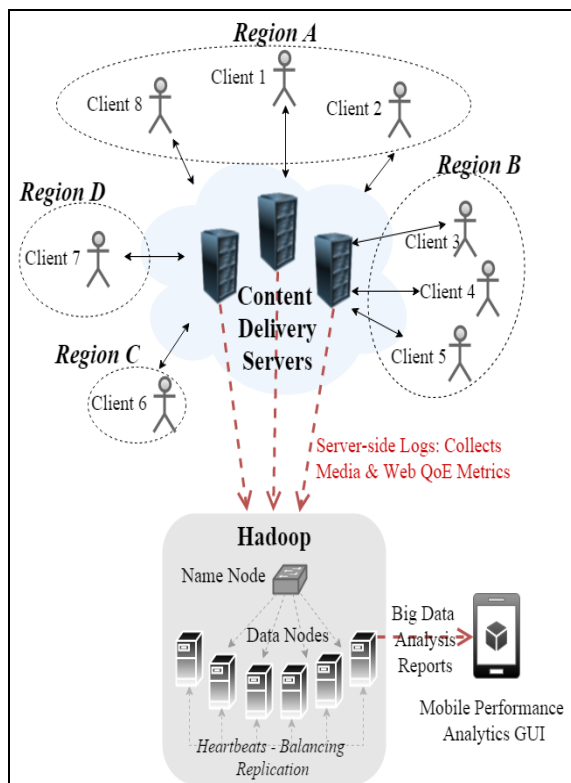


Fig. 4. Proactive ISP Speed Meter System (Content Delivery Category based deployment). It has three components; (i) server side logs collector (running on the server), (ii) a Hadoop based performance analytics tool, and (iii) a GUI dashboard application.

The user experience can be measured from server-side as it is possible to measure frequency of successful packet transmissions to any client device within the ISP’s address range. Moreover, separate frequencies can be measured for each content delivery service, hence giving us the utilization of the respective service. The deployment (Fig. 4) collects performance indicators of network centric measurements and

connection streams on the server-side. The connection logs are processed and analyzed in distributed fashion, which enabled us to easily and quickly process an extremely large amount of speed tests.

Fig. 5 shows the steps of the MapReduce operation which follows the same basic steps as given in Section 3.1.2.

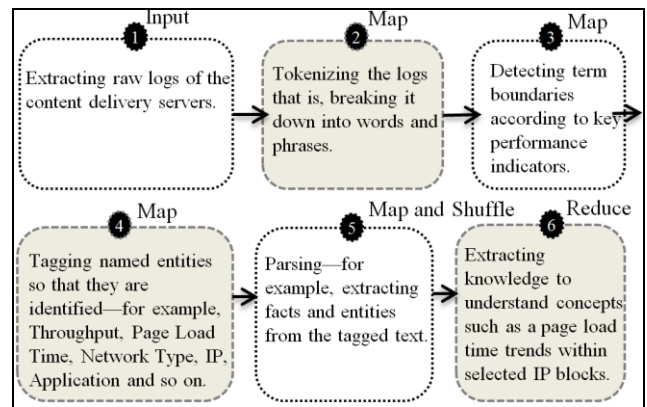


Fig. 5. Flow of MapReduce based Analysis.

Geographic Region Based Deployment requires client-side Speed Meters distributed across every geographic region and each of them needs to execute the speed tests constantly. That deployment has the advantage that user experience can be measured directly and accurately. However, it has the overhead of client-side installations and maintenance. Content Delivery Service Based Deployment, on the other hand, does not need client-side components as the objective with which is to measure performance of servers. It measures server-side initiated end-to-end network speed by listening actual user (subscriber) connections to content delivery servers.

The client-side analysis with Geographic Region Based Deployment provides only network centric KPIs of download/upload speeds and RTT. However, server-side analysis with Content Delivery Service Based Deployment provides us with richer set of KPIs, including congestion, page load time, TCP connection time, dropped frames, content category rate, time to playback, errors, DNS lookup time end, in addition to download/upload speeds and RTT. As a result, KPIs of the Geographic Region Based Deployment are purely network centric, while that of the Content Delivery Service Based Deployment are network centric plus user and competitive intelligence centric. Additionally, server-side analytics is designed to use real user connections. Monitoring web access performances combined with video analytics (video startup time, re-buffering, avg. bit rate) allows mobile ISPs to fully understand how websites load and videos play on the end-user device.

For some specific situations of the network problems, user and network centric KPIs may be analyzed, combined and correlated together. Because of that, server-side analysis system can enable us with analysis on different domains such as network centric, user centric and competitive intelligence centric. For instance, some problems may be solely caused by

network centric, while some others may be caused by combinations of the network centric and user centric. Therefore, both network and user centric analyses offer advantages for the detection of network problems [24]. Like Geographic Region Based Deployment, Content Delivery Service Based Deployment analyzes big data to improve network services, user needs and businesses value [25].

IV. EXPERIMENTAL EVALUATION

The proposed Internet Speed Analysis System is tested on the network of a leading mobile Internet ISP of Turkey. Our test network contains 20,500 base stations offering 3G and 4.5G services. This means that test results arrive in every minute from 20,500 stations and are pooled on the Hadoop platform. It is very difficult for a classical computing system to deal with such a high speed massive volume data. This is why we employed big data processing solutions including advanced analytics. This way we are able to transform the complex big data into actionable intelligence.

We experimented with both of the deployments and highlight a few interesting cases found during the live operation of the system. For the Geographic Region Based Deployment, we implemented the map with province granularity of Turkey which has 81 provinces. However, the implementation can be easily configured at cell, base station, neighborhood, county or specific region levels. In this way, the right granularity level will give more informative results.

A. Geographic Region Based Deployment

The most important feature of Geographic Region Based Deployment is scaling user experience and sentiments for each region separately. Provincial map of Turkey is shown in Fig. 6, where five interesting provinces (regions) are enumerated. The reason will be explained later.

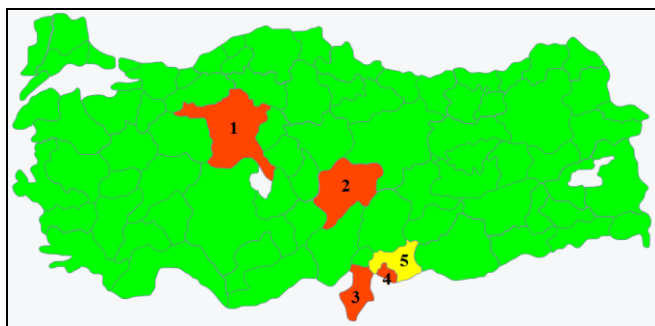


Fig. 6. Provincial map of Turkey, where five interesting provinces have been enumerated.

1) User Experience

The speed characteristics map as shown in Fig. 6 was obtained with plot analysis of test results, an example of which is given in Fig. 7. As shown in Fig. 7, download rates dropped sharply starting from July 29 in Region 1 (the plot for the Region 2 was similar). This is why Regions 1 and 2 in Fig. 6 are colored with red. The cause of the speed drop is being

identified as an operational work (i.e., a maintenance of the base stations) carried out in these regions. These days, we experienced similar plots for Regions 3, 4 and 5. Since the speed drop is sharp in Regions 3 and 4, they are colored with red. However, Region 5 is colored with yellow as the speed drop in this region was moderate. It has been later understood that Syrian refugees caused an excess traffic in Regions 3, 4 and 5. This excess demand exceeded maximum capacity of the base stations in these regions.

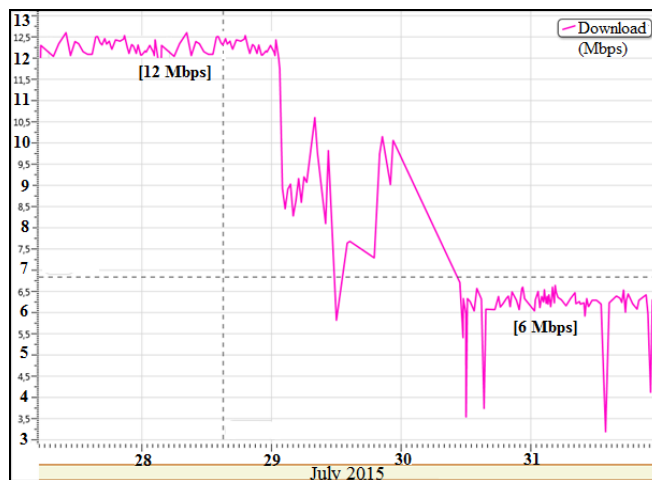


Fig. 7. Download Speed Rate Plot for Region 1.

Fig. 7-8 presents the correlation of our test results with the user feedbacks (complaints) from the Regions 1 and 2 of Fig. 6. Mobile users from the Regions 1 and 2 reported network speed problems on 29-31 July and 1-3 August. In these days, the fall of user satisfaction (Fig. 8) coincides with the decreasing download speed (Fig. 7). The agreement between Fig. 7 and 8 shows that our speed analysis system can proactively scale true user experience and network speed quality.

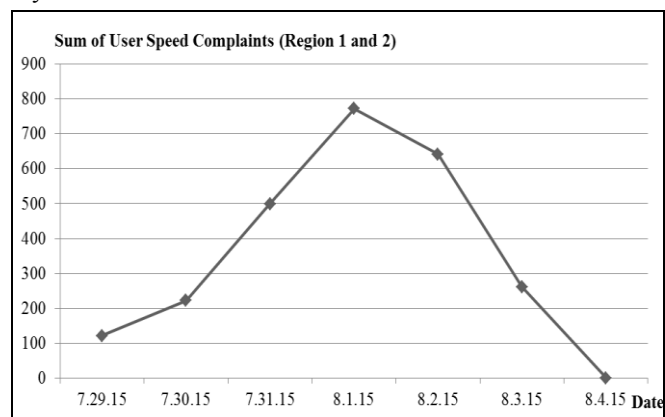


Fig. 8. Speed Satisfaction Feedbacks/Complaints.

2) Detection of Problems Before User Complaints

A Denial-of-Service (DoS) attack targeted the ISP servers providing Internet infrastructure, and as a result users encountered Internet access problems in some regions enumerated in Fig. 9. Users of the Regions 1, 2, 3, 4 and 5 are affected from the DoS attack for 18 minutes. Fig. 10 gives the

network speed change for the Region 1 of Fig. 9. As shown in Fig. 10, the DoS attack started at around 12:00 and the speed was suddenly increased. Shortly after, some packet core network servers were unable to serve to mobile users. Our system was operational at that period and detected this DoS attack through user experiences and raised the red/yellow alarms for the respective regions. In a short time of minutes, system administrators were able to intervene. As a result, the problem is detected and fixed right before massive user complaints arrive. Around 12:18 the Internet usage/speed experience came to normal.

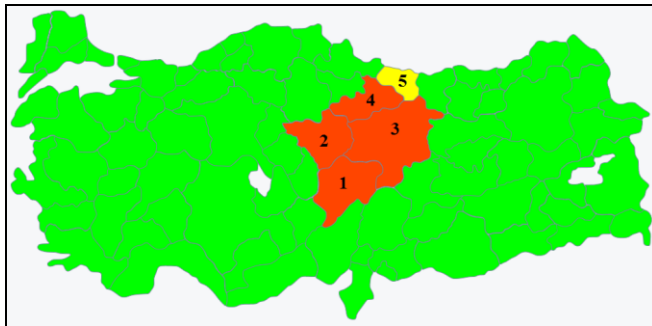


Fig. 9. DoS attack regions.

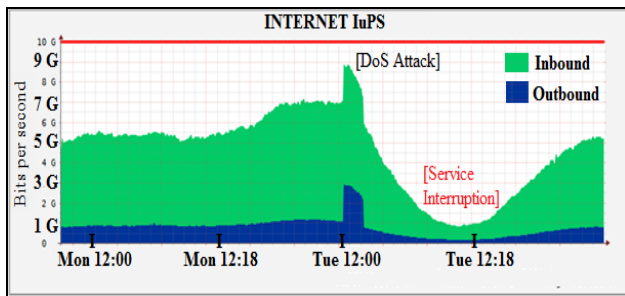


Fig. 10. Internet usage/speed in Region 1.

B. Content Delivery Service Based Deployment

The most important feature of Content Delivery Service Based Deployment is scaling user experience and sentiments for each service separately.

1) User Experience

Some of the web services of packet network were adversely affected after an operational work of maintenance. The affected regions are shown in red in Fig. 11. As a result, Internet usage rates quickly fell in Sunday at around 00:00 (Fig. 12).



Fig. 11. Operational work in Internet service servers. The maintenance is carried out in regions marked with red.

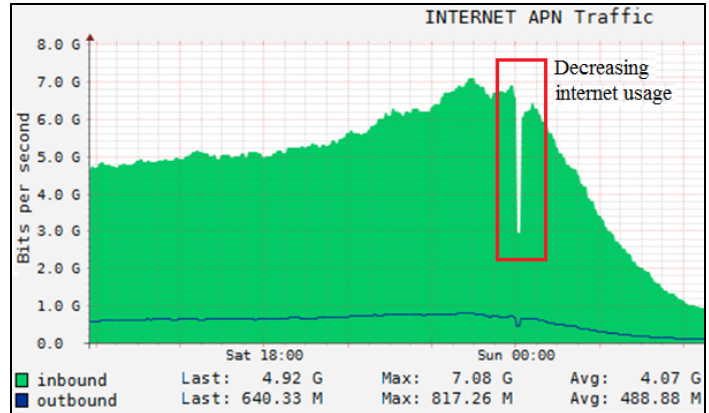


Fig. 12. Internet usage/speed in Region 1.

2) Competitive Intelligence

Competitive intelligence can be gathered from analyzing patterns of customer usage and mobile network data segmentation on the content servers. Our system is able to show reports about patterns of actual customer usage due to real user activities. Therefore, collected logs of the servers include all Internet activities of users such as click patterns, connection streams, performance indicators of network centric measurement, information of data segmentation and page views and similar. Hence, this deployment does not only scale user experience, but also analyses competitive intelligence.

In the months of March, April and May, the deployment was run on the content delivery network to analyze real Internet usage. For our experimental objective, we projected analysis results on different categories of service based, VOIP based, application based, video based and instant messaging based cases in the ISP’s mobile network. The results of competitive intelligence are given in Fig. 13 through Fig. 17. Note that these figures are very informative as far as comparing the demand for the related services/tools/applications.

The competitive intelligence reports provide the ISPs with valuable information of highly demanded services. Hence, ISPs can use this information to decide on their future investment plan.

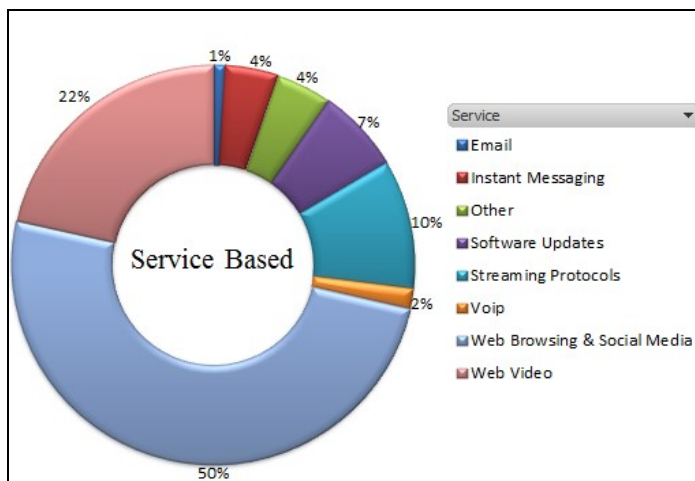


Fig. 13. Service Based Analysis.

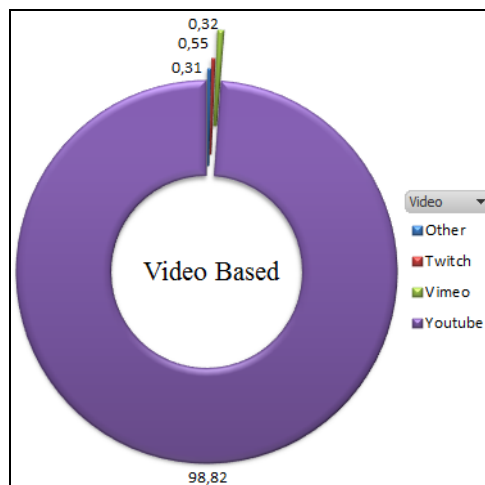


Fig. 16. Video Based Analysis.

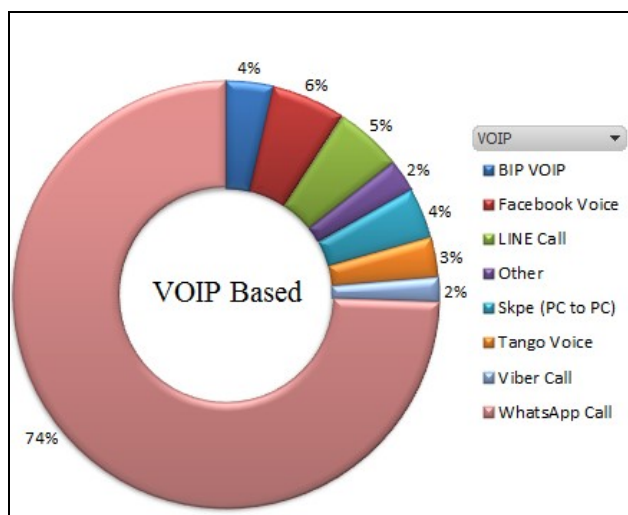


Fig. 14. VOIP Based Analysis.

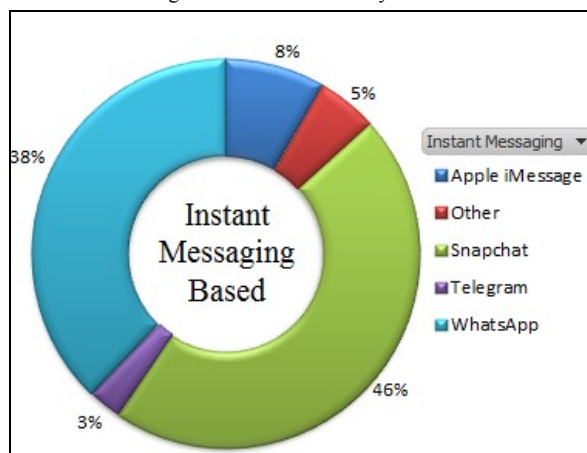


Fig. 17. Instant Messaging Based Analysis.

V. CONCLUSION

Today wider user community services like social media, online gaming, video streaming and VoIP demand more network speed and high bandwidth capacity. For mobile ISPs, Internet service quality is a very strategic issue to satisfy user demands and to be competitive as well. Therefore, constantly monitoring Internet speed and true user experiences have the key role for customer satisfaction. However, correctly measuring the true user experiences in online fashion and promptly intervening the network after any service degradation involve the coordination of right speed measurement, right speed analysis and right decision making facilities.

In this paper, we proposed an Internet speed analysis system for mobile ISPs. The system called Proactive ISP Speed Meter exploits end-to-end network speed measurement to correctly gauge user experience. The system has two deployments: the client-side initiated geographic region based, and the server-side initiated content delivery service based. The former deployment analyzes Internet access speed from each geographic region while the latter deployment analyzes Internet access speed per content delivery category and can additionally provide competitive intelligence.

Big Internet traffic data coming from various clients of the mobile ISP network are processed and analyzed using

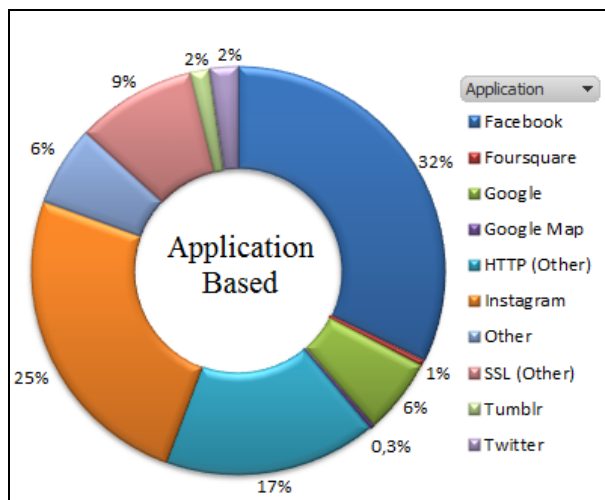


Fig. 15. Application Based Analysis.

MapReduce programming model on Hadoop platform. The analysis results of regional speed characteristic maps and service usage plots are useful for operational excellence. Most importantly, our system is able to predict potential mass user complaints proactively and provide warnings in time.

The utility of our system has been demonstrated on a live network of a leading mobile ISP of Turkey.

REFERENCES

- [1] Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white_paper_c11-520862.pdf, Last accessed Feb 2020.
- [2] H. D. Trinh, N. Bui, J. Widmer, L. Giupponi and P. Dini, "Analysis and modeling of mobile traffic using real traces," 2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), Montreal, QC, Canada, ISBN: 978-1-5386-3531-5.
- [3] R. Pries, F. Wamser, D. Stachle, K. Heck and P. Tran-Gia, "Traffic Measurement and Analysis of a Broadband Wireless Internet Access," Vehicular Technology Conference, 2009. VTC Spring 2009. IEEE 69th, Barcelona-Spain, ISBN: 978-1-4244-2517-4.
- [4] W. Robitza, A. Ahmad, P.A. Kara, L. Atzori, M. G. Martini, A. Raake and L. Sun, "Challenges of future multimedia QoE monitoring for internet service providers," *Multimed Tools Appl* (2017) 76: 22243, <https://doi.org/10.1007/s11042-017-4870-z>.
- [5] M. Uzun and O. Abul, "End-to-end internet speed analysis of mobile networks with MapReduce," 2016 International Symposium on Networks, Computers and Communications (ISNCC), Yasmine Hammamet, Tunisia, ISBN: 978-1-5090-0284-9.
- [6] M. Kyryk, N. Pleskank and M. Pitsyk, "QoS mechanism in content delivery network," 13th International Conference on Modern Problems of Radio Engineering, Telecommunications and Computer Science (TCSET), 2016, Lviv, Ukraine, ISBN: 978-6-1760-7807-4.
- [7] M. Taruk, E. Budiman, Haviluddin and H. J. Setyadi, "Comparison of TCP variants in Long Term Evolution (LTE)," 5th International Conference on Electrical, Electronics and Information Engineering (ICEEIE), 2017, Malang, Indonesia, ISBN: 978-1-5386-0355-0.
- [8] L. Khoshnevisan, F. R. Salmasi and V. Shah-Mansouri, "An adaptive rate-based congestion control with weighted fairness for large round trip time wireless access networks," 24th Iranian Conference on Electrical Engineering (ICEE), 2016, Shiraz, Iran, ISBN: 978-1-4673-8789-7.
- [9] F. Ahmed, J. Erman, Z. Ge, A. X. Liu, J. Wang and H. Yan, "Detecting and Localizing End-to-End Performance Degradation for Cellular Data Services Based on TCP Loss Ratio and Round Trip Time," *IEEE/ACM Transactions on Networking*, Volume: 25, Issue: 6, p: 3709–3722, Dec. 2017, ISSN: 1558-2566.
- [10] P. Skocir, D. Katusic, I. Novotni, I. Bojic and G. Jezic, "Data rate fluctuations from user perspective in 4G mobile networks," 22nd International Conference on Software, Telecommunications and Computer Networks (SoftCOM), 2014, ISBN: 978-9-5329-0052-1.
- [11] A. S. Khatouni et al, "Speedtest-Like Measurements in 3G/4G Networks: The MONROE Experience," 29th International Teletraffic Congress (ITC 29), 2017, Genoa, Italy, ISBN: 978-0-9883045-3-6. <http://www.speedtest.net/>, Last accessed Feb 2020.
- [12] L. Velasco, "Managing services in the telecom cloud: An example for CDN," 18th International Conference on Transparent Optical Networks (ICTON), 2016, Trento, Italy, ISBN: 978-1-5090-1467-5.
- [13] S. Wang, X. Wang, J. Huang, R. Bie and X. Cheng, "Analyzing the potential of mobile opportunistic networks for big data applications," *IEEE Network*, Volume: 29, Issue: 5, September-October 2015, ISSN: 0890-8044.
- [14] J. Dean and S. Ghemawat, "MapReduce: Simplified data processing on large clusters," In *Proceedings of the 6th Symposium on Operating System Design and Implementation (OSDI 2004)*, pages 137-150, San Francisco, California, 2004.
- [15] J. Dean, and S. Ghemawat, "MapReduce: a flexible data processing tool," *Communications of the ACM*, 53(1), pp.72-77, 2010.
- [16] Hadoop, <http://hadoop.apache.org/>, Last accessed Feb 2020.
- [17] P. Zikopoulos and C. Eaton, "Understanding big data: Analytics for enterprise class hadoop and streaming data," McGraw-Hill Osborne Media, 2011.
- [18] Y. Qiao, Z. Xing, Z. M. Fadlullah, J. Yang and N. Kato, "Characterizing Flow, Application, and User Behavior in Mobile Networks: A Framework for Mobile Big Data," *IEEE Wireless Communications*, Volume: 25, Issue: 1, February 2018, ISSN: 1536-1284.
- [19] G. H. M. Almeida, E. H. R. Coppoli, E. N. Goncalves, M. M. Afonso and U. C. Resende, "Traffic flow management in a real mobile phone network using linear optimization," *IEEE Latin America Transactions*, Volume: 16, Issue: 2, Feb. 2018, ISSN: 1548-0992.
- [20] Dumpcap, <https://www.wireshark.org/docs/man-pages/dumpcap.html>, Last accessed Feb 2020.
- [21] R. Blum, "Network Performance Open Source Toolkit: Using Netperf, tcptrace, NISTnet, and SSFNet," John Wiley & Sons, 2003.
- [22] Tcptrace, <http://www.tcptrace.org/>, Last accessed Feb 2020.
- [23] M. Dighiri, G.M. Lee and T. Baker, "Measurement and Classification of Smart Systems Data Traffic Over 5G Mobile Networks," In: Dastbaz M., Arabnia H., Akhgar B. (eds) *Technology for Smart Futures*. Springer, Cham, ISBN: 978-3-319-60137-3, 2018.
- [24] B. Zhou, J. Li, S. Guo, J. Wu, Y. Hu and L. Zhu, "Online Internet Traffic Measurement and Monitoring Using Spark Streaming," *GLOBECOM 2017 - 2017 IEEE Global Communications Conference*, Singapore, Singapore, ISBN: 978-1-5090-5019-2.
- [25] J. Sandoval, A. Ehijo, A. Casals and C. Estevez, "New Model and Open Tools for Real Testing of QoE in Mobile Broadband Services and the Transport Protocol Impact: The Operator's Approach," *IEEE Latin America Transactions*, Volume: 13, Issue: 2, Feb. 2015, ISSN: 1548-0992.

BIOGRAPHIES



METE UZUN received his B.S. and M.S. degrees in computer engineering from the TOBB University of Economics and Technology. His research interests include mobile networks and machine learning.



OSMAN ABUL is a professor of Computer Science at TOBB University of Economics and Technology, Ankara, Turkey. He received his Ph.D. degree in Computer Engineering from Middle East Technical University, Ankara, Turkey. He held visiting posts in University of Calgary, Norwegian University of Science and Technology, and Italian Institute of Information Science and Technology. His research interests include data mining, privacy and bioinformatics.

Semi-conductor Applications to Printed Circuits on Flexible Surfaces

Nazmi Ekren and Ali Samet Sarkin


Abstract— The most common type of identification system today is RFID. RFID circuits are used as covered with plastic. With the increase in usage areas, it is also used on metal, wood, paper, and plastic product. In this study, the behavior of the same circuit on different surfaces was investigated. The surface impedance and signal reflection coefficients of RFID tag antennas were investigated based on paper, plastic, and textile surfaces. According to the results of the electrical and mechanical tests, the best results in terms of reflectance coefficients and surface impedances of RFID tags are on PET surfaces. The surface impedance and the reflection coefficients were high on paper surfaces. The lowest values were measured on textile surfaces. According to the results, it has been seen that RFID antenna application on plastic, paper, and textile surfaces is possible and usable.

Index Terms— Flexible surface, Identification, Reflection coefficient, RFID, Surface impedance


I. INTRODUCTION

AUTOMATIC IDENTIFICATION and data capture (AI/DC) systems are the identification of information or feature by matching with a person or object [1]. Identification processes are made with barcode, optical character recognition, magnetic stripe, smart cards, RFID, and biometric systems. The oldest type of automatic identification and data capture systems is the barcode system. Barcodes can be encoded by Code 39, EAN (European Article Numbering) 8, EAN 13, and QR Code standards. In the Code 39 coding system, 29 letters, 10 numbers, and 8 special characters can be defined in the barcode [2]. The EAN standard works with the logic of the binary code system. By encoding thin and thick bars as numerical logic, a series of numbers is obtained, and information is kept in the database as the equivalent of this

NAZMI EKREN, (*corresponding author), is with the Department of Electrical and Electronical Engineering, Marmara University, Istanbul, Türkiye, (e-mail: nazmiEkren@marmara.edu.tr).

 <https://orcid.org/0000-0003-3530-9262>

ALİ SAMET SARKIN, is with Department of Electrical and Energy, Osmaniye Korkut Ata University, Osmaniye, Türkiye, (e-mail: sametsarkin@osmaniye.edu.tr).

 <https://orcid.org/0000-0001-6261-0531>

Manuscript received March 28, 2022; accepted June 22, 2022.

DOI: [10.17694/bajece.1094805](https://doi.org/10.17694/bajece.1094805)

number. All commercial products are made under the EAN standard. A large amount of information cannot be entered into the content of the barcode and no direct information is available. Barcode systems are very low costs, rapid, and easy to implement. In addition to their advantages, barcode systems cannot respond to long-distance needs and collective readings due to the need for individual processing and the need to be read from a short distance. In addition, the barcode can be scratched, torn, and unreadable.

The RFID system can respond to these needs. Magnetic stripe and optical character recognition systems are other systems used in product/identity identification. In these systems, reasons such as tearing and scratching in the barcodes may occur due to deformation, where the identity cannot be identified. In biometric systems, retina, iris, fingerprint, face, voice, hand geometry, vascular map, and signature are used [3]. However, for the application of these methods, a person and close distances are required [4]. In addition, they cannot be applied to commercial products and goods.

The RFID system is expressed as Radio Frequency Identification. The RFID system works based on radiofrequency. It consists of an antenna called a tag, a processor where the data is recorded, a transmitter-transceiver, and an interface program where the data is processed (Fig.1.) [5]. When the signal sent by the transmitter is captured by the antenna on the RFID tag, the unique ID number in the processor on it is sent back to the transceiver device via the magnetic field [3]. Electrical circuits are located on the PCB surface. The RFID antenna, on the other hand, can be found in a flexible form without the need for a hard surface PCB.



Fig.1. RFID Components[5]

RFID systems can be classified according to usage purposes. Its usage areas have increased considerably today,

and it is frequently used in daily life. It is common in areas such as supply chain and tracking, animal identification, libraries, cargo, bridges and highways, parking lots, container tracking, city bus cards, bank cards, and hotels [6].

RFID systems are divided into active, passive, and semi-active. Passive RFID, examples of which are given in Figure II., interacts with the power of the incoming signal without a power supply. They are inductively coupled, and AC current is induced. The induced voltage, which is converted to DC current with a simple rectifier, enables the chip to work [7]. For coupling to occur, the antenna and the reader must be close enough to each other. The range capacity is the reading distance. To increase the reading distance, it is necessary to increase the frequency and power.

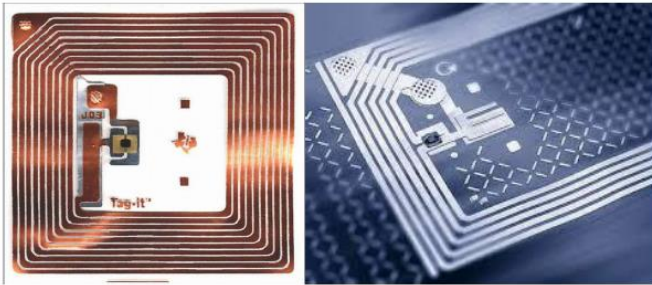


Fig.2. Passive RFID antennas[8]

Passive RFIDs are usually used at distances of less than 15 mt. It is the most cost-effective and simple type of RFID. For this reason, it has wide uses as identity control, contactless bank cards, bus cards, animal ID, and security alarms for goods. They contain a power supply in the active and semi-active systems. Because it has a power source, they do not need an electrical signal wave from the reader for communication. The power supply supplies energy to the chip on which the data is located and provides the energy for communication. Semi-active tags provide readings up to 50-100 mt, while active tags can be used in applications at a longer distance [9,10].

There are 4 operating frequencies used in the active, semi-active, and passive RFID systems. These frequencies are shown in Table I.

TABLE I
RFID FREQUENCY AND USAGE AREAS

Frequency Type	Frequency Range	Usage
Low Frequency	125-134 kHz	Short Distance- up to 15 cm
High Frequency	13.56 MHz	Short Distance- up to 1 mt
Ultra-High Frequency	860-960 MHz	Supply Chain- up to 10 mt
Microwave	5.8 GHz	Long-Distance and active RFID- up to 100 mt

RFID applications use the frequency ranges in Table I. RFID tags are used not only on flat and rough surfaces but also on flexible and bendable surfaces. Similar studies in the literature were investigated, and the effect of RFID tags operating at UHF (865-868 MHz) frequency on the adhesion of chips to the textile surface and textile antennas was performed in a study. Three different adhesive interfaces were used. The reading distances were compared, and it was seen that the RFID tag was read from a distance of 3.41 m in the

laboratory environment and a distance of 2.05 m when measured on a human. Although the human body reduced the reading distance, the antennas on the textile showed stable performance [11]. An RFID antenna in the 902-908 MHz UHF band was designed and tested on curved and flat surfaces. The antenna performance was evaluated through simulations of input impedance, reflection coefficient, gain and read range. According to the results of the analysis, it has been proven that it can be used on flexible and curved surfaces by obtaining a reading distance of more than 1 m on cylindrical surfaces [12]. In another study, tensile and bending tests were performed by placing the RFID tag in glass epoxy resin, and the effect of the RFID tag on the bending test was determined. It was observed that the RFID tag did not affect the tensile test [13]. Another study, it was aimed to design and produce flexible UHF RFID tags used in bad environmental conditions. Flexible labels are intended to perform well even underwater and chemical exposure when applied to non-metallic surfaces. Polyamide surface was chosen as the surface to be printed. After the RFID circuit was applied to the surface, the reflection coefficient (dB) was examined, and it was found at which frequency the best result was measured [14].

II. MATERIAL AND METHOD

Within the scope of the Master's thesis entitled "Semiconductor Applications to Printed Circuits on Flexible Surfaces", the effects of RFID tag application on flexible surfaces such as textiles, fabrics, and plastics were investigated [15]. Avery Dennison brand Passive UHF inlay antennas/tags were provided for the first phase of the thesis work. The protective plastic layer on the labels was dissolved by applying chromic acid ($H_2SO_4+K_2Cr_2O_7$) in an amount of 1 ml for 7 minutes. Afterward, the passive RFID inlay adhered to the paper and the knitted fabric textile surface with a butadiene-acrylonitrile-based adhesive at 150-200°C for 3 minutes and at 150-200°C for 5 minutes, respectively. It was adhered to the plastic surface without applying heat. Akramin BA-N adhesive was preferred because of its high binding capacity, high mechanical stability, and softness and elasticity to the prints [16, 17].

The obtained samples for each surface are shown in Fig.3., Fig.4., and Fig.5.



Fig.3. Application on the paper surface



Fig.4. Application on the plastic surface



Fig.5. Application on the textile surface

After the antennas were applied to the surfaces, the surface resistances were measured firstly with a multimeter. Then, with the Rohde Schwarz ZVL brand Network Analyzer capable of measuring in the 9 kHz – 13.6 GHz frequency range, the frequencies of the RFID tag samples operating at maximum efficiency, that is, minimum loss, and their impedances at these frequencies were measured. A 500-ohm resistive cable was used in the Network Analyzer and the device was set to S11. In the last experimental setup, a transmitting antenna and a receiving antenna setup were set up. A dipole antenna is placed at 1 meter from the transmitting antenna as the receiving antenna. RFID antennas were connected to the Anritsu MT 2605B signal generator. The dipole antenna used as the receiving antenna was connected to the Anritsu MS710C spectrum analyzer that can measure in the range of 10 kHz – 23 GHz. The signal generator was adjusted to the frequency value of each sample measured in the network analyzer. In this way, the signal at the frequency at which the RFID antenna operates was sent to the tag. The frequency applied to the RFID antenna and the loss of this frequency in dBm of the dipole antenna were examined. The dipole antenna was fixed, but the RFID antenna was 0, 45, 90, 135, 180, 225, 270, and 315 degrees directional difference was created. This method allowed two antennas to compare power loss in a single direction.

III. RESULTS AND DISCUSSIONS

After the applications were made on 3 different surfaces, the surface resistances were measured. The surface resistances of 3 samples prepared from each surface were measured and the

average surface resistances of each sample are given in Table II.

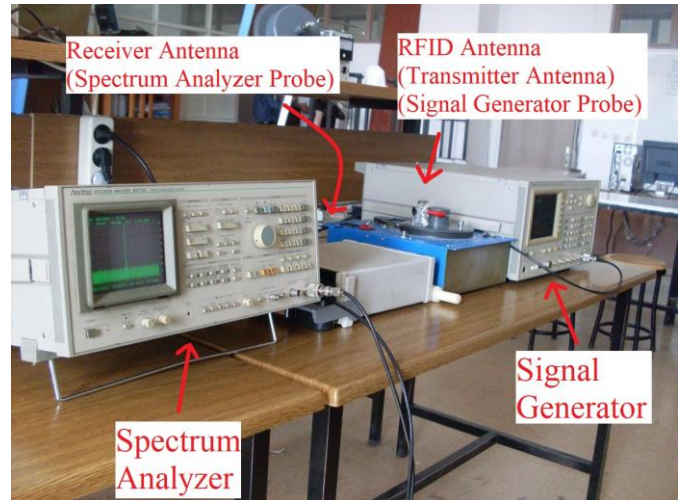


Fig.6. Experimental setup

TABLE II
SURFACE RESISTANCES

Surface	Sample-1 (Ω)	Sample-2 (Ω)	Sample-3 (Ω)	Average (Ω)
Plastic	0.39	0.40	0.38	0.39
Paper	0.80	0.90	0.88	0.86
Textile	1.2	0.9	1.3	1.13

Although the same RFID antenna was used in the 9 samples, different surface resistances were measured on different surfaces due to the surface effect when measuring surface resistances. While the surface resistance was very low on the plastic surface, more resistance was measured on the paper surfaces. It has been evaluated that the surface resistance had increased on all textile surfaces than others, and therefore the RFID reading distance would have been decreased or a higher operating frequency would have been required.

TABLE III
FREQUENCY, RETURN LOSS, AND IMPEDANCE

Surface	Frequency (MHz)	Return Loss (dB)	Impedance (Ω)
Plastic-1	982.600	-7.661	121.74-j4.063
Plastic-2	982.600	-7.786	102.66-j33.239
Plastic-3	982.000	-7.741	113.56-j26.452
Plastic Average	982.400	-7.730	112.65-j21.251
Paper-1	980.000	-7.882	98.194-j24.051
Paper-2	983.000	-8.293	111.04-j15.516
Paper-3	980.600	-8.006	104.22-j18.429
Paper Average	981.200	-8.060	104.484-j19.332
Textile-1	983.200	-9.621	81.922-j31.454
Textile-2	980.000	-7.268	111.60-j34.508
Textile-3	983.000	-8.800	96.650-j32.698
Textile Average	982.066	-8.563	96.724-j32.886

The operating frequencies, return losses, and impedances of the samples on the surfaces were measured with a network analyzer device. According to these results, as can be seen from Table III, there is no surface effect on the operating frequency. In examples of all surfaces, the operating frequency range is in the range of 980-983.2 MHz. It will be able to respond to RFID antennas in this frequency range.

Since the operating frequency range is close and equal to the values before applying, there will be no need to change the frequency on different surfaces. When the reflection coefficients and impedance are evaluated, the loss on plastic

surfaces is lower due to the high surface resistance and impedance. Textile surfaces have more reflection loss due to low impedance and high surface resistance.

TABLE IV
EFFECT OF DIRECTIONAL

Direction (degree)	Plastic-1 (dBm)	Plastic-2 (dBm)	Plastic-3 (dBm)	Plastic-Average (dBm)	Paper-1 (dBm)	Paper-2 (dBm)	Paper-3 (dBm)	Paper-Average (dBm)	Textile-1 (dBm)	Textile-2 (dBm)	Textile-3 (dBm)	Textile-Average (dBm)
0=360	-50	-54	-53.6	-52.53	-46	-46.6	-46	-46.2	-53.1	-51	-54	-52.7
45	-58	-57.4	-56	-57.13	-53	-55	-53.4	-53.8	-56.6	-57.1	-58	-57.23
90	-61	-63	-62.8	-62.27	-56	-59	-57.5	-57.5	-60.2	-65	-61.7	-62.3
135	-57.1	-59.8	-58	-58.3	-53.6	-58.8	-55	-55.8	-57.8	-60	-59	-58.93
180	-53.4	-56	-54	-54.47	-46.5	-50.6	-48.9	-48.67	-52	-56	-53	-53.67
225	-59.3	-60	-59	-59.43	-53.6	-59.2	-55.2	-56	-57.2	-61.4	-60.5	-59.7
270	-61.2	-61.2	-59	-60.47	-54	-63	-57	-58	-57.7	-64	-62	-61.23
315	-57.8	-58	-57	-57.6	-49.8	-54	-52	-51.93	-57.4	-62.6	-58	-59.33

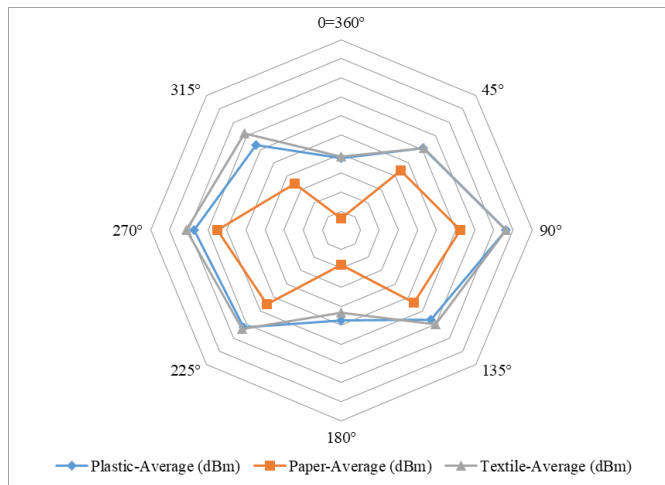


Fig.7. Chart of the reading loss

Table IV. shows the effect of direction on the response of RFID tags to the transmitting source. The results for all the samples are given in Table IV. Fig.7. is derived from Table IV. by taking the average values of each surface. The minimum power loss was observed on paper surfaces. The next minimum loss was on plastic surfaces. The highest power loss was on textile surfaces. It was observed that the loss was minimal on all surfaces in cases where the source and receiver antenna were located exactly opposite (0-180). The highest losses were measured in the reading directions of 90 and 270 degrees. It has been analyzed that the best results are on paper surfaces where power loss is minimal. According to these results, it has been evaluated that RFID tags (tags) can be applied to plastic, paper, and textile surfaces.

Operations and applications to be performed with RFID systems are made more practical and faster. While accuracy and reliability are increasing with RFID systems, system reliability is also increasing because manpower is reduced in the workflow. It is known that other biometric systems have their advantages and disadvantages. But the digital RFID system is also advantageous because they do not allow outside intervention compared to barcode systems.

IV. CONCLUSION

The study aims to examine the effects of the application of RFID systems on different surfaces. There are reading

problems and high reflection losses on metal and wet surfaces. It has been investigated whether such losses also occur on the applied surfaces. It has been observed that the reflection losses of RFID application on all three surfaces are less than on metal or wet surfaces. According to the results on paper and plastic surfaces, no change was observed in the operation of the RFID tag. Although reflection losses are higher in RFID tags on the textile surface, the RFID tag will work normally like other surfaces at shorter distances. According to the analyzed results, it has been evaluated that RFID applications are usable on flexible surfaces.

ACKNOWLEDGMENT

This study is derived from my Master's thesis (MSc) which is titled "Esnek Yüzeylerdeki Baskı Devrelere Yarı İletken Uygulamaları" and in English "Semiconductor Applications to Printed Circuits on Flexible Surfaces" under supervisory Nazmi Ekren.

REFERENCES

- [1] Sweeney P. J., "RFID for Dummies", 1st Edition, Wiley Publishing Inc., Indiana, USA, 2005.
- [2] <https://www.upu.int/UPU/media/upu/files/postalSolutions/programmesAndServices/standards/S10-12.pdf> Online: 15.01.2022
- [3] Bilgin, M., "Biyometrik Seçim Sistemi Tasarımı ve Gerçekleştirilmesi", Thesis of MSc, Selçuk University Institute of Pure and Applied Sciences, Konya, Türkiye, 2008.
- [4] B. Eren, "Biyometrik Teknolojilerin Etkili Tasarlanması ve Uygulanmasında Yeni Bir Öneri: Multimodel Teknoloji", Thesis of MSc, Mimar Sinan Fine Arts University Institute of Pure and Applied Sciences, İstanbul, Türkiye, 2009.
- [5] J.M. Sardroud, "Influence of RFID technology on automated management of construction materials and components", Scientia Iranica, Vol. 19, No. 3, 2012, pp. 381-392.
- [6] X. Zhu, S.K. Mukhopadhyay, H. Kurata, "A Review of RFID Technology and Its Managerial Applications in Different Industries" Journal of Engineering and Technology Management, Vol. 29, No. 1, 2012, pp. 152-167.
- [7] Y. Wang, Y. Huang, Y. Li, P. Cheng, S. Cheng, Q. Liang et. al., "A facile process combined with roll-to-roll flexographic printing and electroless deposition to fabricate RFID tag antenna on paper substrates", Composites Part B: Engineering, Vol: 224, November 2021, pp. 109194.
- [8] J. Shekhar, D. Zerihun, M.A. Haile, B. Shifaw, "Automated classroom monitoring with IOT and virtuino app", International Journal of Advanced Research in IT and Engineering, Vol. 8, No. 3, March 2019, pp. 1-17.

- [9] S. Kalaycı, "Design of A Radio Frequency Identification (Rfid) Antenna", Thesis of MSc, Middle East Technical University Institute of Pure and Applied Sciences, Ankara, Türkiye, 2009.
- [10] D. Dobkin, *The RF in Rfid Passive UHF Rfid in Practise*, 1st Ed., Elsevier, 2008.
- [11] C. Luo, I. Gil, R. Fernandez-Garcia, "Experimental comparison of three electro-textile interfaces for textile UHF-RFID tags on clothes", *AEU - International Journal of Electronics and Communications*, Vol. 146, March 2022, pp.154137.
- [12] C.D.M. Pena, D.B. Oliveira, E.J. Silva, M.W.B. Silva, "Ultra slim and small UHF RFID tag design for mounting on curved surfaces", *AEU - International Journal of Electronics and Communications*, Vol.128, January 2021, pp.153502.
- [13] E. Hardi, M. Veigt, M. Koerd, A.S. Herrmann, M. Freitag, "Use of RFID tags for monitoring resin flow and investigation of their influence on the mechanical properties of the composite", *Procedia Manufacturing*, Vol.24, 2018, pp.305-310.
- [14] J. Alarcon, R. Saba, M. Egels, P. Pannier, "A Flexible UHF RFID Tag for Harsh Environments". IEEE 2012 International Conference on RFID-Technologies and Applications (RFID-TA), France, 2012
- [15] A.S. Sarkin, "Esnek Yüzeylerdeki Baskı Devrelere Yarı İletken Uygulamaları", Thesis of MSc, Marmara University Institute of Pure and Applied Sciences, İstanbul, Türkiye, 2013.
- [16] L. Jian, A. Bing, Q. Jian, W. Yiping, "Nano Copper Conductive Ink for RFID Application". 2011 International Symposium on Advanced Packaging Materials (APM), China, 2011.
- [17] Ü. Sevinç, "Akrilonitril Bütadien (Nbr) Kauçuğunun ve Yapıştırıcısının Sentezi, Karakterizasyonu ve Demir Metaline Yapıştırılması", Thesis of MSc, Bursa Uludağ University Institute of Pure and Applied Sciences, Bursa, Türkiye, 2019.

BIOGRAPHIES

NAZMİ EKREN Esenler, İstanbul, in 1966. He received the B.S. and M.S. degrees in electrical engineering from the Yıldız Technical University, İstanbul, in 1983 and from Marmara University, İstanbul, in 1992. He received a Ph.D. degree from İstanbul University, İstanbul, Türkiye, in 1996.



He has been an Associate Professor with the Electrical and Electronics Engineering Department, Faculty of Technology at Marmara University. He is the author of more than 70 articles, and almost 100 proceedings. His research interests include electrical energy, lighting technology, renewable energy, biomaterials, and nanomaterials.

ALİ SAMET SARKIN Kadıköy, İstanbul, in 1987. He received the B.S. and M.S. degrees in electric education from the Marmara University, İstanbul, in 2010 and 2013, respectively. He received a Ph.D. degree in electrical and electronics engineering from Marmara University, İstanbul, Türkiye, in 2022.



Since 2014, he has been working as a Lecturer with the Osmaniye Korkut Ata University. His research interests include identification systems including RFID, renewable energy, photovoltaic systems, and their applications.

Design Consideration for Active–Only Memcapacitor Emulator Circuits

Hacer Atar Yildiz and Osman Aydin

Abstract—In this paper, a simple memcapacitor emulator circuit using only active elements is presented. Instead of using any bulky passive components, the proposed circuit makes use of the intrinsic capacitors of MOSFETs. As a result, the circuit took up significantly less area on the IC environment. In addition, a modification technique is proposed to extend the operating frequency range of the emulator, which might broaden the circuit's application possibilities. Considering the basic non-idealities of the circuit, a more realistic formulation of the memcapacitance value is derived. Detailed simulations utilizing the 0.18 μm CMOS Cadence design tool are used to validate all theoretical aspects as well as the circuit's appropriate functionality.

Index Terms—Memcapacitor emulator circuit, pinched hysteresis loop, CCCII, electronically adjustable circuit

I. INTRODUCTION

MEMRISTOR was first postulated by L. O. Chua as the missing circuit component in 1971 [1], is shown to be physically realizable by HP in 2008 [2]. Later, with the introduction of new class of mem-elements, memcapacitor and meminductor by Chua [3], the application of these elements has been expanded further for instance in adaptive filters [4], oscillators [5], chaotic circuits [6-7], neuromorphic circuits and non-volatile memories [8-10].

As the memristive elements are characterized by the well-known pinched, or zero-crossing hysteresis twist in current-voltage space, the fingerprint of the memcapacitor is the pinched loop in the voltage-charge space [11]. In some implementation, memcapacitor behavior can be obtained using mutator which converts memristor behavior to memcapacitor [12-13]. These circuits employ perfect memristor to implement the memcapacitor, which may be problematic in

case the limitations of the involved memristor are not well understood or its optimized design is not available. The synthesis of memcapacitor using conventional building blocks, i.e. integrators, summing amplifiers, on the other hand, is a more prudent approach as the optimized basic building blocks are easily accessible [14-17].

At this point, it should be noted a well-known disadvantage of all these mem-element emulators for IC integration point of view, which stems from the implementation of the involved passive elements, suffer from large chip area [18-20]. Active-only circuit design is a very useful technique to address this drawback. In these circuits, passive components are realized using the intrinsic resistors and capacitors of the involved active components [21]. While this approach substantially reduce the chip area occupied by the circuits, the circuit parameters can be adjusted electronically.


In this study an area efficient implementation of an integrator-based memcapacitor emulator (MC) is introduced. In order to have an area efficient solution, a design containing only active components is proposed. The benefits of the circuit are revealed by comparing it with similar circuits presented in the literature. Thanks to the evaluations of the main non-idealities of the proposed circuit, theoretical considerations which may help to assess the limitations of the proposed circuit are presented. Simulation results of the circuit using UMC 0.18 μm CMOS process are presented to illustrate the circuit performance.

In this paper, first the proposed realization of the memcapacitor emulator is presented. Then, the active element nonidealities are modeled and theoretical expressions are derived, which may be useful to assess the circuit important limitations. The operating frequency range of the proposed emulator is considered to be limited, which is partially mitigated by applying a modification technique. Simulation results justifying the proper operation of both the initial and the modified circuits are presented. Finally, a section with extensive comparisons of the suggested circuits to their counterparts is provided.


II. PROPOSED MEMCAPACITOR EMULATOR AND ITS PRELIMINARY DESIGN CONSIDERATIONS

The proposed memcapacitor, consisting of two active integrators built around a second generation current conveyor (CCII), a current controlled current conveyor (CCCII) and two voltage buffers is given in Fig.1. Note that CCCII is actually a

HACER ATAR YILDIZ, is with Department of Electronics and Communication Engineering University of Istanbul Technical University, Istanbul, Turkey, (e-mail: haceryildiz@itu.edu.tr).

 <https://orcid.org/0000-0003-4490-6878>

OMER AYDIN, is with Department of Electronics and Communication Engineering University of Istanbul Technical University, Istanbul, Turkey, (e-mail: oaydin@itu.edu.tr).

 <https://orcid.org/0000-0002-1519-6937>

Manuscript received April 25, 2022; accepted July 17, 2022.

DOI: [10.17694/bajece.1108680](https://doi.org/10.17694/bajece.1108680)

modified CCII, in which the intrinsic x -terminal resistance is included in its defining equation.

This resistance is given by:

$$r_x = \left(\mu_n C_{ox} \frac{W}{L} I_B \right)^{-1/2} \quad (1)$$

where μ_n, C_{ox} are respectively the electron mobility and gate capacitor, while W/L is the transistor dimensions. The fact that this resistance can be controlled via the biasing current I_B is very beneficial from IC integration point of view.

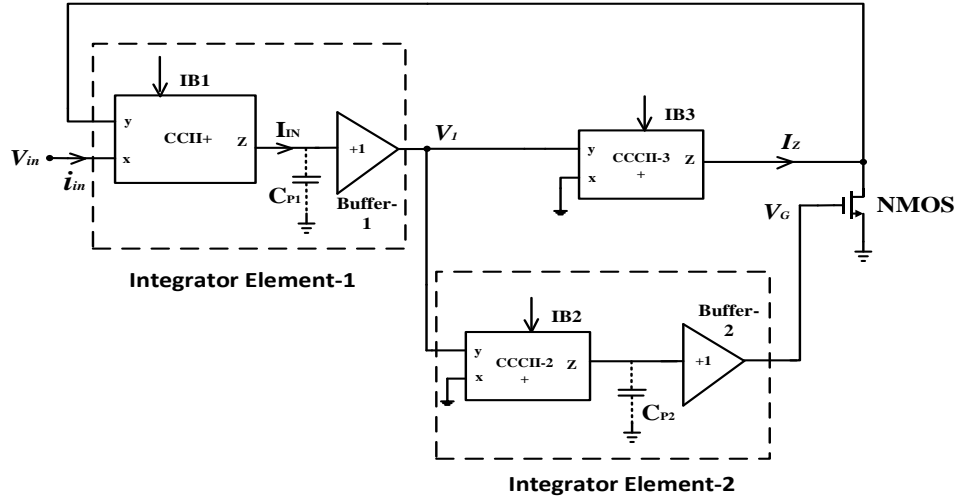


Fig.1. The general topology of proposed MC topology

From Fig. 1, capacitors, C_{P1} and C_{P2} shown within the dashed lines are intrinsic capacitors which appear at the input port of the voltage buffers. While designing the circuit, transistors' dimensions at the input section of the voltage buffer are chosen big enough that these intrinsic capacitors become large and dominate other parasitic capacitors in the circuit. Moreover, in case the voltage buffers are implemented such that properly, these capacitors mainly stem from the gate capacitance of MOS transistors, these capacitors can be considered linear [22]. Given the formal defining equations of CCCII and CCII, respectively [23],

$$\begin{bmatrix} I_y \\ V_x \\ I_z \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 1 & r_x & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} V_y \\ I_x \\ V_z \end{bmatrix}, \quad \begin{bmatrix} I_y \\ V_x \\ I_z \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} V_y \\ I_x \\ V_z \end{bmatrix} \quad (2)$$

and considering the fact that the input voltage V_{in} is small enough to satisfy $(V_{DS} \ll V_{GS} - V_{TN})$, the equivalent memcapacitance value of the proposed MC can be obtained as follows:

$$V_1 = \frac{1}{C_{P1}} \int I_{in}(t) dt = \frac{q_{in}}{C_{P1}} \quad (3)$$

$$\begin{aligned} V_G &= \frac{1}{C_{P2} r_{x2}} \int V_1(t) dt \\ &= \frac{1}{C_{P2} r_{x2}} \int \frac{q_{in}(t)}{C_{P1}} dt \\ &= \frac{g_{m2}}{C_{P1} C_{P2}} \int q_{in}(t) dt \end{aligned} \quad (4)$$

The required multiplication operation is realized by the current-voltage relationship of the NMOS transistor arranged to operate in the triode region:

$$I_Z = \underbrace{\mu_n C_{ox} \left(\frac{W}{L} \right)}_{\beta} (V_{GS} - V_{TN}) V_{DS}, \quad I_Z = \frac{V_1}{r_{x3}} \quad (5)$$

By using (3) and (4) in (5), the equation is obtained as follows:

$$\frac{q_{in}}{r_{x3} C_{P1}} = \beta \left(\frac{g_{m2}}{C_{P1} C_{P2}} \int q_{in}(t) dt - V_{TN} \right) V_{in} \quad (6)$$

The Eq.6 is rearranged as follows:

$$C_M = \frac{q_{in}}{V_{in}} = \frac{\beta C_{P1}}{g_{m3}} \left(\frac{g_{m2}}{C_{P1} C_{P2}} \int q_{in}(t) dt - V_{TN} \right) \quad (7)$$

By considering the charge-voltage relation of the capacitor element $q(t) = C v(t)$, and rearranging the equation, the value of memcapacitor becomes as follows:

$$C_M = \frac{\beta C_{P1}}{g_{m3}} \left(\frac{g_{m2}}{C_{P1} C_{P2}} \int q_{in}(t) dt - V_{TN} \right) \quad (8)$$

$$C_M = \frac{\beta g_{m2}}{g_{m3} C_{P2}} \int q_{in}(t) dt - \frac{\beta_1 C_{P1}}{g_{m3}} V_{TN} \quad (9)$$

From Eq. 9, the value of the memcapacitance consists of two terms, one being time-dependent and the other is constant. Both of the terms can be electronically adjusted by changing g_{m2} and g_{m3} parameters, which may be beneficial in practical applications of the memcapacitance.

In Table I, aspect ratios of the MOS transistors used in active devices are given.

TABLE I
ASPECT RATIOS (in μm) OF THE TRANSISTORS IN FIG.2

MOS transistor	CCCII and CCII
M1, M2	4/0.8
M3, M4	10/0.8
M5	2/0.8
M6	5/0.8
M7	1.5/0.8
M8	4/0.8
M _B	35/0.8

The CCCII active element is depicted in Fig.2a and realized by using conventional translinear topology. To implement the CCII, input transistor with a very low r_x value is selected in accordance with Eq. 1. Voltage buffer is also shown in Fig.2b.

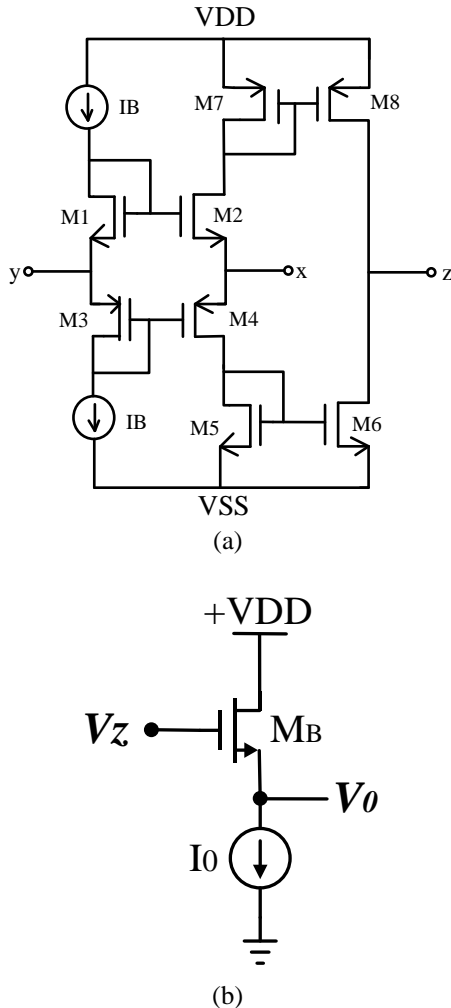


Fig.2. Implementations of (a) the current conveyors, (b) the voltage buffer

III. NON-IDEAL ANALYSIS OF THE MEMCAPACITOR EMULATOR

Non-ideal model of the proposed memcapacitor is shown in Fig.3. Considering the well-known non-idealities of the active

elements, memcapacitor defining equation is obtained as follows [24]:

$$V_1 = [\alpha_1(s)\beta_1(s)\gamma_1(s) \frac{1}{\left(\frac{1}{r_{z1}} + sC_{P1}\right)}] I_{in} \quad (10)$$

In this expression, $\alpha_1(s)$ and $\beta_1(s)$ are respectively, the non-ideal current and voltage gains of CCII, $\gamma_1(s)$ is the non-ideal gain of the buffer, while r_{z1} is the z -terminal parasitic resistance.

Extensive simulation results of the circuit in Fig. 2a shows that frequency dependency of $\beta_1(s)$ and $\gamma_1(s)$ can be neglected, while $\alpha_1(s)$ can be represented with the following simple expression:

$$\alpha_1(s) = \alpha_0 \frac{\omega_z}{s + \omega_z} \quad (11)$$

For practical applications, the output resistance of the CCII satisfy the following: ($1/r_{z1} \ll |j\omega C_{P1}|$), thus Eq.10 can be approximated as:

$$V_1 \cong \frac{\alpha_1(s) I_{in}}{C_{P1} s} = \frac{\alpha_1(s)}{C_{P1}} q_{in} \quad (12)$$

Considering the non-idealities of integrator-2, the equations are obtained as follows:

$$V_G = [\alpha_2(s)\beta_2(s)\gamma_2(s)g_{m2} \frac{1}{\left(sC_{P2} + \frac{1}{r_{z2}}\right)}] V_1 \quad (13)$$

Assuming that ($1/r_{z2} \ll |j\omega C_{P2}|$) and $\beta_2(s)$ and $\gamma_2(s)$ are constant, Eq.13 can be obtained as:

$$V_G \cong \left[\alpha_2(s)g_{m2} \frac{1}{C_{P2}} \right] \frac{V_1}{s} \quad (14)$$

Using (12) into (14) yields:

$$V_G = \left[\alpha_1(s)\alpha_2(s)g_{m2} \frac{1}{C_{P1}C_{P2}} \right] \frac{q_{in}}{s} \quad (15)$$

If the current-voltage relation of the NMOS transistor operating in the linear region is written and the Eq.15 is substituted in this relation, the memcapacitance relation is obtained as follows:

$$I_z - \frac{V_{in}}{r_{z3}} = \beta(V_{GS} - V_{TN})V_{DS}, \quad I_z = \frac{V_1}{r_{x3}} \quad (16)$$

$$\begin{aligned} \frac{V_1}{r_{x3}} - \left[\frac{V_{in}}{r_{z3}} + \frac{V_{in}}{r_{Y1}} + V_{in}(sC_{Y1}) \right] \\ = \beta[(\alpha_1(s)\alpha_2(s)g_{m2} \frac{1}{C_{P1}C_{P2}}) \frac{q_{in}}{s} \\ - V_{TN}]V_{in} \end{aligned} \quad (17)$$

Rearranging this equation, memcapacitance value can be obtained as:

$$\frac{\alpha(s)}{C_{P1}r_{x3}} q_{in} = \left[\beta\alpha^2(s)g_{m2} \frac{1}{C_{P1}C_{P2}} \frac{q_{in}}{s} - \beta V_{TN} + \frac{1}{r_{z3}} + sC_{Y1} \right] V_{in},$$

$$\alpha_1(s) = \alpha_2(s) = \alpha(s)$$

$$C_M = \frac{q_{in}}{V_{in}} = \underbrace{\frac{\beta\alpha(s)g_{m2}}{g_{m3}C_{P2}} \int q_{in}(t)dt}_{\text{Ideal memcapacitor}} - \frac{\beta C_{P1}}{\alpha(s)g_{m3}} V_{TN}$$

$$+ \frac{C_{P1}}{\alpha(s)g_{m3}} \left(\frac{1}{r_{z3}} + \frac{1}{r_{Y1}} + sC_{Y1} \right)$$

As it can be seen from Eq.19, in the non-ideal case, the parasitics C_{Y1}, r_{Y1} and r_{z3} appear in parallel with the memcapacitor. In case, the circuit driving this memcapacitor has a lower output impedance, these effects can be neglected. Furthermore, the frequency dependency of the current gain, $\alpha(s)$ may worsen the memory behavior of the element. Therefore, the circuit operating frequency range should be kept much smaller than the pole in the model of Eq. (19).

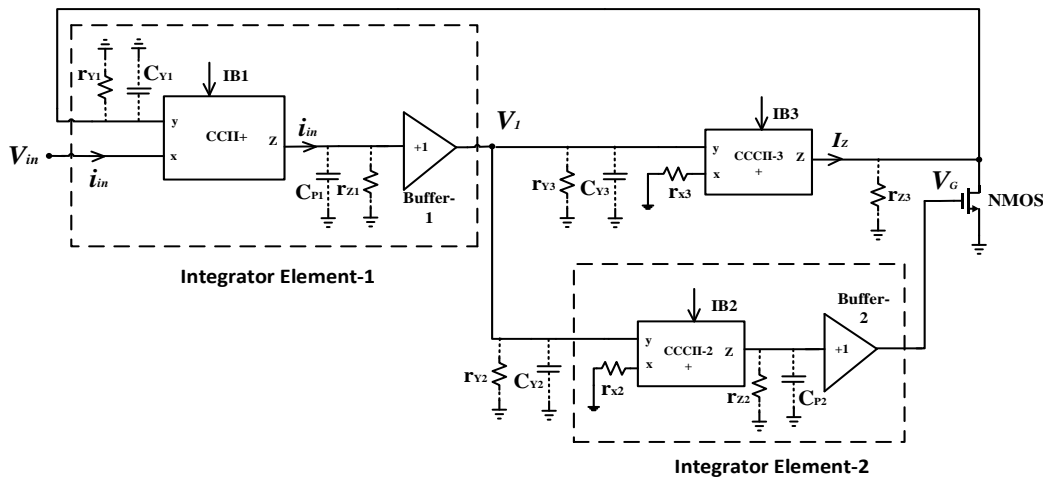


Fig.3. Non-ideal model of proposed memcapacitor

IV. SIMULATION RESULTS

The simulation results of the proposed circuit are produced by using UMC 0.18 μm technology parameters in Cadence design environment. The active elements, CCII, CCCII and voltage buffers are supplied with $\pm 1\text{V}$. Since the input current on the parasitic capacitor cannot be sensed, the hysteresis loop is simulated depending on the input voltage (V_{in}) and $V_1(q_{in}/C_{p1})$ which is the first integrator output.

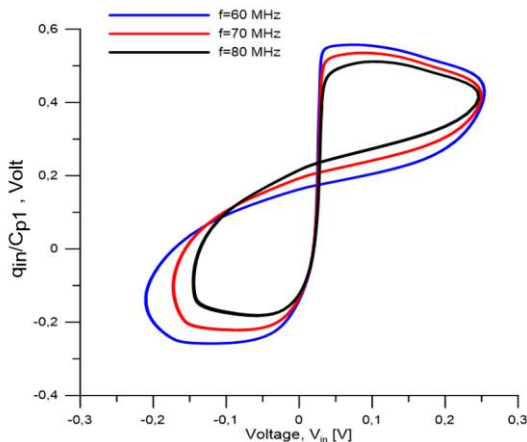


Fig.4. The Voltage-charge characteristic of proposed memcapacitor.

An external sinusoidal signal with an amplitude of 200 mV is applied to the circuit to obtain the emulator characteristic. As shown in Fig. 4, the circuit generates the pinched hysteresis characteristic which is the fingerprint of mem-behavior. The sample characteristics are given at three different frequencies, at 60 MHz, 70 MHz and 80 MHz.

However, the simulation results for different biasing currents are also obtained in Fig.5. From these results, it is seen that the circuit is capable of generating the memristor behavior over a wide range of the biasing current

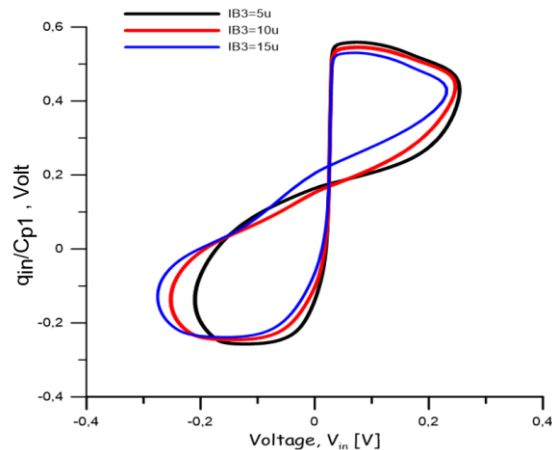


Fig.5. Pinched hysteresis loops for different biasing currents

The change in the circuit performance against environmental temperature variations is also studied. The circuit dynamics at three different temperatures are obtained at 60 MHz and the results are given in Fig. 6. From these results, it is clear that the circuit is able to generate the mem-behavior over a wide range of temperature. In addition, simulation results at different processing corners are obtained at 60 MHz and are given in Fig.7. From these results, the circuit works properly against the process changes.

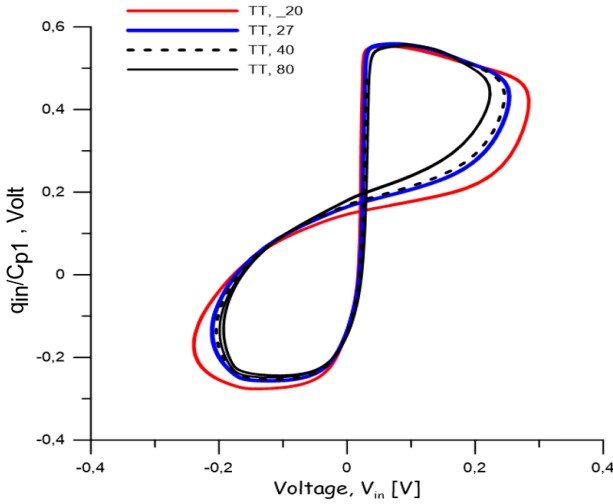


Fig.6. Memcapacitive characteristics at three different temperatures

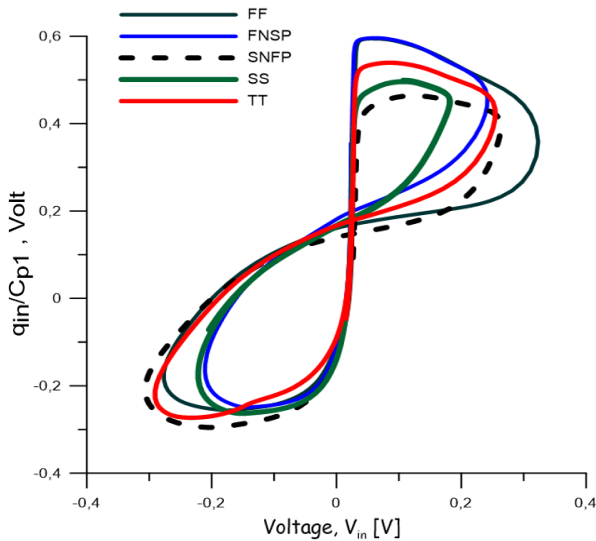


Fig.7. Memcapacitive characteristics for the process corners

In order to reveal the dependency of the memcapacitance to the variation of the input signal, we have obtained the simulation result in Fig.8. The memcapacitance depends on the input signal frequency, which is resulted from the nonlinearity of the memcapacitor.

Since intrinsic capacity is used instead of discrete capacity in this study, the capacity value varies between 100-500 fF. The rate of change is in the same order as circuits using discrete capacity [14].

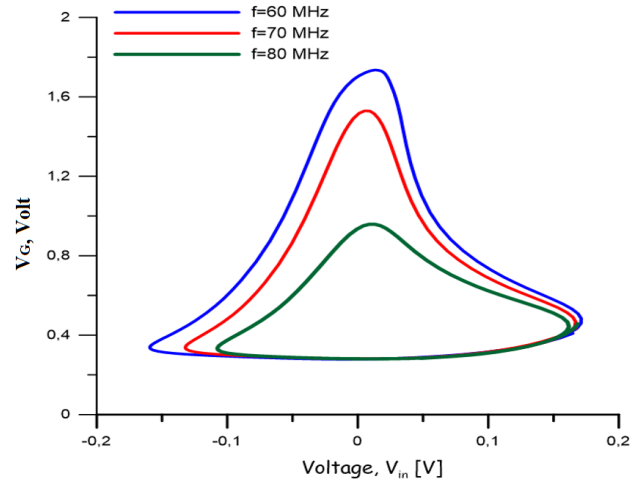


Fig.8. V_G versus V_{in} for different operating frequencies

V. MC MODIFIED FOR WIDE OPERATING FREQUENCY RANGE

The proposed memcapacitor has an operating frequency range proportional to g_m/C_p . As we use an intrinsic capacitor with a small range of values to implement C_p , the operating frequency range of the circuit varies within a narrow range. To remedy this issue, capacitance multiplier is proposed in the literature [25]. In this section, we applied this technique to the proposed emulator, in order to expand its operation frequency range, towards lower frequency region.

A conceptual capacitance multiplier built around CCII- is given in Fig.9. CCII- refers to the negative output current conveyor and the constitutive equation is given as:

$$i_y = 0, \quad V_x = V_y, \quad i_z = -K_C i_x \quad (20)$$

where K_C is the current gain from the x to z terminal currents of the current conveyors and equals 1 for conventional current conveyors. The current gain mentioned in this study should be taken as greater than 1. In this case, the equivalent input capacitance can be calculated using the following equation:

$$C_{eq} = K_C C_p \quad (21)$$

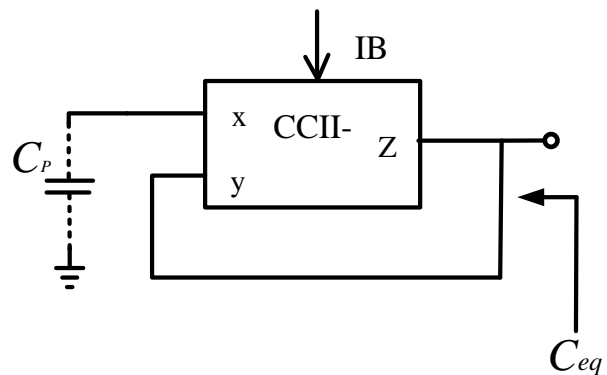


Fig.9. Capacitance multiplier built around a CCII [25]

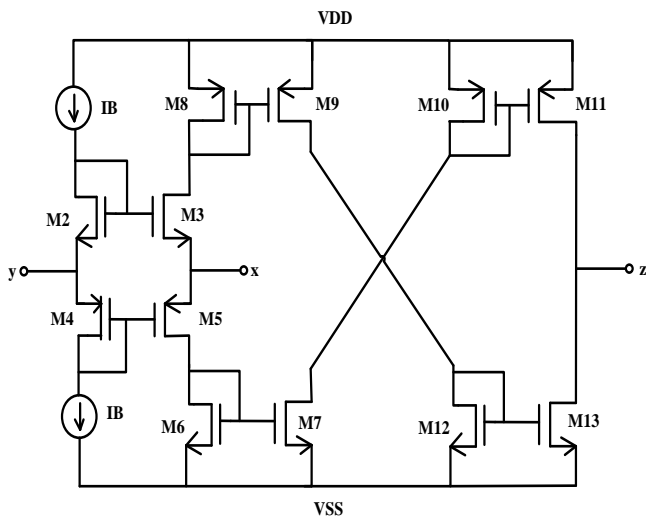


Fig.10. CMOS realization of the CCII- [26]

Therefore, for $K_C > 1$, the equivalent capacity will be greater than the C_p capacity. The K_C value can be set to the desired value by appropriately arranging the size ratios of the M10-M11 and M12-M13 MOS current mirror transistors in the translinear current conveyor which is shown in Figure 10.

By applying this technique to the proposed circuit, we obtain the modified circuit in Fig. 11. Assuming that the input terminal voltages of the current conveyors are equal, the memcapacitance of the circuit in Fig. 11 is as follows:

$$C_M = \frac{\beta g_{m2}}{g_{m3} K_C C_{P2}} \int q_{in}(t) dt - \frac{\beta_1 C_{P1}}{g_{m3}} V_{TN} \quad (22)$$

To verify the feasibility of the approach, simulation results are obtained for the 80 MHz and 100 mV input sinusoid of the modified circuit. The observed pinched hysteresis loops obtained for $K_C = 10$ and $K_C = 1$ are shown in Fig.12.

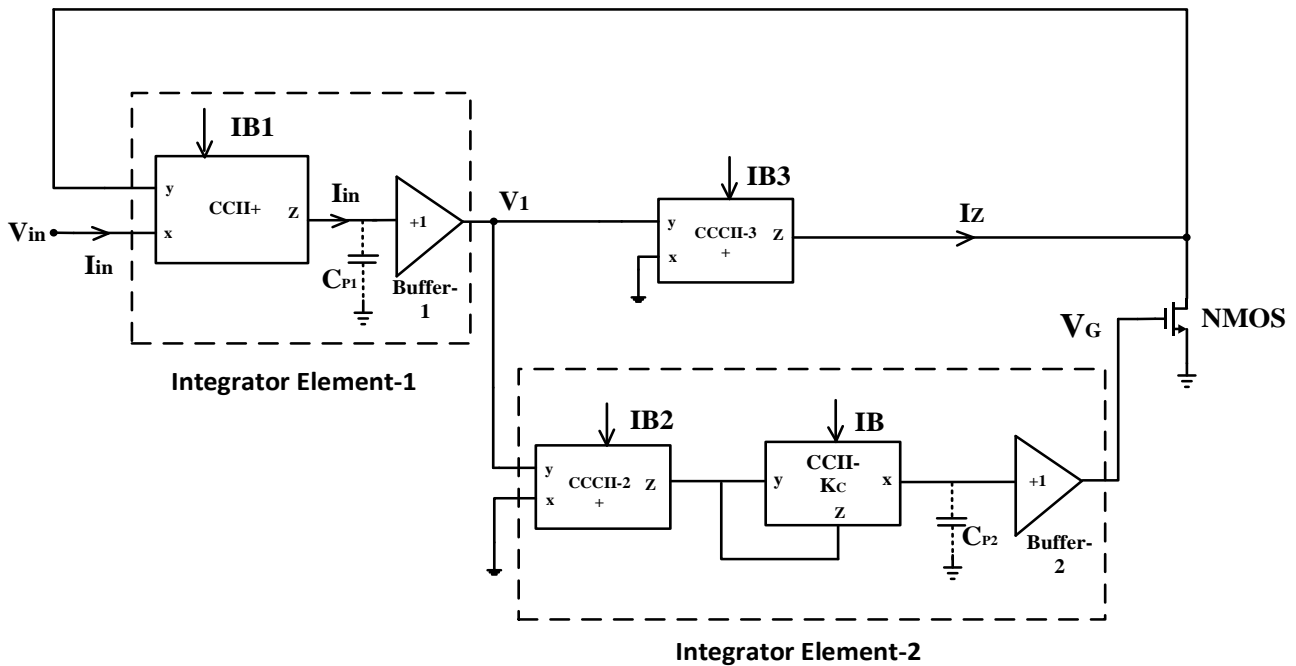


Fig.11. The memcapacitor emulator modified for low operating frequency

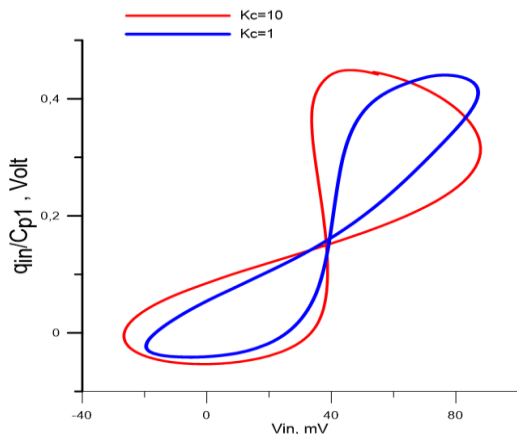


Fig.12. Voltage-current response of the modified MC for $K_C = 1$ and $K_C = 10$

VI. COMPARISON OF THE PROPOSED MC WITH EXISTING CIRCUITS

Some figures which can be useful to compare the proposed circuit compared with its existing counterparts are presented in Table II. All existing counterpart's studies are used passive components, so they are not area efficient design. To the best of the authors knowledge, there is not any other active-only memcapacitor emulator in the literature. It provides a big advantage in terms of the IC design environment, as it takes up less space as there are no discrete components.

TABLE II
PERFORMANCE COMPARISON WITH OTHER STUDIES

References	Number of active components	Operating frequency	Power supply	Number of passive elements	Elec. Tunability	Floating & grounded	Transistor Parameter
[12]	1 CBTA, 1 Memristor	250kHz	NA	1C	Yes	Floating	TSMC 0.18 μm
[13]	5 Opamp, 1 Multiplier	200Hz	$\pm 15\text{V}$	11R, 2C 1Diode	No	Grounded	Discrete components
[14]	2 CCII, 1 Multiplier	48Hz	$\pm 20\text{V}$	3C, 2R	Yes	Grounded	Discrete components
[15]	1 DVCCTA	900kHz	$\pm 0.9\text{V}$	2C, 1R	No	Grounded	TSMC 0.18 μm
[16]	2 OTA, 1 multiplier	10Hz	$\pm 1.25\text{V}$	2C, 2R	Yes	Grounded	TSMC 0.18 μm
[17]	1 DXCCDITA	1MHz	$\pm 1.25\text{V}$	2R, 1C	Yes	Floating	CMOS technology 0.18 μm
[18]	3 Opamp, 2 CFOA	10Hz	NA	3R, 2C, 1LDR	No	Grounded	Discrete components
[19]	2 AD844, 1 Memristor	10Hz	NA	1C, 1R	No	Grounded	NA (simulation)
[20]	1 Multiplier, 4 Opamps, 1CCCS	10Hz	$\pm 10\text{V}$	3C, 2R	No	Grounded	NA (simulation)
Proposed work	2 CCCII, 1CCII and 2 buffers	80MHz	$\pm 1\text{V}$	No passive component	Yes	Grounded	UMC 0.18 μm

As can be seen from Table II, the proposed circuit can operate at much higher frequencies and lower power supply rates than existing circuits. The proposed memcapacitor can be adjusted electronically in this study, as in some studies in Table II. The floating solution offers more flexible usage possibilities, but the grounded solutions are used more in practice because they have a simpler structure.

VII. CONCLUSION

In this study, an integrator-based memcapacitor emulator is presented. The important parameters of the proposed circuit can be adjusted electronically to the desired value. Compared to previous studies presented in the literature, the proposed memcapacitor has a structure consisting of only active building blocks. Therefore, the chip area will be smaller compared to their counterparts. The viability of the circuit is verified via detailed simulation results. Furthermore, a detailed nonideality analysis which reveals the main limitation of the circuit are presented. Finally, a modification technique which can be applied to extend the operating frequency range of the circuit is presented.

REFERENCES

- [1] L. O. Chua, "Memristor—The Missing Circuit Element," *IEEE Trans. Circuit Theory*, vol. 18, no. 5, pp. 507–519, 1971.
- [2] D. B. Strukov, G. S. Snider, D. R. Stewart, and R. S. Williams, "The missing memristor found," *Nature*, vol. 453, no. 7191, pp. 80–83, 2008.
- [3] M. Di Ventra, Y. V. Pershin, and L. O. Chua, "Circuit elements with memory: Memristors, memcapacitors, and meminductors," *Proc. IEEE*, vol. 97, no. 10, pp. 1717–1724, 2009.
- [4] Driscoll T, Quinin J, Klein S, Kim, HT, Kim BJ, Pershin YV, Ventra MD & Basov DN (2010), Memristive adaptive filters. *Applied Physics Letter* 97, pp. 1-3.
- [5] D. Yu, Z. Zhou, H. H. C. Iu, T. Fernando, and Y. H. Hu, "A coupled memcapacitor emulator-based relaxation oscillator," *IEEE Trans. Circuits Syst. II Express Briefs*, vol. 63, no. 12, pp. 1101–1105, 2016.
- [6] Muthuswamy B & Kokate PP, Memristor-Based Chaotic Circuits. *IETE Technical Review* 26, pp. 417-429, 2009.
- [7] Wang G, Zang S, Wang X, Yuan F & Ching Lu HH, Memcapacitor model and its application in chaotic oscillator with memristor, *Chaos* 27, pp. 013110-1-12, 2017.
- [8] M. S. Feali, A. Ahmadi, and M. Hayati, "Implementation of adaptive neuron based on memristor and memcapacitor emulators," *Neurocomputing*, vol. 309, pp. 157–167, Oct. 2018.
- [9] A. A. M. Emara, M. M. Aboudina, and H. A. H. Fahmy, "Non-volatile low-power crossbar memcapacitor-based memory," *Microelectronics J.*, vol. 64, pp. 39–44, Jun. 2017.
- [10] Q. Zhao, C. Wang, and X. Zhang, "A universal emulator for memristor, memcapacitor, and meminductor and its chaotic circuit," *Chaos*, vol. 29, no. 1, 2019.
- [11] Z. Yin, H. Tian, G. Chen, and L. O. Chua, "What are memristor, memcapacitor, and meminductor?" *IEEE Trans. Circuits Syst. II Express Briefs*, vol. 62, no. 4, pp. 402–406, 2015.
- [12] Çam Taşkıran Z. G., Sağbaş M., Ayten U. E., Sedef H., "A new universal mutator circuit for memcapacitor and meminductor elements", *International Journal of Electronics (AEU)*, vol.119, DOI: 10.1016/j.aeu.2020.153180, 2020.
- [13] Yuan F, Li Y, Wang G, Dou G, Chen G. Complex dynamics in a memcapacitor-based circuit. *Entropy*;21(2):188, 2019.
- [14] Yesil A., Babacan Y., "Electronically Controllable Memcapacitor Circuit with Experimental Results", *IEEE Transactions on Circuits and Systems II: Express Briefs*, DOI: 10.1109/TCSII.2020.3030114, 2020.
- [15] Vista J. and Ranjan A., "Design of Memcapacitor Emulator using DVCCTA", *International Conference on Applied Physics, Power and Material Science* 5–6 December 2018.
- [16] M. Konal and F. Kacar, "Electronically Tunable Memcapacitor Emulator Based on Operational Transconductance Amplifiers," *J. Circuits, Syst. Comput.*, pp. 1–8, 2020.
- [17] J. Vista and A. Ranjan, "Simple charge controlled floating memcapacitor emulator using DXCCDITA," *Analog Integr. Circuits Signal Process.*, vol. 104, no. 1, pp. 37–46, 2020.
- [18] Wang X, Fitch A, Iu H, Qi W. "Design of a memcapacitor emulator based on a memristor", *Physics Letters A*;376(4):394–399, 2012.
- [19] XiaoYuan Wang, Herbert H. C. Iu, GuangYi Wang, Wei Liu, "Study on Time Domain Characteristics of Memristive RLC Series Circuits" *Circuits Syst Signal Process*, 35:4129–4138, 2016.
- [20] M.E. Fouda and A.G. Radwan, "Charge controlled memristor-less memcapacitor emulator", *Electronics Letters*, vol. 48, pp. 23, 2012.
- [21] Arslan E, Metin B, Cicekoglu O, "MOSFET-only multi-function biquad filter," *AEU-International Journal of Electronics and Communication*, 69, 1737-1740, 2015.

- [22] HA. Yildiz, S. Ozoguz, A. Toker, On the Realization of MOS-Only Allpass Filters, *Circuits Syst. Signal Process.*, 32 1455–1465, 2013.
- [23] A. Fabre, O. Saaid, F. Wiest, C. Boucheron, High frequency applications based on a new current controlled conveyor, *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, 43(2), 82 – 91, 1996.
- [24] E. Yuce and O. Cicekoglu, The Effects of Non-Idealities and Current Limitations on the Simulated Inductances Employing Current Conveyors, *Analog Integrated Circuits and Signal Processing*, 46(2), 103-110, 2006.
- [25] G.A. Rincon-Mora, Active capacitor multiplier in Miller-compensated circuits, *IEEE J. Solid State Circ.* 35, 26–32, 2000.
- [26] C. Thoumazou, F.J. Lidgey, D.G. Haigh, *Analog Integrated Circuits: the current mode approach*, IEEE Circuit and System Series 2 (1993).

BIOGRAPHIES



HACER ATAR YILDIZ was born in Trabzon, Turkey. She received the B.Sc. and M.Sc. degrees in Electronics Engineering from the Faculty of Electrical and Electronics Eng., Karadeniz Technical University, Turkey in 1997 and 2000, respectively. She received her Ph.D. degree in 2015 in Electronics Engineering from Istanbul Technical University. Her research interests include analogue circuit design, active-only filters and analogue neural networks and VLSI circuit design.



OMER AYDIN has received his B.Sc. and M.Sc. in Electronics Engineering from Istanbul Technical University in 1982, 1985, respectively. He has completed his PhD on 4G and 5G radio power frequency amplifiers in 2016 in Istanbul Technical University. He has more than 20 scientific papers on 5G communication systems. His research interests include 5G communication systems, theoretical and practical aspects of radio frequency power amplifier designs.

The New HEMS Modelling of Human Heart

Fikret Yalcinkaya and Ertem Kizilkaplan


Abstract— The new version of the hydro-electro-mechanical system (HEMS) is modeled via 14 serially connected electrical equivalent circuits resulting in an integrated equivalent circuit. The new model accepts a group of variables and even examines the interaction between them. This paper introduces an improved integrated new model of the heart by replacing the monolithic equivalent structures with segmental comprehensive equivalents. Windkessel Model (WM) is a model of the relationships between aorta, aortic valve and left ventricle. Based on WM, the integrated new model was developed and simulated. The model's main focus is to define the dynamic properties of the system by a set of ordinary differential equations, and solving them using Ode23, a method for the solution of a closed-loop system. Using Matlab based Ode23 method; time-dependency of pressure, volume and flow were obtained. In case, short computation time and high accuracy are needed, then Ode23 is used. The model may be used to analyze complex processes in the heart and blood vessels. The new HEMS model has potential use for hemodynamic simulation of diseases, cardiovascular disorders, and special congenital heart diseases; such as ASD, VSD and PDA.

Index Terms— Time-dependent average pressure, biomedical engineering, medical control systems, medical simulation.


I. INTRODUCTION

TO OVERCOME critical and challenging heart conditions and to understand the functionality of the cardiovascular system (CVS), various models of human heart have been studied using physical models of the real and the artificial systems, based on the reference test domain [1]. The model of Windkessel is a famous example of such a discrete model [2,3]. Due to the complex nature of the human heart, models claiming to be sufficient standalone propose that the basic components and functions of the heart should separately be considered. Although computer analysis of numerical models has replaced physical models in many cases, electrical equivalent circuits and numerical analysis methods are now widely being used in

FIKRET YALCINKAYA, is with Department of Electrical and Electronics Engineering, Kirikkale University, Kirikkale, Turkey, (e-mail: fyalcinkaya@kku.edu.tr).

 <https://orcid.org/0000-0002-2174-918X>

ERTEM KIZILKAPLAN, is with Department of Electrical and Electronics Engineering, Kirikkale University, Kirikkale, Turkey, (e-mail: ertem.kizilkaplan@gmail.com).

 <https://orcid.org/0000-0003-2602-4804>

Manuscript received Jan 22, 2022; accepted July 22, 2022.

DOI: [10.17694/bajece.1061718](https://doi.org/10.17694/bajece.1061718)

addition to mathematical models in testing the control and stability of the cardiovascular system for the sole benefit of the heart and its functions [4]. Because the human cardiovascular system is an extremely nonlinear integrated structure in which electrical, mechanical and hydraulic components all work together in harmony. In this context, 0, 1, 2 or 3 dimensional models have been used in order to provide a better understanding of functionality and the simulation of blood flow in the human cardiovascular system and different calculation techniques have been developed using experimental methods [5]. Although these approaches simulate only certain sections of the cardiovascular system, they often reveal that the systems are quite difficult and unstable, and nearly unsolvable nonlinearity problems arise.

The difficulty of cardiovascular system modelling arises from its multifunctionality and compartmentalized nonlinear structure. In order to overcome this issue, due to the difficulties encountered in integrating modelling of the whole system, instead of modelling all the subsystems of the real system, it is preferred to model some parts of it independently [6], in order to reduce the problem complexity, and developing cardiovascular system new models can produce healthy hemodynamic data that mimic diseases to certain extent [7]. The hemodynamic measurements of a healthy person with a modified Windkessel model were used to analyze the hemodynamic data of congenital heart diseases, such as hypoplastic left heart syndrome (HLHS) [8].

Cardiovascular system modelling requires not only electrical simulation, the integrated electro-mechanical response of the heart, but also its two-dimensionality- hydraulic and pneumatic –as well, as that increases the complexity of the heart system; its complexity increases due to the hydraulic structure of the heart and the pneumatic pressure of the lung on the heart cycle. Modelling attempts integrating or cascading pneumatic effect to the hydro-electro-mechanical effect are a few in the literature. For this purpose, a new version of cardiovascular system model was attempted in this paper, and attention was paid to include more details than the HEMS model and to use quite a lot parameters [9].

The new HEMS model can evaluate and verify the effects of a group of variables at the same time and even make it possible to examine the interaction between more than one variable. Autocad based anatomical 3D drawings of all subsystems of the heart have been produced in order to get the best results specific to modelling and physiological requirements. 3D drawings and models are used to facilitate intermediate version development to achieve the integrated HEMS model. Because, in order to

integrate nonlinear complex biological structures and independent models, the structures described as interface and their models are needed [10].

In order to examine the hemodynamic of the cardiovascular system (CVS), first of all, the hydro-electro-mechanical system of the heart was modelled through 14 serially connected electrical equivalent circuits which demonstrates the integrated form of all subsystems developed. In the new HEMS model, a total of 28 equations (Eq. 4 – Eq. 31) were developed and used for the pressure and flow values at the outlets of all segments. In addition, pressure, volume and flow graphs for the whole circuit were obtained using the Matlab Ode23 method.

II. METHOD

The cardiovascular system requires the integration of four basic structures; these vital structures are the heart and triple vascular network. This network system is a network of arteries, veins, and capillaries. Blood circulates in the cardiovascular system through this network with the effect of pressure produced by the heart. Blood is a non-Newtonian fluid that completes its cycle through and by these structures. Events and processes occurring during the cardiac cycle are shown in The Wigger's diagram [11].

It is necessary to simulate the flow waveforms to describe the two different phases of the cardiac cycle, the systole and diastole. Under normal conditions, 60 to 100 impulses per minute originate from the sinus node (SA node) to trigger the heartbeat. The heartbeat time (Heart Period-HP) used in our model is 0.83 seconds (corresponds to 72 beats / minute). During this period, the systole time was 0.33 s and the diastole time was 0.50 s.

The time cycle in the heart begins with the activation of the heart muscle around the right atrium near the upper vena cava. The heart muscle surrounding the atriums and ventricles undergoes contraction and relaxation during a cycle. This contraction and relaxation in the heart is associated with intracellular calcium (Ca^{+2}) ion movement.

The activation function has been obtained in atrium over a heart cycle using MATLAB. The general idea is to discretize the given function, evaluating the function at different times separated by the time interval dt . dt is the time interval between two consecutive samples and it has been selected to be 0.01 s. The total number of samples for which the function evaluated depends on the heart period and the number of heart cycles represented. For dt -equals 0.01 s and period HP-equals 0.83 s then the number of samples is determined. This time interval seems ideal for the representation of both the flow and the pressure as these variables vary along the heart cycle. The total number of samples for which the activation function generated in the atria and ventricles will be evaluated depending on the heart period and the cardiac cycle time shown. A graphical representation of the activation function is given below for the first cardiac cycle. (Fig. 1)

The walls in the ventricles are characterized by time-varying elastic functions that relate instantaneous pressure and volume. These flexibility functions provide a smooth transition from a nonlinear diastolic pressure-volume curve (EDV) to a linear

systolic pressure-volume relationship (ESV). In our model, in order to represent the contraction and relaxation changes in each cardiac chamber, a time dependent sinus wave activation function was created for the ventricle and atria. The right and left ventricles are modelled as variable capacitors to calculate the relationship of pressure, volume and time, as well as viscous losses due to their characteristic flexibility and activation function. The time-dependent end-diastolic elastance (its hardness) of the right and left ventricles is shown in E_d , equation 1 with its general expression [12].

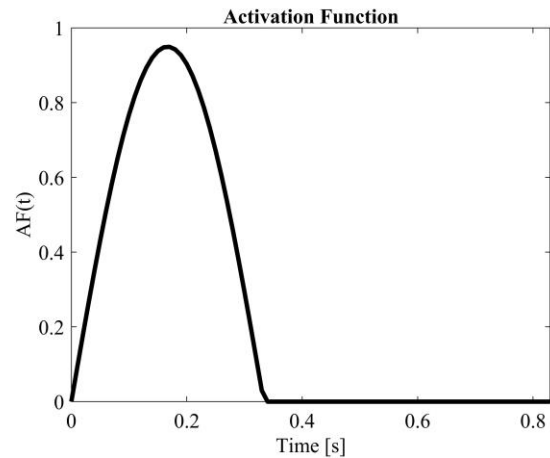


Fig. 1. Activation function for the first cardiac cycle

$$E_d(t) = AF(t) \cdot E_{DV} + [1 - AF(t)] \cdot E_F \quad (1)$$

According to this formula, $E_d(t)$ is determined as a time-varying parameter of flexibility according to the maximum and minimum values of the activation function. E_F refers to the harmonic factor created for the ventricles in this change, E_{DV} is the diastolic stiffness created for the right and left ventricles, and $AF(t)$ is the generated activation function.

The blood flow and its dynamics are used to study the hemodynamics of healthy and diseased blood vessels, vascular specific conditions, and the cardiovascular system. Hemodynamics can be defined as the physical principles that govern blood flow based on the fundamental laws of physics. Of these basic laws, Ohm's law can be expressed for the cardiovascular system as seen below. In addition, the relationship between the electrical, mechanical and hydraulic parameters is given as in TABLE I.

$$\Delta Q = \frac{\Delta P}{R} \quad (2)$$

TABLE I. ELECTRICAL, MECHANICAL AND HYDRAULIC PARAMETERS

	Symbol	Element	Unit
Electrical System	R	Resistance	ohm
	L	Inductance	s·ohm
	C	Capacitance	s/ohm
	U	Voltage	volt
	q	Charge	amper·s
	I	Current	amper

Mechanical System	B	Damping Rate	N·s/m
	M	Mass	kg
	K	Spring Rate	N/m
	F	Force	N
	x	Displacement	m
	v	Velocity	m/s
Hydraulic System	R	Blood Viscosity	mmHg·s/ml
	L	Blood Inertance	mmHg·s ² /ml
	C	Wall Compliance	ml/mmHg
	P	Pressure	mmHg
	V	Volume	ml
	Q	Blood Flow	ml/s

Windkessel modeling is used to represent the connection between hydraulic structure and electrical structure equivalent circuits. As the Windkessel model is the most accepted form of discrete modeling, it allows us to understand the relationship between blood pressure and blood flow in the aorta (Fig. 2 (a)- Fig. 2 (b)).

The similarity relationship between the electrical, mechanical and hydraulic parameters of the circuit elements to be created for modeling the cardiovascular system is shown in TABLE I, and with this analogy, parts of the system model can be combined with separate components to model the physical systems with interconnected components. According to this analogy; in case of electrical systems, these elements include resistors (R), capacitors (C), and inductors (L). For mechanical systems these include inertia (masses-M), springs (K) and shock absorbers (or friction elements-B). For hydraulic systems, these include flow resistance/viscosity (R), inertness (L), and chamber compliance (reservoir capacity) (C). Each segment of the circulatory cycle is modeled through a series of equations, describing the relationships, that give the correlations between the parameters of pressure (P, mmHg), volume (V, ml), and flow (Q, ml/s) associated with that specific segment. However, the desired average pressure and volume distribution determine the parameter values of compliances and resistances. For example, the time-dependent pressure and flow equations at the nodes that govern the aortic hydraulic components are obtained as follows:

$$\Delta P = L \frac{dQ}{dt} = \frac{\Delta V}{C} \tag{3}$$

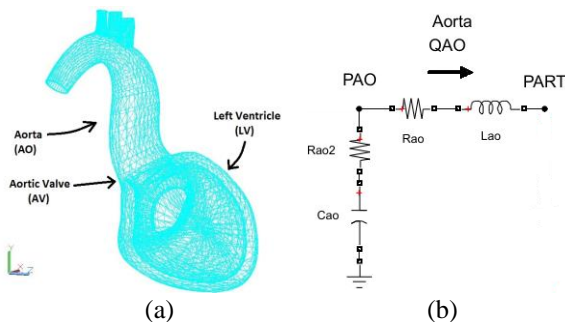


Fig. 2 (a) Autocad 3d representation of Aorta-Left Ventricle. (b) Hydraulic circuit of the Windkessel effect (model) between the aorta - left ventricle.

In equations (4) and (5), time dependent average pressure (P_{AO}) and flow (Q_{AO}) equations are obtained for the aortic part of the heart:

$$P_{AO}(t) = \frac{V_{AOI}(t) - V_{AOU}(t)}{C_{ao}} + R_{ao2} \cdot [Q_{LVI}(t) - Q_{AOI}(t)] \tag{4}$$

$$\frac{dQ_{AO}(t)}{dt} = \frac{P_{AO}(t) - P_{ART}(t) - [R_{ao} \cdot Q_{AOI}(t)]}{L_{ao}} \tag{5}$$

In equations (6) and (7), time-dependent average pressure (P_{ART}) and flow (Q_{ART}) equations are obtained for the heart arteries, where P_{CAPS} is time-dependent average pressure for the heart capillaries.

$$P_{ART}(t) = \frac{V_{ARTI}(t) - V_{ARTU}(t)}{C_{art}} \tag{6}$$

$$Q_{ART}(t) = \frac{P_{ART}(t) - P_{CAPS}(t)}{R_{art}} \tag{7}$$

The components of 3-D functional anatomy in which all the parts of the new version of hydro-electro-mechanical system developed are independently modeled and are given below. Representation sequencing of blood flow in the heart segments is drawn in 3-D forms by using Autocad of four compartments and other subsystems components, shown in Fig. 3.

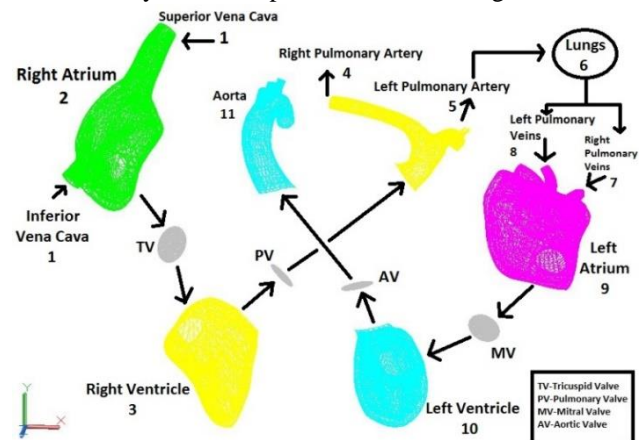


Fig. 3. Representation of blood flow in heart sub-segments drawn in 3D with Autocad

The analytically solvable hydraulic, electrical and mechanical models are designed to study the entire subsystem of specific organs (eg, the heart). Because of the multi-systems character of the human body total system and the multi-subsystems of each system make biological modeling extremely complex, as there are numerous variables that affect the functions, properties, and response of the circulatory system. Experimentally, it is impossible to include all known variables in a single system. However, by modeling the new HEMS model segmentally instead of monolithic approach; it also gives the opportunity to evaluate a group of variables at the same time and even to examine the interaction between the variables.

The partial description of dynamic models in the new HEMS is usually made with a set of differential equations using

hydraulic parameters. Different behaviours are observed in various parts of the closed loop. As a result, it is possible to evaluate a group of variables at the same time and even to examine the interaction between variables. The new version of the hydro-electro-mechanical system of human heart is modelled through integrated 14 series circuits of electrical equivalent of sub-segments of heart, the integrated form of all subsystems making up the total heart and its total functionality.

The total and integrated equivalent circuit of this new HEMS model, a fully equivalent closed-circuit, was developed from the cardiovascular system structure given in Fig. 4 and physiological values used for this model are given in Table II. The pressure and flow values of the time-dependent outlets, using the values in Table II, are given by the following equations; derived for different segments of our model, respectively.

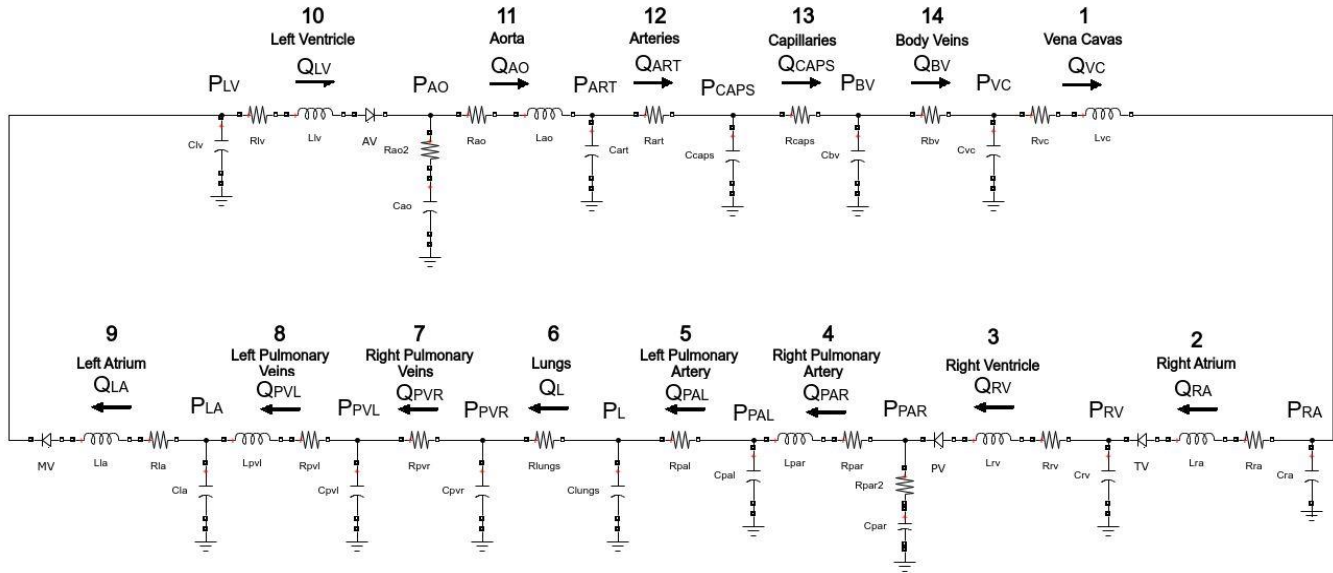


Fig. 4. Integrated Equivalent Circuit of the HEMS

In equations (8) and (9), time-dependent average pressure (P_{VC}) and flow (Q_{VC}) equations are obtained for the superior and inferior vena cava sections of the heart:

$$P_{VC}(t) = \frac{V_{VCI}(t) - V_{VCU}(t)}{C_{vc}} \quad (8)$$

$$\frac{dQ_{VC}(t)}{dt} = \frac{P_{VC}(t) - P_{RA}(t) - [R_{vc} \cdot Q_{VCI}(t)]}{L_{vc}} \quad (9)$$

In equations (10) and (11), time-dependent average pressure (P_{RA}) and end-diastolic flow (Q_{RAD}) equations are obtained for the right atrium of the heart:

$$P_{RA}(t) = \frac{V_{RAI}(t) - V_{RAU}(t)}{C_{ra}} \quad (10)$$

$$\frac{dQ_{RAD}(t)}{dt} = \frac{P_{RA}(t) - P_{RV}(t) - [R_{ra} \cdot Q_{RAI}(t)]}{L_{ra}} \quad (11)$$

In equations (12) and (13), time-dependent average pressure (P_{RV}) and end-diastolic flow (Q_{RVD}) equations are obtained for the right ventricular segment of the heart, where C_{rv} is the compliance of the right ventricle.

$$P_{RV}(t) = \frac{V_{RVI}(t) - V_{RVU}(t)}{C_{rv}} \quad (12)$$

$$\frac{dQ_{RVD}(t)}{dt} = \frac{P_{RV}(t) - P_{PAR}(t) - [R_{rv} \cdot Q_{RVI}(t)]}{L_{rv}} \quad (13)$$

In equations (14) and (15), time-dependent average pressure (P_{PAR}) and flow (Q_{PAR}) equations are obtained for the right pulmonary arteries:

$$P_{PAR}(t) = \frac{V_{PARI}(t) - V_{PARU}(t)}{C_{par}} + R_{par2} \cdot [Q_{RVI}(t) - Q_{PARI}(t)] \quad (14)$$

$$\frac{dQ_{PAR}(t)}{dt} = \frac{P_{PAR}(t) - P_{PAL}(t) - [R_{par} \cdot Q_{PARI}(t)]}{L_{par}} \quad (15)$$

In equations (16) and (17), time-dependent average pressure (P_{PAL}) and flow (Q_{PAL}) equations are obtained for the left pulmonary arteries:

$$P_{PAL}(t) = \frac{V_{PALI}(t) - V_{PALU}(t)}{C_{pal}} \quad (16)$$

$$Q_{PAL}(t) = \frac{P_{PAL}(t) - P_L(t)}{R_{pal}} \quad (17)$$

In equations (18) and (19), time-dependent average pressure (P_L) and flow (Q_L) equations are obtained for the lungs segment of the heart:

$$P_L(t) = \frac{V_{LI}(t) - V_{LU}(t)}{C_{lungs}} \quad (18)$$

$$Q_L(t) = \frac{P_L(t) - P_{PVR}(t)}{R_{lungs}} \quad (19)$$

In equations (20) and (21), time-dependent average pressure (P_{PVR}) and flow (Q_{PVR}) equations are obtained for the right pulmonary veins:

$$P_{PVR}(t) = \frac{V_{PVRl}(t) - V_{PVRu}(t)}{C_{pvr}} \quad (20)$$

$$Q_{PVR}(t) = \frac{P_{PVR}(t) - P_{PVL}(t)}{R_{pvr}} \quad (21)$$

In equations (22) and (23), time-dependent average pressure (P_{PVL}) and flow (Q_{PVL}) equations are obtained for the left pulmonary veins:

$$P_{PVL}(t) = \frac{V_{PVLl}(t) - V_{PVLu}(t)}{C_{pvl}} \quad (22)$$

$$Q_{PVL}(t) = \frac{P_{PVL}(t) - P_{LA}(t)}{R_{pvl}} \quad (23)$$

In equations (24) and (25), time-dependent average pressure (P_{LA}) and end-diastolic flow (Q_{LAD}) equations are obtained for the left atrium of the heart:

$$P_{LA}(t) = \frac{V_{LAl}(t) - V_{LAU}(t)}{C_{la}} \quad (24)$$

$$\frac{dQ_{LAD}(t)}{dt} = \frac{P_{LA}(t) - P_{LV}(t) - [R_{la} \cdot Q_{LAl}(t)]}{L_{la}} \quad (25)$$

In equations (26) and (27), time-dependent average pressure (P_{LV}) and end-diastolic flow (Q_{LVD}) equations are obtained for the left ventricular segment of the heart, where C_{lv} is the compliance of the left ventricle.

$$P_{LV}(t) = \frac{V_{LVl}(t) - V_{LVU}(t)}{C_{lv}} \quad (26)$$

$$\frac{dQ_{LVD}(t)}{dt} = \frac{P_{LV}(t) - P_{AO}(t) - [R_{lv} \cdot Q_{LVl}(t)]}{L_{lv}} \quad (27)$$

In equations (28) and (29), time-dependent average pressure (P_{CAPS}) and flow (Q_{CAPS}) equations are obtained for the capillaries:

$$P_{CAPS}(t) = \frac{V_{CAPSl}(t) - V_{CAPSu}(t)}{C_{caps}} \quad (28)$$

$$Q_{CAPS}(t) = \frac{P_{CAPS}(t) - P_{BV}(t)}{R_{caps}} \quad (29)$$

In equations (30) and (31), time-dependent average pressure (P_{BV}) and flow (Q_{BV}) equations are obtained for the body veins:

$$P_{BV}(t) = \frac{V_{BVl}(t) - V_{BVU}(t)}{C_{bv}} \quad (30)$$

$$Q_{BV}(t) = \frac{P_{BV}(t) - P_{VC}(t)}{R_{bv}} \quad (31)$$

TABLE II. PARAMETERS OF THE HEMS

	R [mmHg.s/ml]	L [mmHg.s ² /ml]	C [ml/mmHg]	P [mmHg]	Initial V _i [ml]	Unstressed V _u [ml]	Initial Q _i [ml/s]
Vena Cavas	14	1	0.0453	7.1	2375.8	1947	85
Right Atrium	4	2	0.0451	5.54	442.23	1949	0
Right Ventricle	6	1	Variable	7.5	154.93	11.1	5.3
Right Pulmonary Artery	11	1	0.00015	7.1	9.29	7.87	0
Right Pulmonary Artery 2	12	-	-	-	-	-	-
Left Pulmonary Artery	41	-	0.00032	6.98	26.36	23.42	-
Lungs	79	-	0.00273	6.5	234.16	210.5	-
Right Pulmonary Veins	31	-	0.0012	4.35	74.96	68	-
Left Pulmonary Veins	9	1	0.0011	3.52	75.16	70	31
Left Atrium	5	1	0.01117	4.35	866.97	814.4	0
Left Ventricle	5	2	Variable	3.9	136.91	10.1	0
Aorta	10	3	0.00017	63.3	49.49	35.15	4.3
Aorta 2	10	-	-	-	-	-	-
Arteries	163	-	0.00024	63	105.16	85	-
Capillaries	1002	-	0.001825	62	861.85	711	-
Body Veins	91	-	0.0212	12.5	1261.3	908	-

III. SIMULATION RESULTS

Pressure waveforms are signals that suggest significant anomalies in the human circulatory system and are used in the diagnosis of cardiovascular diseases, such as heart valve diseases and aneurysms. The left ventricle and aortic pressure-time graph, which is in conformance with the physiological waveform of the circulatory system, shows the time-dependent change of aortic pressure and left ventricular pressure, and as a result of our Matlab based model simulation given in Fig. 5a. This change in pressure difference is related to the resistance parameter of the aortic valve. This structural change was detected by our model simulation. In addition, the comparison of the reference parameters with the theoretical simulation of HEMS model is given in Table III.

The fill in pressures of the ventricles are important in the evaluation of mechanical functions. The fill in pressures of the ventricles due to the blood volume are generally related to the end-diastolic volume (EDV). However, it is also linked to the systolic and diastolic function of the heart. The volume values of left (V_{LV}) and right (V_{RV}) ventricles were calculated based on Matlab based simulation and their time depended graphical changes are given in Fig. 5b. As it can be seen, the stroke volume pumped out of each ventricle is about 60 ml/beat and corresponds to a cardiac output (CO) of 72 ml/s (4.32 l/min) at a heartbeat of 0.83 s. Fill in and stroke volumes are not equal for both ventricles. It is possible that one ventricle passes a little more blood than the other for a heartbeat. However, on average, the fill in and stroke volume must be equal for the left ventricle (LV) and the right ventricle (RV). Although these values were found to deviate slightly from the statistical measurements, the final systolic volume (ESV) and the final diastolic volume (EDV) on average were 150 ml. This means that the value of the ejection fraction $[(EDV - ESV) / EDV]$ is about 50%. This value means that the heart does not function by one half as much as it needs; but it discharges only 50% of the blood into the veins every time it contracts. Ejection fraction (E_F) is a parameter related to stroke volume (SV). Normally, the heart throws more than half of the blood in the ventricles into the veins each time it contracts. E_F shows how much blood is pumping back to the heart with each heartbeat [13].

For our model, Matlab based plots of the change occurring in the pressure and volume in the left ventricle during a cardiac cycle is shown in Fig. 5c. To assess ventricular performance, a pressure-volume cycle parallel to the normal cardiac cycle was developed. According to this cycle, left ventricular contraction begins at the end of the diastole. The contraction energy is first used to increase the intra-cavity pressure without changing the left intra-ventricular volume. At this stage, the aortic and mitral valves close. The pressure increase continues until the diastolic pressure of the aorta is reached. At this point, the aortic valves open. Myocardial fibers continue to shorten at this stage and intra-ventricular blood is thrown towards the aorta. At this point, intra-ventricular pressure remains constant, but the volume gradually decreases due to ejection. At the end of systole, left ventricular contraction reaches its peak, relaxation begins in myocardial fibers, and when intra-ventricular pressure falls below aortic diastolic pressure, aortic valves close, ejection ends. After that, ventricular relaxation gains speed and

intra-ventricular pressure decreases rapidly without any change in volume. When the ventricular pressure falls below the pressure of the left atrium, the mitral valve opens and the diastolic filling begins. At this stage, the left ventricular volume begins to increase, but a slight increase in pressure occurs, and at the end of this stage, the left ventricular pressure-volume ring is completed.

In Fig. 5c, 4 points A to D are related to ventricular beat generation. Mitral valve closure and isovolumic contraction at point (A), aortic valve opening and ejection phase at point (B), aortic valve closure and isovolumic relaxation at point (C) and ventricular filling occur with mitral valve opening at point (D). The end-diastolic volume (EDV) below indicates the inactive ventricle, and the end-systolic volume relationship (ESV) indicates the active phase of the ventricle.

The amount of blood to be pumped in a continuous cycle determines the amount of blood in the cycle entering the heart from the vena cava. Due to the increase in the amount of online blood entering the ventricles, the ventricles go under high stress due to the specialized muscular structure of the heart, and they contract strongly and increase the amount of blood pumped by the heart. This cycle follows the Frank-Starling Law, which states that only the amount of quantity of blood entering the heart can only be pumped out into the cycle. In this way, the heart can pump only the amount of blood it already possess [14].

The basic mechanism underlying the Frank-Starling law is proportional to the contractility levels in the heart muscle fibers. The factor that increases the end-diastolic volume (EDV) is caused by the increase in venous return to the heart; this increase expands the ventricles in volume, stretches the heart muscle fibers, increases stroke volume and consequently cardiac output (CO). In addition, the increase in sympathetic stimulation increases the contraction force of the ventricle, so the ventricles pump more blood in systole and the end systolic volume (ESV) decreases. In cases where sympathetic stimulation increases, for example due to the effects of exercise, the contraction force of the ventricles increases and more blood is pumped, consequently the amount of blood remaining in the ventricles decreases. In line with this basic physiological condition, Matlab based system simulations showing the temporal variation of left ($SLV=1/C_{LV}$) and right ($SRV=1/C_{RV}$) ventricular stiffness from heart components were created and the results are given in Fig. 5d.

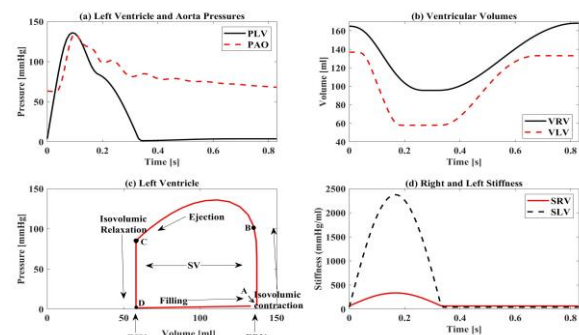


Fig. 5. (a) Simulated left ventricular pressure (P_{LV} -mmHg) and aortic pressure (P_{AO} -mmHg). (b) Simulated right ventricle volume (V_{RV} -ml) and left ventricle volume (V_{LV} -ml). (c) Simulated of pressure-volume change in the left ventricle. (d) S_{LV} : Left Ventricular Stiffness (mmHg / ml) and S_{RV} : Right Ventricular Stiffness (mmHg / ml).

TABLE III. COMPARISON BETWEEN NORMAL VALUES [15] AND THE NEW HEMS VALUES IN SOME PARTS OF THE HEART

Hemodynamic Parameters		
Parameter	Normal Range	HEMS Range
Cardiac Output (CO)	4-8 l/min	4.32 l/min
Stroke Volume (SV)	60-100 ml/beat	60 ml/beat
Ejection Fraction (EF)	40-60%	50%
Right Ventricular Volume (V_{RV})	100-160 ml	100-154.93 ml
Left Ventricular Volume (V_{LV})	60-140 ml	60-136.91 ml
Arterial Blood End-Diastolic Pressure (P_{ARTD})	60-90 mmHg	63 mmHg
Right Atrial End-Diastolic Pressure (P_{RAD})	2-6 mmHg	5.54 mmHg
Right Ventricular End-Diastolic Pressure (P_{RVD})	2-8 mmHg	7.5 mmHg
Left Atrial End-Diastolic Pressure (P_{LAD})	4-12 mmHg	4.35 mmHg

IV. CONCLUSIONS

In this paper, the main focus of the new HEMS modelling is to define the dynamic properties of the cardiovascular system with mathematical equations and the system behaviour based on a strong theoretical background. The dynamic properties of the cardiovascular system are determined by a set of ordinary differential equations. The system was solved using an ordinary differential equation method, known as Ode23, proper enough to be used in the solution of a closed-loop cardiovascular system until it converges. In addition, the whole circuit of the heart cycle acts as a time vector, with initial values taken from the system, it was simulated for 0.83 second with Matlab based Ode23 method and were obtained time-dependent curves of the outputs of pressure, volume and flow, accordingly. If computation time has priority and a high level of accuracy is desired, then Ode23 is the best alternative for such applications. For our model, the Ode23 is more efficient choice than the other types of ODEs [16].

In addition, Matlab Ode23, which is preferred for our model, is within the scope of a performance improvement with other odes (TABLE IV). According to this comparison, Ode23 compared to other odes; successful steps and failed attempts in the calculation seem to be superior.

Ode solvers, estimate the local error in different components of the solution at each step. This error is considered as a function of the specified relative tolerance (RelTol) and the specified absolute tolerance (AbsTol). Accuracy values were created according to different error values in our model, and convergence values were calculated according to these error values in Table IV. According to this table, as the relative tolerance (RelTol) and absolute tolerance (AbsTol) values are both decreased, then the calculation time decreases and the performance of Ode23 increases with respect to the performance of other odes.

TABLE IV. COMPARISON OF ODE23, ODE45 AND ODE113 FOR THE NEW HEMS.

Solver	Elapsed Time (s)	Successful Steps	Failed Attempts	Function Evaluation
RelTol: 1e-3 - AbsTol: 1e-6				
Ode23	0.581493	2970	707	8266
Ode45	0.726103	914	1080	9727
Ode113	1.356708	1675	1551	7492
RelTol: 1e-12 - AbsTol: 1e-14				
Ode23	0.398419	2970	707	8266
Ode45	0.398793	914	1080	9727
Ode113	0.882071	1675	1551	7492

The new developed version of cardiovascular system (HEMS model) given in this paper is a new version of the HEMS model, using many variables and their interactions accordingly. Autocad based anatomical 3D drawings of all subsystems in the heart system were drawn targeting specific modelling and physiological-anatomical requirements. This is applied to facilitate intermediate version development to achieve the more integrated HEMS model possible, such as its multi-layer equivalents using electrical, mechanical and hydraulics parameters all together in an integrated model. Because, in order to integrate nonlinear complex biological structures and independent modelling, structures that can be described as interface models are needed. Biological and anatomical structures that act as multi-functional intelligent elements or systems can only be examined in a single form only as an electrical, mechanical or hydraulic system, by ignoring their multi-functionalities and based on their functions reduced to a single dominant function.

Our model defines the start time of the cardiac cycle as the time instant at which contraction of the right atrial wall becomes active due to action potentials. In order to represent the contraction and relaxation changes in each cardiac chamber, a time-dependent activation function was created for the ventricle and atria. Right and left ventricles are modelled as time-varying capacitors due to their flexible characteristics and activation function. The flow waveform was simulated to describe the systole and diastole phases of the cardiac cycle. The new HEMS model can integrate the cardiac dynamics and evaluate a group of variables, at the same time creating an efficient platform to study combinations of physiological properties as well as the effects of modifications in the system dynamics. The new version of the hydro-electro-mechanical system is modelled through 14 series of electrical circuits in which the integrated forms of all subsystems that make up the whole heart are produced by creating unique equivalent electrical circuits. By comparing our model with available hemodynamic models [17, 18], our proposed model is a more comprehensive model than other minimal models that can be used to evaluate the real time implementations.

Time-dependent average pressure measurement is a key tool to diagnose various heart diseases, which is one of the main reasons that this paper focused on the mathematical description of the average pressure of each segment of the whole the cardiovascular system (CVS) resulting in 28 equations (Eq. 4 –

Eq. 31). Additionally, Fig.5 gives the time-dependency of the critical four outputs which are graphically represented.

The graphics and values obtained are in conformance with the physiological values of the healthy subject and literature data available. As a result, it is predicted that the physiological system parameters based new HEMS model will act as a new method in the diagnosis of diseases and cardiovascular disorders involving the human heart through pressure-volume relationships and hemodynamic parameters.

V. FUTURE WORK

Engineering equivalent new models of the cardiovascular system are expected to make deeper understanding of complex processes possible occurring in the heart and blood vessels under the normal and pathophysiological conditions. All outputs obtained from or derived from pressure-volume relations or evaluated hemodynamic variables may serve as a useful assistant tool for diagnosing an illness and then recommending the new way of medical treatment of the cardiovascular disorders. The new HEMS model is thought to be potentially beneficial for hemodynamic simulation of diseases, cardiovascular disorders as well as congenital heart diseases, for example ASD, VSD and PDA, involving the human heart and its real-time simulation. The new HEMS model will be further developed and enhanced to improve the understanding of complex hemodynamic states and validation through patient data.

Abbreviations

The following abbreviations are used in this article:

HEMS	Hydro-electro-mechanical system
WM	Windkessel Model
CVS	Cardiovascular System
SA	Sinus Node or Sinoatrial Node
HP	Heart Period
AF	Activation Function
EDV	End-Diastolic Volume
ESV	End-Systolic Volume
E_F	Ejection Fraction
CO	Cardiac Output
SV	Stroke Volume
SLV	Left Ventricular Stiffness
SRV	Right Ventricular Stiffness
ASD	Atrial Septal Defect
VSD	Ventricular Septal Defect
PDA	Patent Ductus Arteriosus

REFERENCES

- [1] O. Frank, "Die Grundform des arteriellen Pulses. Erste Abhandlung. Mathematische Analyse." *Z. Biol.* 37, 483–526, 1899. DOI: 10.1016/0022-2828(90)91459-k.
- [2] Rosalia L., Ozturk C., Van Story D., Horvath M.A. and Roche E.T., "Object-Oriented Lumped-Parameter Modeling of the Cardiovascular System for Physiological and Pathophysiological Conditions.", *Adv. Theory Simul.*, 4: 2000216, 2021. DOI:10.1002/adts.202000216.
- [3] Westerhof B.E., Van Gemert M.J.C. and Van den Wijngaard J.P., "Pressure and Flow Relations in the Systemic Arterial Tree Throughout Development From Newborn to Adult.", *Front. Pediatr.* 8:251, 2020. DOI: 10.3389/fped.2020.00251.
- [4] Ajmal A., Tananant B., Andres J. R., Vinh N. D. L., Jessica C. R., "Investigation of optical heart rate sensors in wearables and the influence of skin tone and obesity on photoplethysmography (PPG) signal," *Proc. SPIE 11638, Biophotonics in Exercise Science, Sports Medicine, Health Monitoring Technologies, and Wearables II*, 1163808, 2021. DOI:10.1117/12.2578023.
- [5] Padmos R.M., Józsa T.I., El Bouri W.K., Konduri P.R., Payne S.J., Hoekstra A.G., "Coupling one-dimensional arterial blood flow to three-dimensional tissue perfusion models for in silico trials of acute ischaemic stroke.", *Interface Focus* 11:20190125, 2021. DOI:10.1098/rsfs.2019.0125.
- [6] Morany, A., Lavon, K., Bluestein, D. et al., "Structural Responses of Integrated Parametric Aortic Valve in an Electro-Mechanical Full Heart Model.", *Ann Biomed Eng* 49, pp. 441–454, 2021. DOI: 10.1007/s10439-020-02575-0.
- [7] Roy D., Mazumder O., Sinha A. and Khandelwal S., "Multimodal cardiovascular model for hemodynamic analysis: Simulation study on mitral valve disorders." *Plos One*, Vol. 16(3):e0247921, 2021. DOI:10.1371/journal.pone.0247921.
- [8] S. Shimizu, D. Une and T. Kawada, "Lumped parameter model for hemodynamic simulation of congenital heart diseases", *The Journal of Physiological Sciences*, Vol. 68, pp. 103–111, 2018. DOI: 10.1007/s12576-017-0585-1.
- [9] F. Yalcinkaya, E. Kizilkaplan and A. Erbas, "Mathematical modelling of human heart as a hydroelectromechanical system", 8th International Conference on Electrical and Electronics Engineering (ELECO), Bursa, 2013. DOI:10.1109/ELECO.2013.6713862.
- [10] Trzaska Z., "Study of mixed-mode oscillations in a nonlinear cardiovascular system." *Nonlinear Dyn* 100, 2635–2656, 2020. DOI: 10.1007/s11071-020-05612-8.
- [11] Jamie R. M., and Jiun W., "Expanding application of the Wiggers diagram to teach cardiovascular physiology.", *The American Physiological Society*, Vol. 38:2, pp. 170-175, 2014. DOI: 10.1152/advan.00123.2013.
- [12] V. S. Vasudevan, K. Rajagopal, J. F. Antaki, "Application of mathematical modeling to quantify ventricular contribution following durable left ventricular assist device support", *Applications in Engineering Science* 2022, Vol. 11, ISSN 2666-4968. DOI: 10.1016/j.apples.2022.100107.
- [13] Otto A. Smiseth, John M Aalen, Helge Skulstad, "Heart failure and systolic function: time to leave diagnostics based on ejection fraction?", *European Heart Journal*, Vol. 42:7, 14 February 2021, pp. 786–788, 2021. DOI: 10.1093/eurheartj/ehaa979.
- [14] June-Chiew H., Denis L., Andrew T. and Kenneth T., "Re-visiting the Frank-Starling nexus", *Progress in Biophysics and Molecular Biology*, Vol. 159, pp. 10-21, 2021. DOI:10.1016/j.pbiomolbio.2020.04.003.
- [15] Cattermole G.N., Leung P.Y., Ho G.Y., et al., "The normal ranges of cardiovascular parameters measured using the ultrasonic cardiac output monitor." *Physiol Rep.* 2017; 5(6):e13195. DOI:10.14814/phy2.13195.
- [16] Comunale G., Peruzzo P., Castaldi B. et al., "Understanding and recognition of the right ventricular function and dysfunction via a numerical study", *Sci Rep* 11, 3709, 2021. DOI:10.1038/s41598-021-82567-9.
- [17] Bahnasawy S., Al-Sallami H., Duffull S., "A minimal model to describe short-term haemodynamic changes of the cardiovascular system." *Br J Clin Pharmacol* 2021; 87: 1411–1421. DOI:10.1111/bcp.14541.
- [18] Roy D., Mazumder O., Sinha A. and Khandelwal S., "Multimodal cardiovascular model for hemodynamic analysis: Simulation study on mitral valve disorders." *PLoS ONE* 2021; 16(3):e0247921. DOI:10.1371/journal.pone.0247921.

BIOGRAPHIES



Fikret YALCINKAYA received his B.Sc. degree in Electrical and Electronics Engineering from METU, Turkey (1984). He received his integrated Ph.D. degrees in Biomedical Engineering from Engineering and Applied Sciences department of the University of Sussex, UK (1994-

1998).

Since 2001, he has been an Assistant Professor with Electrical and Electronics Engineering Department, Kirikkale University, Kirikkale, Turkey.

His research interests include nanobiosensors, medical instrumentation, physiological modelling (cardiovascular system, respiratory system and auditory system) and medical imaging (specifically CNN and Deep Learning based melanoma early detection).



Ertem KIZILKAPLAN received his B.Sc. degree in Electrical and Electronics Engineering from Kirikkale University, Kirikkale, Turkey (2010) and M.Sc degree in Electrical and Electronics Engineering from Kirikkale University, Kirikkale, Turkey (2022). Since 2010, he has been working as an Electrical and Electronic

Engineer at MSK Immobilien Invest GmbH.

His research interests include mathematical modelling, physiological modelling (cardiovascular system), heart diseases and biomedical research.

Axial Flux Motor Design for Ventilation Fans Used in The Automotive Industry

Tuba Kocaer and Yasemin Öner

Abstract— In this paper, it was aimed to carry out the design, electromagnetic analysis of a 3 phase and dual 3 phase, axial flux permanent magnet (AFPM) synchronous motor with power level of 240W and 3000rpm speed. The geometrical dimensions of motors were designed to be the same and the motors are intended to be used in ventilation applications in the automotive industry. The performance of 3-phase 12 slot / 10 pole AFPM motor, and a dual 3-phase 12 slot / 10 pole AFPM motor were analyzed. Finite element analysis (FEA) was used to observe the characteristics of both motors. Torque values, torque ripples, motor losses, magnetic flux density and current density were examined and the results were compared.

Index Terms— Axial flux permanent magnet motor (AFPM), automotive ventilation fan, performance comparison.

I. INTRODUCTION

THE VEHICLES used in transportation, which cover an important part of human life, are shaped in line with human needs. Standards and customer acceptance tests in automotive applications, including electric vehicles, burden significant noise emission limitations for automotive applications due to the users' comfort in driving and the pollution in the environment, and as an obvious result, it is aimed to reduce the noise level in the system.

Axial fans are used in many on-road and off-road vehicles for ventilation, heating, refrigeration, and engine cooling. With the developments in the automotive sector, motors used in axial fans are desired to be long-life, low noise, high-efficiency, small-volume, low cost, and low weight. You can see a brushless direct current motor (BLDC) fan used in ventilation systems in Fig. 1. [1]

Permanent magnet synchronous motors (PMSM) and permanent magnet (PM) brushed DC motors, which are frequently used in automotive ventilation systems, are a major

source of noise. As it is known, one of the most important disadvantages of brushed DC motors is that they have a torque ripple that includes acoustic noise. The torque ripple changes periodically depending on the position of the rotor and its adverse effects occur mainly at low speeds and light loads. [2]




Fig.1. BLDC fan [1]

Great exertion has been made to reduce acoustic noise and examine its causes. For this purpose, several studies have been conducted on different models to be used to reduce acoustic noise, such as in automobile applications. With these studies, many different types of the motor such as reluctance motor, IPM; motor designs, and different voltage levels were studied [3], [4],[5].


In [6], best torque density, torque quality, and acoustic noise were investigated by applying different slot/pole combinations on the performance of a fractional concentrated winding spoke type synchronous motor.

In this study, two motor designs were made for the automotive industry, which occupy a smaller volume, have high power density and efficiency, and can be used in axial flux fans. 3D analysis of the designed motors were made and the analysis results were compared. It is aimed to reduce torque ripple and losses in motors designed as 3-phase and 6-phase.

TUBA KOCAER, is with Department of Electrical Engineering University of Yıldız Technical University, Istanbul, Turkey, (e-mail: tubakocaer@gmail.com).

 <https://orcid.org/0000-0003-4614-7158>

YASEMİN ÖNER, is with Department of Electrical Engineering University of Yıldız Technical University, Istanbul, Turkey, (e-mail: yoner@yildiz.edu.tr).

 <https://orcid.org/0000-0002-5310-6875>

Manuscript received May 23, 2022; accepted July 25, 2022.
DOI: [10.17694/bajece.1120298](https://doi.org/10.17694/bajece.1120298)

II. AXIAL FLUX MOTOR

When we examine the history of electrical machines, it is seen that the oldest electrical machines are axial flux machines (AFM) [7]. The first records of studies on axial flux motors belong to M. Faraday, and it was named as Faraday disk in 1831 [8]. In 1889, N.Tesla patented this disk-shaped electric machine. However, the use of axial flux motors has not become widespread due to the strong axial magnetic attraction force between the stator and the rotor, the challenge and cost of producing the lamination used in the stator, the difficulty of assembling the machine, and the difficulties in providing that the air gap between the stator and rotor disk is uniform and stable everywhere. [7]

Today, with the developments in magnet technology, studies on axial flux motors and their use have increased. The development of permanent magnets, the discoveries of new magnetic materials, and the global cheapness of magnets and studies on electric vehicles will contribute to the widespread use of axial flux motors.

Axial flux motors can be used instead of radial flux motors due to their higher power density and more compact structure. These motors are generally used in electric vehicles, pumps, axial fans, robots.

In axial flux motors, the magnetic flux goes directly perpendicularly through the air gap. These motors consist of a ring-shaped stator disk with windings and a rotor disk with magnets. The torque is produced at the radial distance between the stator inner diameter and the outer diameter. The number of poles can be increased without changing the radial length, thus reducing the axial length and increasing the power density. Therefore, axial flux motors designed by increasing the number of poles have the ability to produce high torque at low speed. Axial flux motors are generally named according to the stator and rotor arrangement and their number, the placement of the magnets, use of the core and their slot structure.

Single-sided axial flux machines are simpler than other axial flux motor structures. However, the torque production capacity of these machines is lower than other axial flux machine types. Fig. 2 shows the general structure of the single-sided axial flux machine, in which the stator core is wound and PMs are placed on the surface of the rotor disc.

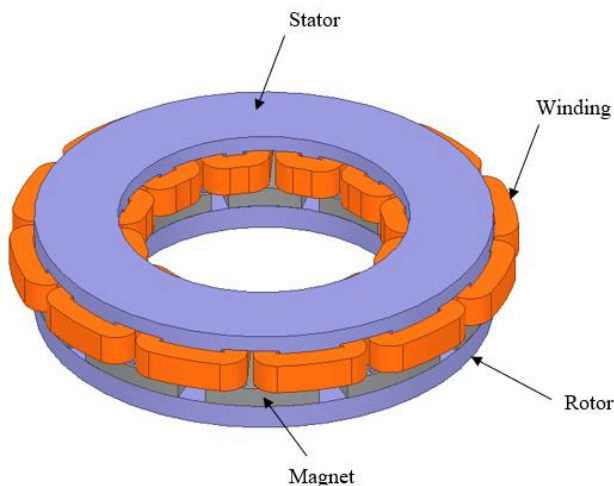


Fig.2. Construction of the single-sided AFM

Single-sided AFM provide a volume advantage at the same power when compared to radial flux motors. It is suitable for fan and pump applications where low power, less volume, less weight low cogging torque, and low noise level are required.

III. DESIGN OF AXIAL FLUX PERMANENT MAGNET MOTOR

This study, it is aimed to enable the rotor part of the AFM to be mounted directly to the impeller of the fan, thus providing a space advantage and making a thinner fan in a more compact structure. Since axial flux motors have higher power densities than radial flux motors, a smaller motor design than radial flux motors of the same power have been proposed.

In the study, the torque ripple, acoustic noise caused by the AFPM was tried to be reduced. Therefore, two AFPM designs with dual 3-phase winding set and 3-phase winding configurations with the same motor dimensions, slot/pole combination, and materials used, were designed and the results with the analysis output were compared with each other.

The 12 slot 10 poles machines are quite popular in product and technical literature [9]. By using double-layer winding in the motor structure with 12/10 slot/pole combination, the values of the fifth and seventh harmonics that significantly affect the torque ripple are reduced. In addition, it is intended to achieve low cogging torque and torque ripple from the machine. The cogging torque is proportional to the least common multiple (LCM) of the number of poles and the number of slots and the selected slot and pole combination is suitable for low cogging torque.

The winding structure of an AFPM is determined by the slot-pole combination. Also, the winding structure of PM motors is in two main forms, integrated slot winding and fractional slot winding. Fractional-slot machines with a tooth winding and PMs have certain advantages such as high power density and efficiency, short end windings, low cogging torque, high slot fill factor and no skew [10], [11].

If the value of q , which expresses that the number of slots per pole per phase, is fraction the winding is called fractional slot winding.

$$q = \frac{Q_s}{2pm} \quad (1)$$

In Equation 1, Q_s refers to the number of slots, m refers to the number of phase and $2p$ refers to pole pairs.

If a three-phase system is doubled by increasing the number of phases, it halves the number of slots per pole per phase. As can be seen from the Eq1, when only the number of phases is changed in a motor, q values of the three-phase motor is 0.4, meanwhile the q values of the six-phase motor is 0.2.

With the progress of the motor control techniques and power electronics equipment used to drive the motor, the FSCW multiphase motor has significant advantages due to its high power densities and short end winding lengths [12],[13]. Multiphase motors have been studied in the literature because of their effects such as improvement in EMF waveform, higher performance, lower phase current, increased efficiency

due to lower rotor losses, reduced acoustic noise, reduced torque ripples and increased torque density.

TABLE I
DESIGN PARAMETER

Parameters	Values
Output Power (W)	240
Output Torque (Nm)	0.6
Phase Voltage (V)	26
Speed (rpm)	3000
Target Efficiency (%)	90

The parameters of the motor to be mounted inside the axial fan are given in Table 1. Fans used for ventilation in automotive applications are powered by 24V or 12V batteries and due to the battery output voltage, motor tests are usually performed with 26V. For this reason, the motor design was made according to the 26V supply voltage.

The design parameters and geometric dimensions of the motor, whose design and analysis have been made, are given in Table 2.

TABLE II
MOTOR DESIGN PARAMETERS AND DIMENSIONS

Parameters	3-Phase AFPM Values	Dual 3-Phase Set AFPM Values
Phase Number	3	6
Pole Pairs	5	5
Slot Number	12	12
Operation Voltage	26 V	26 V
Outer Diameter of Stator	110 mm	110 mm
Inner Diameter of Stator	70 mm	70 mm
Stator Material	M250-50A	M250-50A
Slot Opening	3 mm	3 mm
Slot Height	10 mm	10 mm
Stator Tooth Width	3.5 mm	3.5 mm
Stator Axial Length	14 mm	14 mm
Airgap	0.7 mm	0.7 mm
Magnet Thickness	6 mm	6 mm
Magnet Material	Ferrit Y30	Ferrit Y30
Rotor Axial Length	8 mm	8 mm
Rotor Material	Steel 1008	Steel 1008
Number of Turns	21	21
Wire Diameter	1 mm	1 mm
Number Of Strands	2	2
Coil Pitch	1	1

In this study, the two motors are designed to give the same outputs and the dimensions of the motors are identical. The number of turns in a coil is the same in both motors. However, the number of turns in series per phase is half of the 3-phase motor in a dual 3-phase set motor.

The phase windings of the 12-slot, 10-pole motor are divided into two sets and an appropriate phase shift is made between both phase sets, as seen in Fig. 3 (b). First set 3-phase windings expressed as A1, B1, C1 and the second set 3-phase windings as A2, B2, C2. The winding structure of the two sets 3-phase motor has a phase shift of 120 degrees between phases A1, B1, C1 and between phases A2, B2 and C2. However, there is a 90° phase shift angle between the first and second winding set. In other word, there is a phase-shifting of 90° between the A2 and A1 phases. You can see the winding arrangement of the motors in Fig. 3.

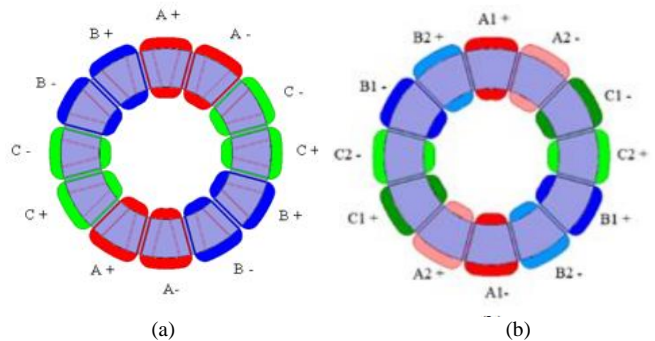


Fig.3. Winding arrangement for a 12 slot / 10 pole geometry. (a) 3-phase winding. (b) Dual 3-phase winding

A. Electromagnetic Analyzes

Electromagnetic analyzes were carried out in 3D with the finite element method using Ansys Maxwell software. In these analyzes made with sinusoidal current supply, moment, output power, current densities, flux densities, cogging torques, losses are examined.

B. Torque Analysis

The two motors were operated with the same current and voltage value. You can examine the average torque values and torque ripple of the motors from Table 3.

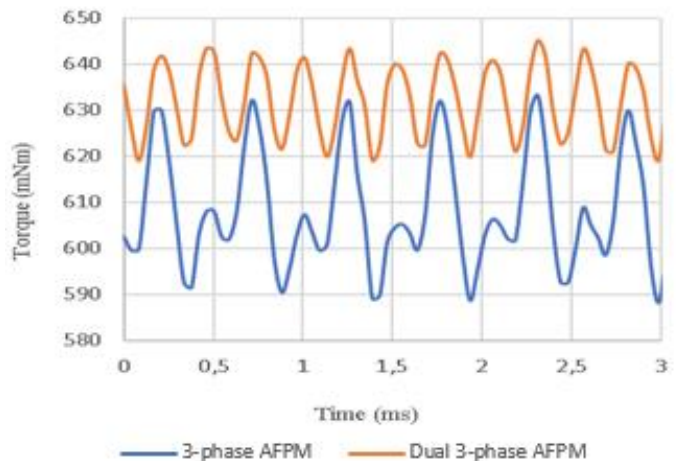


Fig. 4. The torque graph of AFPM

When the torque plots of the dual 3-phase motor and the 3-phase motor are compared, it is seen that the average torque value of the dual 3-phase motor is 2.5% higher than the 3-phase motor. In addition, It has been observed that the torque ripple value of the 3-phase motor is 3.24% higher. You can see the torque graph of both AFPM in Fig.4.

TABLE III
TORQUE VALUES

Torque Values	3-Phase AFPM	Dual 3-Phase AFPM
Average Torque (mNm)	607.98	632.41
Torque Ripple (%)	7.39	4.15

C. The EMF Waveforms

The EMF waveform of the AFPM is shown in Fig. 5 and Fig. 6. While the rms value of induced voltage in one phase of the dual 3-phase motor is 6.85V, that of the 3-phase motor is 14.12V.

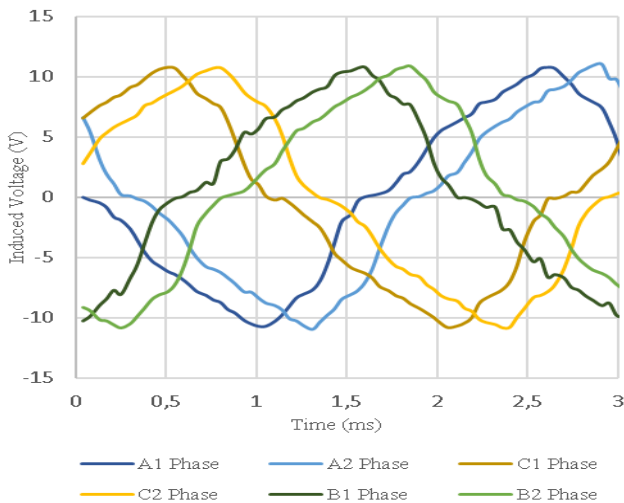


Fig.5. The induced voltage at load condition of dual 3-phase AFPM

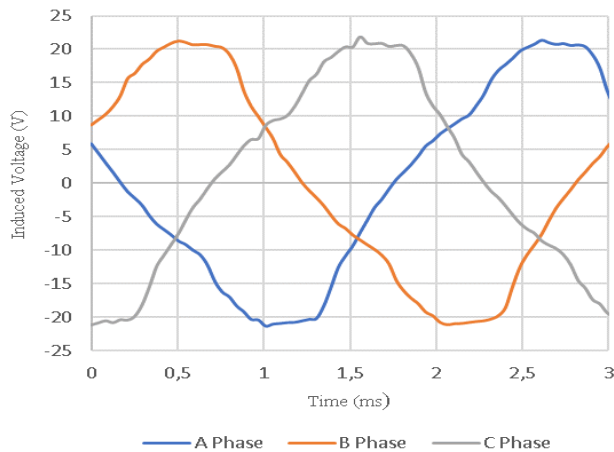


Fig.6. The induced voltage at load condition of 3-phase AFPM

D. Flux Density Analysis

Fig.7 shows the flux density of both motors. No magnetic saturation was observed in the stators and rotors of both motors. According to the results of the analysis, the maximum magnetic flux density observed in the motors is 1.7T and it was observed in the stator teeth.

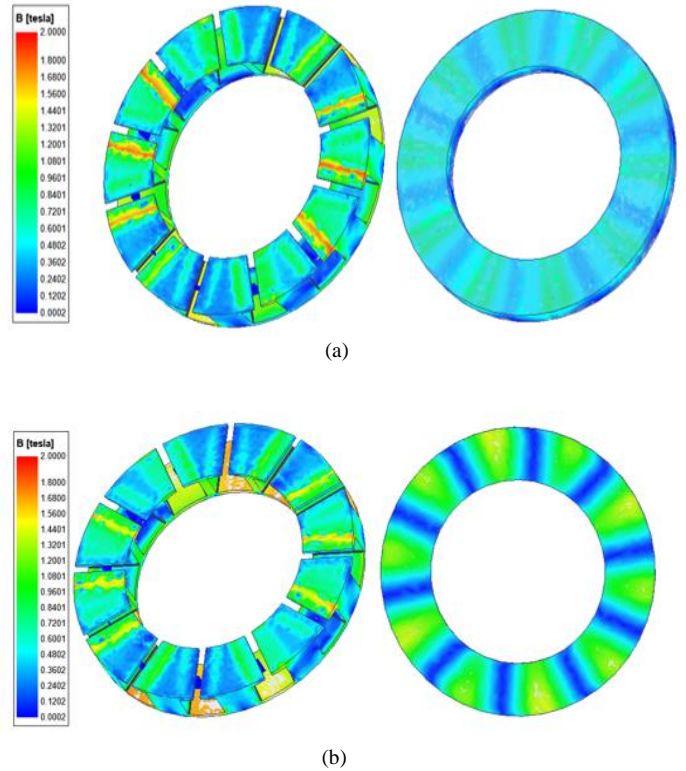


Fig.7. The flux density of AFPM (a) Dual 3-phase motor. (b) 3-phase motor

E. Current Density Analysis

You can see the current density values in the coils of both motors in Fig. 8. The maximum current density value in the coils is 3.75A. The current density values in Fig. 8 show that the motors can operate in S1 mode without any cooling requirement under nominal load conditions.

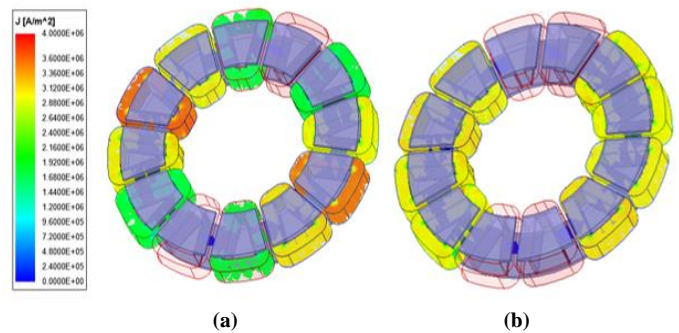


Fig.8. The current density of AFPM (a) Dual 3-phase motor, (b) 3-phase motor

F. Loses

According to the analysis results, it was seen that the efficiency of the dual 3-phase motor was higher. You can see the losses of both AFPM from Table 4.

TABLE IV
MOTOR LOSES

Loses	3-phase AFPM	Dual 3-phase AFPM
Core Loss	4.69 W	4.38 W
Eddy Current Loss	2.42 W	2.26 W
Hysteresis Loss	1.91 W	1.77 W

As seen in Fig. 6, Tables 3 and 4, the windings of a multiphase machine produce an EMF waveform closer to the sinusoidal so the winding factor for the main harmonic is high. This results in an increase in torque per ampere for the machine. Furthermore, higher phase number winding improved armature EMF waveform as it will contain lower harmonic contents and reduced amplitude-frequency of the torque ripple thus reducing acoustic noise and reducing rotor loss by this means increased efficiency of the motor

IV. CONCLUSION

This paper presented two AFPM motor has 12/10 slot-pole combination with FSCW configurations that can be used for an automotive ventilation fan application. It is designed in such a way that the geometric dimensions of the two motors are the same, but the number of phases is different. The losses, current density, magnetic flux density, EMF waveform and torque data of the two motors designed and analyzed in 3D were compared. According to the analysis results, when compared to the three-phase AFPM motor, the dual 3-phase AFPM motor has higher torque density, lower torque ripple, and motor losses. Based on the results presented, it can be said that dual 3-phase AFPM is more suitable for the application in terms of torque ripple, acoustic noise, and efficiency.

REFERENCES

- [1] "Automotive - BL-DC fans for Commercial Vehicles", Ebm-Papst product catalog 2019-10.
- [2] B. Stumberger, G. Stumberger, M. Hadziselimovic, I. Zagradsnik, "Torque ripple reduction in exterior-rotor permanent magnet synchronous motor", Journal of magnetism and magnetic materials, Elsevier,
- [3] G. F. Lukman, K. Jeong, J. Ahn "Comparative Performance Analysis of Switched Reluctance Motors for Automotive Fan", 2019 22nd International Conference on Electrical Machines and Systems (ICEMS), 2019.
- [4] J.Choi, I.Jung, J.Hur, H. Sung, B.Lee "On the Feasibility of the Brushless DC (BLDC) Motor and Controller for 42V Automotive Cooling Fan System" 2007 IEEE International Electric Machines & Drives Conference,2007
- [5] A. S. Abdel-Khalik , S. Ahmed , and A. Massoud "A Six-Phase 24 Slot-10 Pole Permanent-Magnet Machine with Low Space Harmonics for Electric Vehicle Applications", IEEE Transactions on Magnetics, vol.52, June 2016.

- [6] E.Carraro, N.Bianchi, S.Zhang M. Koch, "Design and Performance Comparison of Fractional Slot Concentrated Winding Spoke Type Synchronous Motors With Different Slot-Pole Combinations", IEEE Transactions on Industry Applications, vol.54, May-June 2018.
- [7] Jacek F. Gieras, Rong-Jie Wang, Maarten J. Kamper "Axial Flux Permanent Magnet Brushless Machines" Second Edition, 2008
- [8] Atherton W.A. From Compass to Computer; A History of Electrical and Electronics Engineering. London, UK: The Macmillan Press Ltd., 1984.
- [9] J. R. Hendershot, TJE Miller "Design of Brushless Permanent-magnet Machines", 2010
- [10] N. Bianchi, S. Bolognani and M. Dai Pre, "Magnetic Loading of Fractional-Slot Three-Phase PM Motors With Nonoverlapped Coils," IEEE Transactions on Industry Applications, vol. 44, no. 5, pp. 1513-1521, Sept.-Oct. 2008.
- [11] A. M. El-Refai, "Fractional Slot Concentrated Windings Synchronous Permanent Magnet Machines: Opportunities and Challenges," IEEE Trans. on Ind. Elec., vol. 57, no. 1, pp. 107-121, Jan. 2010.
- [12] L. J. Wu, Z. Q. Zhu, J. T. Chen, Z. P. Xia, and G. W. Jewell, "Optimal split ratio in fractional-slot interior permanent magnet machines with nonoverlapping windings," IEEE Trans. Magn., vol. 46, no. 5, pp. 1235-1242, May 2010.
- [13] M. Barcaro, A. Faggion, N. Bianchi, S. Bolognani, "Sensorless rotor position detection capability of a dual three-phase fractional-slot IPM machine," IEEE Trans. Ind. Appl., vol. 48, no. 6, pp. 2068-2078, Nov./Dec. 2012.

BIOGRAPHIES



TUBA KOCAER received the B.S. degree in electrical engineering from the Yıldız Technical University, İstanbul, in 2019. She has started M.Sc. at 2019 at Yıldız Technical University electrical engineering department, electrical machines and power electronics program. She has been working at Kormas Elektrikli Motor San.ve Tic. A.Ş R&D department as electrical design engineer since February 2020. Her research and professional interests include electrical motor design & optimization, and advanced electromagnetic analysis.



YASEMİN ÖNER received the B.Eng., M.Sc., and Ph.D. degrees in electrical engineering from Yıldız Technical University, İstanbul, Turkey, in 2007, 2009, and 2013, respectively. She is currently Associate Prof. Dr. in Yildiz Technical University. Her research interests include analytical solutions for electrical machines.

Graph Visualization of Cyber Threat Intelligence Data for Analysis of Cyber Attacks

Mücahit Sülü and Resul Daş


Abstract—Threat intelligence enables us to make faster, more informed, data-backed security decisions and change their behavior from reactive to proactive in the fight against threat actors. Cyber threat intelligence sources include open-source intelligence, social media intelligence, human intelligence, technical intelligence, device log files, forensically acquired data or intelligence from the internet traffic, and data derived for the deep and dark web. In this study, graph visualization is discussed for the intelligible and accurate analysis of complex cyber threat intelligence data, including network attacks. The processes of collecting, cleaning, organizing, and visualizing cyber intelligence data in different formats and contents on a single platform are given step by step. Dynamic graphs play an active role in these systems, where the attack locations and targets from different points are constantly variable. Therefore, research on dynamic graph solutions and visualization in the visual analysis of cyberattacks is presented.


Index Terms—cyber security, graph visualization, dynamic graph, cyber threat intelligence, cyber attack visualization, big data.

I. INTRODUCTION

GRAPHS HAVE created models for solving and analyzing many unsolved problems throughout history. These models have a wide variety of types and formats. Graph visualization is of great importance as it provides convenience in terms of understanding and tracking problems. Today, graphs have become so popular that it is possible to encounter a problem or an algorithm modeled with graphs in almost every field in the scientific world. Effective solutions through graphs are seen in the literature in almost all fields such as computer sciences [1], social sciences[2], linguistics [3], engineering[4], mathematics [5], biology and genetics [6].

Every day in meeting rooms, people use graph techniques to label relationships and create diagrams to explain their thoughts to others. Graphs can express relatively complex concepts that other visualizations cannot. Data expressed with graphs can be analyzed with graph techniques. The right technique, when chosen wisely, can give the simplest and most intuitive expression of a particular type of knowledge. When poorly selected, a graph can be painfully abstract and broad. Graph visualization may seem confusing. The standard

 **Mücahit SÜLÜ** is with the Department of Computer Programming, Organized Industrial Zone Vocational School, İnönü University, Malatya, 44280 TÜRKİYE e-mail: mucahit@inonu.edu.tr

 **RESUL DAŞ** is with the Department of Software Engineering, Technology Faculty, Firat University, Elazığ, 23119 TÜRKİYE e-mail: rdas@firat.edu.tr

Manuscript received March 19, 2022; accepted July 28, 2022.
DOI: [10.17694/bajece.1090145](https://doi.org/10.17694/bajece.1090145)

pie charts and bar charts you would normally find in a cyber control panel are not the subject of graph theory. Graph theory is the visualization of connections and relationships in data. Cyber attacks can be made to any device connected to a local network or internet. It is vitally important to detect these attacks and take action accordingly. Graph visualization methods and graph techniques can be used to make sense of and analyze cyber attacks.

In this article, the steps of analyzing cyber threat intelligence data in different formats through various stages and visualizing them as graphs are examined. It is a concise study that guides researchers working in this field. In this context, the second part of the article examines the cyber intelligence data. In the third part, graph visualization process steps and processes are explained. In the last part, the general results of the study are presented.

II. BACKGROUND

Intelligence is tactical, technical or predictable information that is specially collected and analyzed through certain filters for presentation to military or political higher authorities. We can describe it as valuable data that has been unearthed by combining interrelated parts over raw information obtained from almost any source. Intelligence activities are indispensable for states and require the processing of information and documents compiled from various sources in response to the needs determined by the state. In history, the effectiveness of intelligence data has always come to the fore in the destruction or establishment of states, and in winning or losing wars. For this reason, almost all heads of state have made an effort to collect sound and reliable intelligence data and use them in the most effective way. When we take a look at Turkish history, we see that Sultan Abdülhamit II was a sultan who understood the necessity of intelligence and applied it, and benefited from intelligence activities by establishing the Yıldız Intelligence Organization. Additionally, we see in history that he took a cautious approach towards the intelligence data and thus, made decisions after verifying the reliability of the source and the integrity of the data. Considering the types of intelligence according to their fields, there are various types of intelligence data such as military, political, economic, geographical, social, biographical, technological, transportation and communication. However, in this study, we focus on the content, scope, analysis and visualization of cyber intelligence data within the scope of technological intelligence data.

A. Cyber Intelligence

Cyber intelligence is the collection and discovery of threats from electronic media that can harm institutions and organizations, business elements and security at any level. It is a type of intelligence that enables early measures to be taken by detecting the aims, methods or attack types of the attackers as a result of analyzing the data collected and enriched from electronic media through a process. When we look at the data breaches we have encountered in recent years, it has been revealed that the measures taken during or after the cyber attacks do not always work. Due to the cyber world we live in and the rapid development of technology, institutions can face different threats and can receive hundreds of attacks at any time. It is not easy to follow the cybercriminals and their techniques targeting critical systems in the internet world, and it is a situation that requires large budgets and dealing with big data. At this point, cyber intelligence emerged and started to play an important role in cyber attacks. While cyber attackers carry out their cyber activities in a highly motivated way, they improve their attack methods and diversity day by day. Because of this increasing cyber threat, it is critical to be aware of an attack before it happens. The power of predicting attacks accelerates the decision-making process of institutions. At that point, the importance of Cyber Intelligence is increasing day by day.

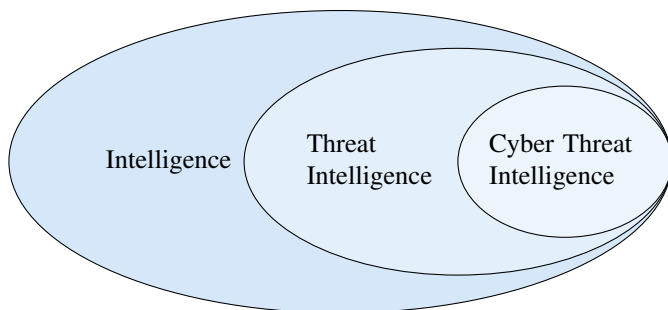


Fig. 1. Cyber Threat Intelligence

B. Cyber Threat

A cyber threat is an attempt by malicious persons or organizations to gain unauthorized access to control system devices or network, disrupting the network structure or rendering it unusable. Cyber threats can originate from various different points/places, people, institutions or organizations. This is where cyber intelligence comes into play. The actions taken by cyber threat sources such as terrorists, hackers, commercial competitors, spies, hostile states, unhappy employees, organized crime groups with the aim of causing harm are called cyber threats. These threats provide insight into what kind of scenario attackers might follow when attacking their victims. Malware, Spyware, Malvertising, Man in the Middle (MITM), Wiper Attacks, Distributed Denial of Service (DDoS), Ransomware, Botnets, Trojans, Phishing, Data Breaches, Worms, Keyloggers, Backdoors, Advanced Persistent Threats are important examples of cyber threats.

C. Cyber Threat Intelligence

Cyber threat intelligence - CTI) is knowledge, skills, and experience-based information concerning the occurrence and assessment of both cyber and physical threats and threat actors that are intended to help mitigate potential attacks and harmful events occurring in cyberspace. The purpose of cyber threat intelligence is to help institutions and organizations understand the risks of cyber attacks or cyber threats. Examples of these attacks are 0-day attacks, crypto viruses, APT (Advanced Persistent Threat), botnets or exploits. These threat elements are reported with the intelligence activities revealed using various software tools and presented to the relevant institutions and organizations together with special protection methods, thus providing guidance for an active defense. Such attacks can cause serious damage to institutions and organizations. Thanks to cyber threat intelligence, extensive and deep analysis data is used to help protect institutions and organizations from such attacks.

1) *Cyber Threat Intelligence Data*: Advances in attack methods make it extremely difficult to identify the attacker and the attack. Traditional security measures such as firewalls, signature registration, and intrusion detection system (IDS) fail to prevent these new types of attacks. To meet these challenges, the emerging field of cyber threat intelligence uses artificial intelligence and machine learning techniques to intelligently detect, learn and overcome advanced cyber attacks. There is an increasing trend in the use of Machine Learning (ML) and data mining techniques in the static and dynamic analysis of malware, due to their efficiency and powerful network anomaly detection. In addition to these, different mechanisms such as honeypot are used to deceive the attackers. In such methods, security professionals use fake information or sources that appear to be legitimate to lure attackers, monitor the attackers' activities, and detect the attack and its type. Cyber threat intelligence data mining has an increasing popularity today.[7]

TABLE I
CYBER OBSERVABLE SPECIAL DATA TYPES [8]

Type	Description
boolean	True or False.
float	One IEEE 754 [IEEE 754-2008] double-precision number.
hashes	One or more cryptographic hashes.
integer	Integer.
list	An ordered array of values.
open-vocap	Type from a STIX or suggested word value.
string	Unicode character string.
timestamp	A time value (date and time).
binary	A byte array.
hex	A decimal number at the base of eight.
dictionary	Set of key-value pairs.
object-ref	Cyber observable reference.
observable-objects	One or more cyber observable objects.

The data is collected, analyzed and organized from the deep/dark web, blogs, social media and forums, often with artificial intelligence support. Although there are many companies that provide this data today, companies that attach importance to security provide it themselves. Structured Threat Information Expression (STIX) is an application that enables organizations to share cyber threat intelligence data with each other in a consistent and readable way. An example of STIX data is given by the list 1. Cyber observable custom data types are given in Table I. The Figure 7 shows STIX domain object relationships. Here, many relationship objects such as targets, users, viruses, threats are visualized. Figure 8 shows a simple cyber attack graph structure.

```

1 {"type": "bundle",
2  "id": "bundle--5d0092c5-5f74-4287-9642-33f4c354e56d"
3  ,
4  "spec_version": "2.0",
5  "objects": [{
6    "type": "indicator",
7    "id": "indicator--8e2e2d2b-17d4-4cbf-938f...",
8    "created_by_ref": "identity--f431f809-377b...",
9    "created": "2021-04-29T14:09:00.000Z",
10   "modified": "2021-04-29T14:09:00.000Z",
11   "object_marking_refs": [{"marking-definition..."}],
12   "name": "Poison Ivy Malware",
13   "description": "This file is part of P",
14   "pattern": "[file:hashes.'SHA-256' =
15   'aec070645fe53ee3b3763059376134f058cc3372...' ]"
16   },{
17   "type": "marking-definition",
18   "id": "marking-definition--34098fce-860f-48...",
19   "created": "2021-09-01T00:00:00.000Z",
20   "definition_type": "tlp",
21   "definition": {
22     "tlp": "green"}
23   }
24 ]}
25 \caption{}

```

Listing 1. A sample snippet from the STIX 2 package.

D. Data Processing

There are two types of data collection methods, the first is collected over a period of time and the second is constantly flowing from one device to the next.

1) *Batch Data Processing*: Batch Data Processing is an efficient way to process large amounts of data collected over a period of time. It also helps reduce the operational costs that businesses can spend on their workforce, as it doesn't need dedicated data entry officers to support its operation. It can be used offline and gives administrators full control over when they start processing, whether overnight, on a weekend, or at the end of a pay period.

As with anything, batch processing has a few downsides. One of the biggest problems businesses see is that these systems can be difficult to debug. If you don't have a dedicated IT team or specialist, trying to repair the system when an error occurs can be detrimental and require an outside consultant. In addition, if the business needs quick returns, for example, if the customer searches for a product on an e-commerce website, the customer will have bought that product because the business's system will analyze and offer this customer weeks or days later.

2) *Streaming Data Processing*: Streaming (Real-Time) Data Processing is the process of analyzing data streaming from one device to another almost instantly. This method of continuous calculation takes place as data flows through the system with no mandatory time limitation on output. With nearly instantaneous streaming, systems do not require large amounts of data to be stored. Stream processing is very useful if the events you want to watch happen frequently and are close together in time. It is also best to use if the incident needs to be detected immediately and responded to quickly. It is also useful for tasks such as stream processing, fraud detection, and cybersecurity. If transaction data is stream processed, fraudulent transactions can be identified and stopped before they are completed. In addition, instant suggestions can be offered to the customer. One of the biggest challenges organizations face with stream processing is that the long-term data throughput rate of the system must be the same or faster than the long-term data-input rate, otherwise the system will start to experience storage and memory related problems. Another challenge is trying to find the best way to deal with the huge amount of data being generated and moved. To keep the data flow in the system at an optimal level, organizations need to create a plan for how to reduce the number of copies, how to target the compute cores, and how to make the best possible use of the cache hierarchy.

III. DYNAMIC COMPLEX GRAPHS AND SOME TYPES

It is difficult to represent a time dimension with graphs. It is accepted in scientific circles that simultaneous representations of states are more informative than sequential representations. For example, a time series bar chart is much better at showing behavior over time than animating changes in a single bar. In the latter case, repeatedly scrolling the animation back and forth in time would be a means of getting the essence of the change; however, it will not be easy to detect correlations in behavior when comparing, as it will not be as accessible as a time series. It is more effective to be able to see the values over time simultaneously. However, it is not clear how this principle can be applied to graphs. Since data in dynamic graphs changes over time, visualization can be made with data collected by batch data processing (given by II-D1) or Streaming (Real-Time) Data Processing (given by II-D2). Now we will examine their advantages and disadvantages.

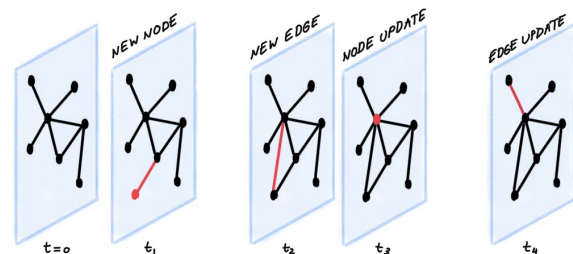


Fig. 2. Dynamic graph sample [9]

A. Implementation of Dynamic Graphs

A dynamic graph evolves and changes over time, so it can be viewed as a timed sequence of events. Figure 2 shows an example of a dynamic graph that changes over time. In this example, a new node and a new edge are added at t_1 moment, only a new edge is added at t_2 moment, the node degree is updated at t_3 moment, and the edge is updated at t_4 moment. In the edge update, the weight assigned to the edge can be changed, a label can be assigned or a new edge can be added. In a node update, it can be a node's degree or a label update.

V (set of nodes), E (set of edges), f (weights of nodes) and g (weights of edges) [10];

- In a node dynamic graph, the set V changes with time. Therefore, some nodes can be added or removed. When the nodes are removed, the edges formed with them are also removed.
- In an edge dynamic graph, the set E changes with time. Thus, edges can be added or removed from the graph.
- In a node-weighted dynamic graph, the f function changes with time; so the weights at the nodes also change.
- In an edge-weighted dynamic graph, the g function changes over time.
- In a full-weighted dynamic graph, both f and g functions may change over time.

that are drawn in thinner indicate that the migrations are less intense.



Fig. 4. An example of circle graph. [14]

With Figure 5, the hierarchy of task dependencies in Cyber entities is given, the grouping operation here makes the graph more understandable.

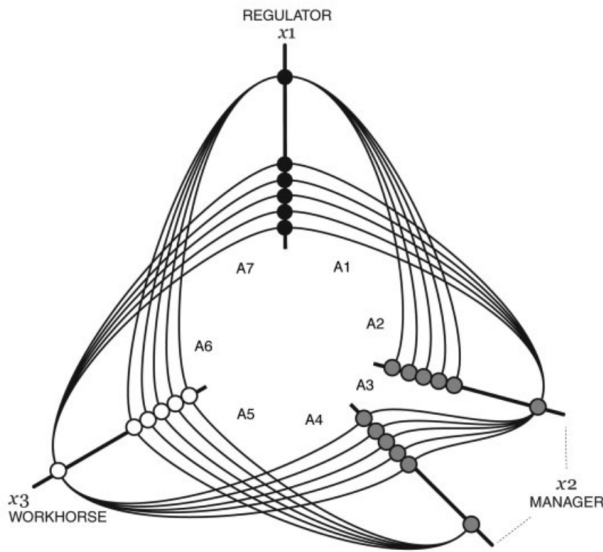


Fig. 3. An example of Network Hive Graph [11]

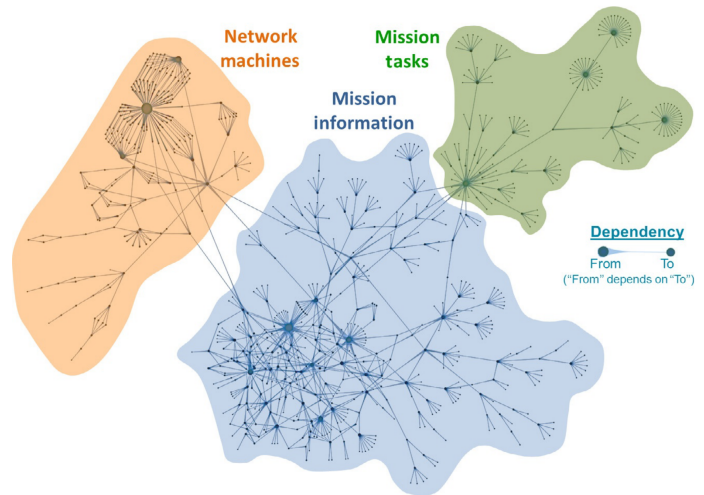


Fig. 5. Hierarchy of mission dependencies on cyber assets. [13]

B. Types of Some Complex Graphs

Graphs are a great way to visualize data and display statistics. Popular graph types include Hive Graphs[11], Circle Graphs[12], Hierarchy Graphs [13]. For illustrate, Hive graphs transparently models the network structure, are easy to understand and can be easily modified to identify patterns of interest. An example network structure model is given in Figure 3. In Figure 4, estimates of migration flows between and within regions for the period from 2005 to 2010 are given as a circle graph. In the graph given in Figure 4, the lines drawn in bold indicate that the migrations are more intense, and the lines

IV. GRAPH VISUALIZATION PROCESS STEPS OF CYBER THREAT INTELLIGENCE DATA

In order for cyber intelligence data to be analyzed effectively, it must first be expressed as graphs. There are two major challenges for viewing and analyzing potentially very large and complex graphs of cyber attacks. The first is the real-time processing of large amounts of data and converting it into visual format. The second is to visualize complex graphs, including all possible attack paths, by keeping them manageable.[15] In order to visualize complex graphs,

graph visualization steps given in figure 6 are applied and data is made meaningful. These data can then be analyzed using graph analysis methods. Graph algorithms[16] and graph partitioning[17], [18] are some of the graph analysis methods.

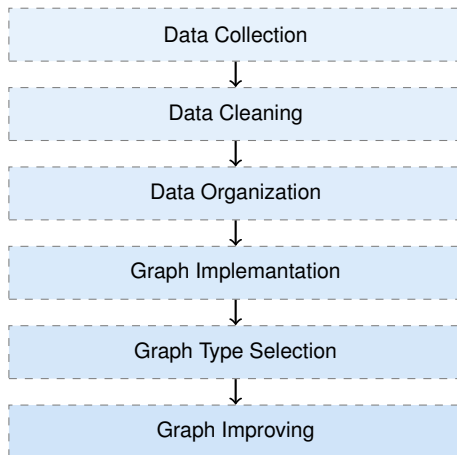


Fig. 6. Steps of graph visualization of cyber threat intelligence data

A. Data Collection

Threat intelligence is data that is collected, processed, and analyzed to understand a threat actor's motives, targets, and attack behaviors. Obtaining large-scale cyber intelligence data in real time is a challenging process. Especially in a network system, it is more difficult to detect the attack data from the data traffic flowing and to capture and collect only the relevant parts. In addition, the intelligibility of these data is another challenge. Therefore, in the collection of cyber intelligence data, it is first necessary to determine what type of data will be collected from which source or sources. For this, it is necessary to have a deep and comprehensive networking knowledge. One of the frequently repeated quotes in network analysis is this: Before you connect the dots, you need to collect them. Your first challenge may be to determine what data to collect.

Cyber threat intelligence data can be versatile according to the threat element. These datasets usually contain such as URL, host, IP address, e-mail account, hashes (MD5, SHA1, and SHA256), common vulnerabilities and exposures, registry, file names ending with specific extensions, and the program database path. This information can be found in different log files kept by the server and network devices [19].

B. Data Cleaning

After collecting complex and large-scale cyber threat intelligence data, the next step is to clean the data and make it usable. Cleaning up text-based cyber intelligence data in different formats, large sizes and containing various parameters is a very difficult process. This data, where log files are actively used, is frustratingly messy. Unfortunately, most graph software tools are not designed to work with this messy data and must be cleaned and prepared before reading the data into graph software.

The following points should be provided regarding the elimination and cleaning of the difficulties experienced in the cleaning process of the complex data obtained;

- *Inconsistent Node Names*: A node should not be represented by more than one name.
- *Refreshed Nodes*: Each node should appear only once in the node dataset.
- *Refreshed Edges*: Some types of graph visualization and analysis software do not work well with multiple edges between the same pair of nodes and need to be consolidated.
- *Self-Loop*: Some graph software do not handle self-loops.
- *Isolated Node*: Datasets may have nodes with no connections, disconnected nodes may cause problems with graph visualization.
- *Edges Connected to Non-Existing Nodes*: In some datasets, an edge can be defined between two nodes where one of the nodes is not in the node list.
- *Invalid Data*: Real world data can have null or invalid data. The numeric data column can have text entries such as N/A or ERROR. These entries should be cleaned or removed.
- *Units*: All numeric data needs to be normalized to the same numeric units.

There are many approaches to correcting invalid, incomplete and inconsistent data. A simple approach would be to remove the specific problematic record, but it may be more beneficial to use other approaches that involve entering missing values or normalizing the data. Cyber attack data is generally not regular data because it is very repetitive and different methods (VPN, etc.) are used to deflect the source. In addition to the steps we mentioned above to clean this data, the relationships must be captured correctly so that the source device or devices can be expressed clearly.

C. Data Organization

It is very effective to define and organize complex and big data into a set of nodes and edges. This clear separation will enable data exploration with a wider variety of tools. Cyber-attack data consists of attacking devices, attacked devices and their relationships and attacks. For example, we can organize our data by expressing the attacking and the attacked devices as nodes, the relationships between these devices and the attack data as weighted edges.

D. Graph Implementation

After the graph data is cleaned and organized, it can be analyzed with software tools [21], [22], [23] where graph analyzes are made, or it can be analyzed and visualized by coding with programming languages. The next goal is to make the graph more understandable. In this context, the following questions are examined and the process proceeds.

- Are the assigned nodes all interconnected or in many separate pieces?
- Is the resulting graph a hierarchy?
- Is the resulting graph sparse? Or is it heavily linked?
- Are there any obvious clusters in the resulting graph?

Statistics can provide large and multidimensional information using data and answer questions about size, density, and

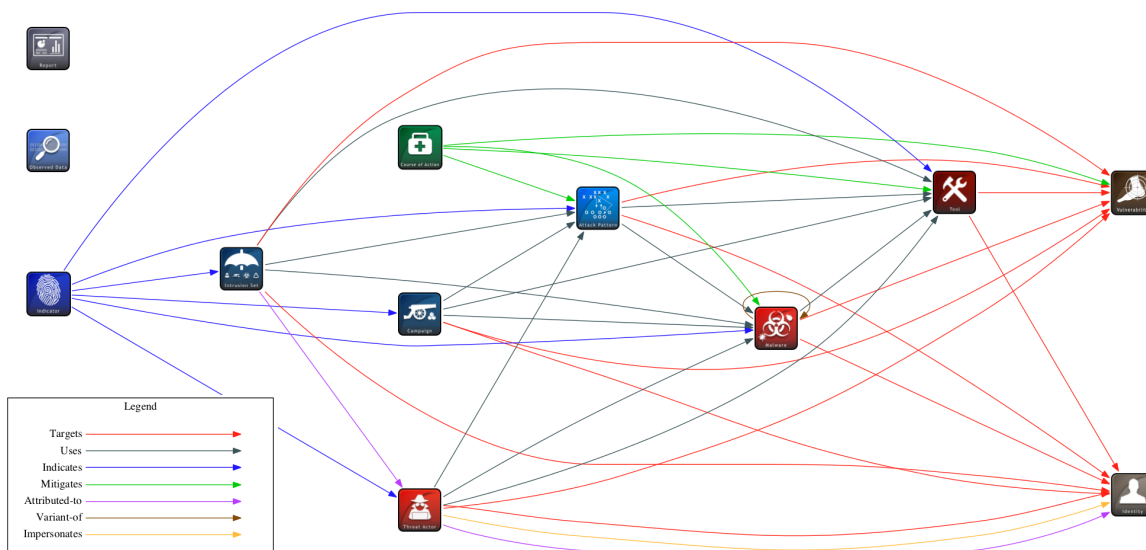


Fig. 7. Visualized STIX domain object relationships [20]

number of discrete graphs. Layouts are an important visual technique for understanding graph structure. Different layouts will reveal different aspects of the graph, allow for different types of analysis, and support different types of stories. A wide variety of node and link layouts can provide different ways of revealing links, groupings, and sequences in graphs. Other graph layout types focus on other properties of the graph, displaying lowest values, hierarchies, or multiple attributes.

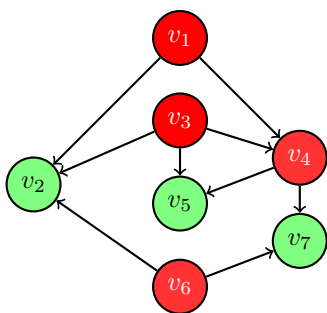


Fig. 8. A simple cyber attack graph

E. Graph Type Selection

During the graph type selection phase, graph images suitable for the type, property and size of the data can be determined. Graph types are increasing day by day. These can be graph visualization types such as Circle Graph [12], Sankey Diagram [24], Hive Graph[11], it is necessary to choose the graph that is most suitable for our data. We can choose different types of graphs according to the desired features in highlighting the data. For example, when cyber threat intelligence data is expressed as a hive graph, only attack density can be analyzed. Relationships can be seen more clearly when using the circle graph.

F. Graph Improving

In the graph visualization process, since we have chosen a graph that is suitable for our data, we can now improve the features we want to emphasize, with methods such as coloring, ghost effect, fading, labeling so that our graph can become more understandable. In this processes, the symbols of the related device or devices can be used in the symbols of the nodes. Thus, the graphs to be created will be more meaningful and more understandable by people who are not closely interested in the subject.

V. CONCLUSION

Graphs are of great importance as they form a model for solving and analyzing many unsolved problems in history. Today, graphs have become so popular that it is possible to encounter a problem or an algorithm modeled with graphs in almost every field. We encounter graphs in almost all fields such as computer science, social sciences, linguistics, engineering, mathematics, and medicine. Cyber threat intelligence benefits organizations of all shapes and sizes by helping process threat data to better understand their attackers, respond faster to incidents and proactively get ahead of a threat actor’s next move. For SMBs, this data helps them achieve a level of protection that would otherwise be out of reach. On the other hand, enterprises with large security teams can reduce the cost and required skills by leveraging external threat intel and making their analysts more effective. In this study, the visualization of complex cyber threat intelligence data as graphs is examined for accurate analysis of network attacks. Within the scope of graph visualization, colouring, segmentation and presenting with different patterns is a very important and technical issue. In this study, the graph visualization processes of cyber threat intelligence data, in which cyber attacks can be detected, are presented in detail. This study opens the floodgates to the monitoring and visualization of network attacks in the Internet environment, the world’s largest information network. Further

studies could be done on the acquisition of critical real-time data and its visualization with dynamic graphs.

REFERENCES

- [1] Y. Bürhan and R. Daş, "Co-author link prediction from academic databases," *Gazi University Journal of Polytechnic*, vol. 20, no. 4, pp. 787–800, Dec. 2017. [Online]. Available: <http://dergipark.gov.tr/download/article-file/387477>
- [2] L. Yang, E. Cheng, and Z. M. Özsoyoğlu, "Efficient path-based computations on pedigree graphs with compact encodings," *BMC Bioinformatics*, vol. 13, no. S3, p. S14, Dec. 2012. [Online]. Available: <https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-13-S3-S14>
- [3] Q. Guo, X. Qiu, X. Xue, and Z. Zhang, "Syntax-guided text generation via graph neural network," *Science China Information Sciences*, vol. 64, no. 5, p. 152102, May 2021. [Online]. Available: <http://link.springer.com/10.1007/s11432-019-2740-1>
- [4] B. Xie, C. Qi, H. Ben, and W. Yu, "The applications of graph theory in electric network," in *2019 International Conference on Sensing, Diagnostics, Prognostics, and Control (SDPC)*. Beijing, China: IEEE, Aug. 2019, pp. 780–784. [Online]. Available: <https://ieeexplore.ieee.org/document/9168962/>
- [5] D. P. Sinha, "A pairing between graphs and trees," *arXiv:math/0502547*, Oct. 2006, arXiv: math/0502547. [Online]. Available: <http://arxiv.org/abs/math/0502547>
- [6] S. A. M. A. Junid, N. M. Tahir, Z. A. Majid, and M. F. M. Idros, "Potential of graph theory algorithm approach for DNA sequence alignment and comparison," in *2012 Third International Conference on Intelligent Systems Modelling and Simulation*. Kota Kinabalu, Malaysia: IEEE, Feb. 2012, pp. 187–190. [Online]. Available: <http://ieeexplore.ieee.org/document/6169697/>
- [7] J. Zhao, Q. Yan, J. Li, M. Shao, Z. He, and B. Li, "TIMiner: Automatically extracting and analyzing categorized cyber threat intelligence from social data," *Computers & Security*, vol. 95, p. 101867, Aug. 2020. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S0167404820301395>
- [8] OASIS, "STIX™ version 2.0. part 3: Cyber observable core concepts."
- [9] E. Rossi, B. Chamberlain, F. Frasca, D. Eynard, F. Monti, and M. Bronstein, "Temporal graph networks for deep learning on dynamic graphs," *arXiv:2006.10637 [cs, stat]*, Oct. 2020, arXiv: 2006.10637. [Online]. Available: <http://arxiv.org/abs/2006.10637>
- [10] F. Harary and G. Gupta, "Dynamic graph models," *Mathematical and Computer Modelling*, vol. 25, no. 7, pp. 79–87, Apr. 1997. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S0895717797000502>
- [11] M. Krzywinski, I. Birol, S. J. Jones, and M. A. Marra, "Hive plots—rational approach to visualizing networks," *Briefings in Bioinformatics*, vol. 13, no. 5, pp. 627–644, Sep. 2012. [Online]. Available: <https://academic.oup.com/bib/article-lookup/doi/10.1093/bib/bbr069>
- [12] R. Das and I. Turkoglu, "Creating meaningful data from web logs for improving the impressiveness of a website by using path analysis method," *Expert Systems with Applications*, vol. 36, no. 3, Part 2, pp. 6635–6644, Apr. 2009. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0957417408005952>
- [13] S. Noel, E. Harley, K. Tam, M. Limiero, and M. Share, "Chapter 4 - cygraph: Graph-based analytics and visualization for cybersecurity," in *Cognitive Computing: Theory and Applications*, ser. Handbook of Statistics, V. N. Gudivada, V. V. Raghavan, V. Govindaraju, and C. Rao, Eds. Elsevier, 2016, vol. 35, pp. 117–167. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0169716116300426>
- [14] "Global international migration flows | Wittgenstein Centre." [Online]. Available: http://download.gsb.bund.de/BIB/global_flow/
- [15] G. Chen, "Information fusion and visualization of cyber-attack graphs," *SPIE Newsroom*, 2007. [Online]. Available: <http://www.spie.org/x14562.xml>
- [16] M. Alshammari and A. Rezgüi, "An all pairs shortest path algorithm for dynamic graphs," *International Journal of Mathematics and Computer Science*, p. 20, 2020.
- [17] J. R. Nascimento, U. S. Souza, and J. L. Szwarcfiter, "Partitioning a graph into complementary subgraphs," *Graphs and Combinatorics*, vol. 37, no. 4, pp. 1311–1331, Jul. 2021. [Online]. Available: <https://link.springer.com/10.1007/s00373-021-02319-4>
- [18] S. V. Patil and D. B. Kulkarni, "K-way spectral graph partitioning for load balancing in parallel computing," *Bharati Vidyapeeth's Institute of Computer Applications and Management*, Aug. 2021. [Online]. Available: <https://link.springer.com/10.1007/s41870-021-00777-w>
- [19] M. Baykara, R. Daş, and G. Tuna, "Web sunucu erişim kütüklerinden web ataklarının tespitine yönelik web tabanlı log analiz platformu," *Firat Üniversitesi Mühendislik Bilimleri Dergisi*, vol. 28, pp. 291 – 302, 2016.
- [20] "Visualized SDO relationships," Sep. 2021. [Online]. Available: <https://oasis-open.github.io/cti-documentation/examples/visualized-sdo-relationships>
- [21] S. Majeed, M. Uzair, U. Qamar, and A. Farooq, "Social Network Analysis Visualization Tools: A Comparative Review," in *2020 IEEE 23rd International Multitopic Conference (INMIC)*. Bahawalpur, Pakistan: IEEE, Nov. 2020, pp. 1–6. [Online]. Available: <https://ieeexplore.ieee.org/document/9318162/>
- [22] S. Hussain, L. Muhammad, and A. Yakubu, "Mining social media and DBpedia data using Gephi and R," *Journal of Applied Computer Science & Mathematics*, vol. 12, no. 1, pp. 14–20, 2018. [Online]. Available: http://www.jacsm.ro/view/?pid=25_2
- [23] G. Drakopoulos, A. Baroutiadi, and V. Megalooikonomou, "Higher order graph centrality measures for Neo4j," in *2015 6th International Conference on Information, Intelligence, Systems and Applications (IISA)*, Jul. 2015, pp. 1–6.
- [24] E. Curmi, R. Fenner, K. Richards, J. M. Allwood, B. Bajželj, and G. M. Kopec, "Visualising a stochastic model of californian water resources using sankey diagrams," *Water Resources Management*, vol. 27, no. 8, pp. 3035–3050, Jun. 2013. [Online]. Available: <http://link.springer.com/10.1007/s11269-013-0331-2>



Mücahit Sülü is a lecturer at İnönü University Organized Industrial Zone Vocational School Computer Programming Department. He graduated from the Mathematics Department of the same university in 2007. He continues his doctoral studies at the Department of Software Engineering at Firat University. Current research interests include graph theory, software design and architecture, IoT/M2M applications, software quality and assurance, big data analysis and visualization. He has been actively developing software (web-based automation systems) at İnönü University since 2008.



Resul Das is a full professor and Chair in the Department of Software Engineering, Technology Faculty, Firat University, where he has been a faculty member since 2011. He graduated with B.Sc. and M.Sc. degrees from the Department of Computer Science at the Firat University in 1999 and 2002 respectively. Then he completed his Ph.D degree at the Department of Electrical-Electronics Engineering at the same university in 2008. He served as both lecturer and network administrator at the Department of Informatics in Firat University from 2000 to 2011. In addition, he is the CCNA and CCNP instructor and the coordinator of the Cisco Networking Academy Program since 2002 at this university. He worked between September 2017 and June 2018 as a visiting professor at the Department of Computing Science at the University of Alberta, Edmonton, Canada supported by TÜBİTAK-BİDEB 2219 Post-Doctoral Fellowship. He has many journal papers and international conference proceedings. he served as Associate Editor for the Journal of IEEE Access and the Turkish Journal of Electrical Engineering and Computer Science-TUBITAK from 2018 to 2021. He entered the 2% of the "World's Most Influential Scientists" list published by Stanford University researchers in 2020 and 2021. His current research interests include computer networks and security, cyber-security, software architectures, software testing, IoT/M2M applications, complex networks, graph visualization, knowledge discovery, and data fusion.

A New Anonymization Model for Privacy Preserving Data Publishing: CANON

Yavuz Canbay, Seref Sagiroglu, Yilmaz Vural

Abstract—Data privacy is a challenging trade-off problem between privacy preserving and data utility. Anonymization is a fundamental approach for privacy preserving and also a hard trade-off problem. It enables to hide the identities of data subjects or record owners and requires to be developed near-optimal solutions. In this paper, a new multidimensional anonymization model (CANON) that employs vantage-point tree (VP-tree) and multidimensional generalization for greedy partitioning and anonymization, respectively, is proposed and introduced successfully for the first time. The main concept of CANON is inspired from Mondrian, which is an anonymization model for privacy preserving data publishing. Experimental results have shown that CANON takes data distribution into consideration and creates equivalence classes including closer data points than Mondrian. As a result, CANON provides better data utility than Mondrian in terms of GCP metric and it is a promising anonymization model for future works.

Index Terms—Data privacy, anonymization, data publishing, CANON.

I. INTRODUCTION

WITH the development in technology, the amount of data is increasing day by day. Internet of things, smart grid, wearable devices, mobile applications, social media, smart cities, e-commerce, health applications, smartphones etc. enable to collect more data than ever.

Today, privacy is a hot topic especially in the digital world. Any violation on sensitive data causes harm on the reputation of individual and they also may lead to discrimination. Hence, protecting privacy of individual in real and digital world is important and requires more effort [1].

Data holders or curators publish data publicly or with a limited set of researchers [2–4]. However, if the data contains sensitive information about individuals (e.g. genomic information [5]), privacy concern becomes one of the major issues to be addressed [6]. “Informational self-determination” [7] and “the appropriate use of responders’ information and the ability to decide what information of a responder goes where” [8] are some of the definitions for data privacy in the literature. Recently, due to the increase in the collection of person-specific information, data privacy has become a major

© **Yavuz CANBAY** (corresponding author) is with the Department of Computer Engineering, Faculty of Engineering and Architecture, Sıtcu Imam University, Kahramanmaraş, Turkey e-mail: yavuzcanbay@ksu.edu.tr

© **Seref SAGIROGLU** is with the Department of Computer Engineering, Faculty of Engineering, Gazi University, Ankara, Turkey e-mail: ss@gazi.edu.tr

© **Yilmaz VURAL** is with the Department of Computer Science, College of Engineering, University of California, Santa Barbara, USA, e-mail: yilmazvural@ucsb.edu

Manuscript received Jan 23, 2022; accepted July 17, 2022.
DOI: [10.17694/bajece.1061910](https://doi.org/10.17694/bajece.1061910)

need and a requirement for data publishing or data mining [9, 10].

Anonymization is a utility-based privacy preserving approach that hides the identity of data subject and, in meantime, provides data utility [11]. In the literature, there exist some anonymization models which enable data curators to publish sensitive data while preserving data privacy. k -anonymity, l -diversity and t -closeness are the most known and frequently used privacy preserving models in data publishing [12]. These models are explained briefly as below.

- k -anonymity ensures that a record in any equivalence class is similar to at least $k - 1$ other records within the same equivalence class and it provides a solution for record linkage attack [13].
- l -diversity guarantees the diversity of sensitive data in each equivalence class and proposes a solution for attribute linkage attack [14].
- t -closeness provides a balance for the distribution of sensitive data between an equivalence class and entire table, and also presents a solution for skewness attack [15].

The reference [16] provides a comprehensive list of attacks and preserving models for privacy. In addition to those, differential privacy is introduced by Dwork et al. [17] as a recent solution. Differential privacy guarantees the results of any analysis will be almost the same if an individual participates the dataset or not, and it presents a solution to background attacks. While k -anonymity, l -diversity and t -closeness are directly proposed for privacy preserving data publishing, differential privacy is mainly used to perturb the results of statistical queries.

k -anonymity, l -diversity and t -closeness are mostly used in privacy preserving data publishing. Since we think that CANON is a base model for future works, we preferred to apply k -anonymity in this paper. It is well-known that, k -anonymity has an exponential relation between the input size and solution space. This situation proves that k -anonymity is an NP-Hard problem and near-optimal solutions are always required as stated in [18–25].

In the literature, there exist many algorithms or models achieving k -anonymity which are compared and presented in Table I. The evaluations of these algorithms or models are summarized and criticized as given below;

- Optimal algorithms require exponential search spaces. Hence, these given algorithms do not provide acceptable solutions in a reasonable time if the size of records increases.

TABLE I: Comparing models/algorithms achieved k -anonymity

Model/Algorithm	Optimality	Dimension	Direction	Partitioning Strategy	Splitting Value Det.
MinGen [26]	Optimal	Single	Bottom-up	Hierarchical	N/A
Incognito [27]	Optimal	Single	Bottom-up	Hierarchical	N/A
Flash [28]	Optimal	Single	Bottom-up	Hierarchical	N/A
Datafly [29]	Near-Optimal	Single	Bottom-up	Hierarchical	N/A
BUG [30]	Near-Optimal	Single	Bottom-up	Hierarchical	N/A
TDS [31]	Near-Optimal	Single	Top-Down	Hierarchical	N/A
Mondrian [32]	Near-Optimal	Multiple	Top-Down	Splitting	Frequency based
CANON (this study)	Near-Optimal	Multiple	Top-Down	Splitting	Distance based

- Near-optimal algorithms provide acceptable solutions in a reasonable time. However, working on single dimension causes more information loss compared to multiple dimensions.
- Hierarchical partitioning requires hierarchy trees. These trees are being constructed by researchers according to their needs under some constraints. Hence these trees may cause undesired information loss because of having a fixed structure. Note that these trees do not present a utility-aware partitioning, they just define the ranges of any related values.
- Among these algorithms, Mondrian stays one step ahead because of supporting multiple dimension. However, it splits data space by employing a frequency based approach which does not consider distribution of data and hence absorbs potential data utility.
- Finally, the presented model (CANON) provides a distance based approach for data space partitioning and increasing data utility by considering data distribution.

CANON is a distribution-aware anonymization model which splits data space by considering the distribution of data points. It creates equivalence classes by grouping closer data points than Mondrian. Hence, CANON provides more data utility compared to Mondrian.

This paper was organized as follows. In Section II, a literature review was presented. In Section III, Mondrian anonymization model was briefed. Some preliminary information about this study was provided in Section IV. Section V introduced VP-tree based greedy partitioning algorithm. The proposed model was presented and introduced in Section VI. Experimental results conducted in this paper was provided in Section VII. Finally, the conclusion was given in Section VIII.

II. LITERATURE REVIEW

This section briefs some current studies about k -anonymity and Mondrian.

Although k -anonymity was proposed in 2002 by Sweeney [13], it is still employed in many current studies. Kacha et al. [33] proposed a metaheuristic method by using black hole algorithm to provide k -anonymity. They employed adult data set and NCP metric in the experiments and presented a comparison for the results. Finally, it was seen that their model provided more utility than the others. Bhati et al. [34] focused on the anonymization of transport user data. A k -anonymity based anonymization model considering both numerical and categorical data was proposed. In addition, they introduced a

new normalization technique and some utility metrics. A real-world dataset and information distortion metric were utilized in the experiments and then successful results were obtained. Mahanan et al. [35] presented a new k -anonymity algorithm based on a heuristic approach. Their algorithm enhanced the performance of existing optimization algorithms by providing a heuristic search for generalization lattice. Uber, Jester and T-drive datasets were used in the experiments and they provided a comparison for the results of the proposed algorithm and other existing algorithms. Finally, they reported that their algorithm provided an efficient anonymization. Adrew and Karthikeyan [36] combined k -anonymity and laplace differential privacy for big data anonymization. They firstly generalized tabular data and then applied laplace noise to provide differential privacy. Experiments are performed on Adult dataset and then NCP metric was used to measure data utility. They reported that the proposed model achieves better utility than other existing models.

Mondrian, which was introduced in 2006, is still being used by some current studies. For example, Wang et al. [37] enhanced Mondrian by applying Self-Organizing Map, Andrew et al. [38], Nezarat et al. [39] and Ashkouti [40] applied Mondrian for privacy preserving big data publishing. In addition, Tang et al. [41] proposed an extended version of Mondrian. They optimized the lowest value of partitions and provided equivalence classes with lower sizes than classical Mondrian. In the experiments, CM, DM and NCP metrics were employed to measure data utilities and their model presented more utility than classical Mondrian. Liu et al. [42] optimized multidimensional k -anonymity model by enhancing Mondrian model. They focused on attribute weighting and then provided a new algorithm. Census dataset was employed and NCP metric was preferred to measure data utility. In the experiments, the proposed model provided better results than classical Mondrian. Gong et al. [43] considered incomplete data anonymization and proposed two algorithms to achieve high data utility on incomplete data. They used NCP metric to measure data utility and Adult dataset. Finally, it was observed that the proposed algorithm gave better utility than classical Mondrian. Nergiz et al. [44] proposed a multidimensional hybrid k -anonymization algorithm based on Mondrian. They aimed to decrease the negative effect of generalization process on anonymization and introduced a hybrid approach using both generalization and data relocation. Census dataset was employed in the experiments and they obtained successful results. LeFevre et al. [45] proposed three different models based on

Mondrian. In their models, they focused on enhancing classical Mondrian with entropy, least square deviance and imprecision. They employed syntactic, Census and Adult datasets in the experiments and obtained better results than the other models in the literature.

This section indicates that k -anonymity is one of the anonymization models used by current studies and Mondrian is still being used to achieve k -anonymity. Since anonymization is one of the hot topics today, there is always a need for novel models and algorithms. This paper proposes a novel anonymization model, called CANON, which is an alternative to Mondrian. We hope that CANON will be a base model for further studies.

III. MONDRIAN ANONYMIZATION MODEL

Mondrian [32] is a frequently used near-optimal anonymization model in the literature. It splits data space by using two different partitioning strategies; strict and relaxed. While strict partitioning creates non-overlapping regions, relaxed partitioning creates potentially overlapping regions. Both strategies accept that the lower bound is k . In addition, $2d(k-1)+t$ and $2k-1$ are accepted as the upper bounds of strict and relaxed partitioning, respectively [32].

Mondrian is a multidimensional anonymization approach that exploits multidimensional generalization and k -Dimensional tree (KD-tree) space partitioning approach. Multidimensional generalization provides higher data utility than other anonymization operators and KD-tree provides an acceptable time complexity of $O(n \log n)$. However, KD-tree has some general disadvantages or weaknesses [46–49] which can be listed as below;

- 1) frequency based splitting value determination,
- 2) non-flexible space partitioning,
- 3) inefficiency in high dimensional data and,
- 4) producing a high unbalanced tree due to skewed data.

If these weaknesses are evaluated in the context of anonymization, it can be clearly seen that only (1) directly affects data utility. Because, frequency based splitting value determination does not consider data distribution and this causes a decrease in data utility. Therefore, if this weakness can be eliminated, a more powerful anonymization model may be obtained. Hence, CANON focuses to eliminate this weakness and provides more data utility than Mondrian.

IV. PRELIMINARY

This section briefs the hardness of k -anonymity, gives some background information about KD-tree and VP-tree, and finally introduces some definitions used in this paper.

A. On the hardness of k -anonymity

In order to reveal the hardness of anonymization problem, we reviewed some papers and presented a summary of these works in this section. In the literature, there exist a number of papers focusing on the NP-Hardness of k -anonymity. Meyerson and Williams [18] investigated the computational complexity of k -anonymity by using a reduction from k -dimensional perfect matching problem. They indicated that if

there is no restriction on the alphabet size, k -anonymity is NP-Hard for $k \geq 3$ and the maximum number of suppressed cells is $n(m-1)$, where n and m are the number of vertexes and edges, respectively. However, Aggarwal et al. [19, 20] employed a reduction from edge partition into triangles and reduced the alphabet size as 2, but the number of suppressed cells remained as $9m$, where m represents number of triangles. Similarly, Sun et al. [21] employed edge partition into 4-clique and represented an alphabet with the size of 2, but the number of suppressed cells was obtained as $48m$, where m is some integer. On the other hand, the proof presented by Scott et al. [24] used a reduction from c-hitting set problem and indicated that anonymization of k -attribute is also NP-Hard even for $k \geq 2$. Finally, LeFevre et al. [32] used another reduction to prove the NP-Hardness of optimal k -anonymous multidimensional partitioning. They utilized discernibility metric to approximate the optimal solution.

From the literature review, it can be clearly understood that anonymization is a hard problem and near-optimal solutions are always required.

B. KD-tree and VP-tree

Mondrian, which is a frequently preferred near-optimal anonymization model, employs KD-tree for data space partitioning. KD-tree [50] splits data points based on their projections into some lower dimensional spaces. A node in KD-tree contains four information as presented below;

- Dimension: is the axis of dataset will be divided,
- Splitting value: median value of dataset on dimension,
- Left-hand side: the left subtree including the data points which are smaller than or equal to the splitting value,
- Right-hand side: the right subtree including the data points that are greater than the splitting value.

The main steps of a KD-tree construction are shown as below [50];

- 1) choose a dimension,
- 2) calculate frequencies of a dataset on the dimension,
- 3) find median of these frequencies,
- 4) accept this median as a splitting value,
- 5) partition data space into two subspaces either horizontally or vertically according to the splitting value,
- 6) go to step 1 until no data point is left.

A VP-tree [51] is a metric tree and a balanced binary tree that recursively divides the space into two partitions based on a median of the distances between a vantage-point and the others. A hypersphere is employed for splitting data into n -dimensional metric space. In this tree structure, a node contains four information as;

- Vantage-point: a point which is selected from dataset,
- Radius: a distance defining the range of vantage-point,
- Left-hand side: the left subtree including the data points which are smaller than or equal to the radius of a vantage-point and,
- Right-hand side: the right subtree including the data points that are greater than the radius of a vantage-point.

The main steps of a VP-tree construction are presented as below [46], [51–55];

- 1) choose a vantage-point,
- 2) calculate the distances between the vantage-point and the others,
- 3) find the median of these distances,
- 4) accept the median as a splitting value,
- 5) according to the splitting value, partition data space into two subspaces,
- 6) go to step 1 until no data point is left.

The illustrations of KD-tree and VP-tree decompositions on a sample dataset are presented in Figure 1. A sample dataset, a KD-tree decomposition and a VP-tree decomposition are shown in Figure 1a, Figure 1b and Figure 1c, respectively. In KD-tree, the frequencies of data on any dimension are calculated and then the median of these frequencies is determined to split dataset. Therefore, a splitting operation which is regardless of data distribution is performed and this situation leads a decrease in potential data utility. But, VP-tree calculates the distances between a vantage-point and the other data points then partitions data space according to the median of these distances. Hence, a distribution-aware splitting is performed to dataset and better data utility is obtained.

In the literature, there are some works comparing KD-tree and VP-tree [46, 49, 56–58]. Based on the results of these works, it can be seen that VP-tree data structure is more successful than KD-tree. Hence, we proposed a new anonymization model that employs VP-tree for space partitioning and also adopts some basic functions of Mondrian. The proposed model is inspired and borrowed some functionalities from Mondrian, but has some advantages compared to Mondrian.

C. Some definitions for the proposed model

In this section, we borrowed some definitions from [32], redefined them in order to provide the theoretical background of the proposed model. These definitions are given below.

A Multidimensional Cut; is a vantage-point centered cutting that produces two disjoint multisets of points, for a multiset of points P .

Allowable Multidimensional Cut; is the state of being multidimensional partitionability. In a d -dimensional space, a cut for the region R_i with radius r_i is allowable if and only if $Count(R_i.P_i > r_i) \geq k$ and $Count(R_i.P_i < r_i) \geq k$.

Non-allowable Multidimensional Cut; in a d -dimensional space, a cut for the region R_i with radius r_i is non-allowable if $Count(R_i.P_i > r_i) < k$ or $Count(R_i.P_i < r_i) < k$.

Multidimensional Partitioning; means cutting the space into multidimensional sub-regions R_1, \dots, R_n that cover all attribute domains.

Minimal Multidimensional Partitioning; let R_i be the i^{th} region created by a multidimensional partitioning and contains multiset P_i of points. If $|R_i.P_i| \geq k$ then this partitioning is minimal and there exist no allowable multidimensional cut for R_i .

In the proposed model, we accepted that the upper bound of minimal multidimensional partitioning is $2k - 1$ and the lower bound as k since our model inspires from relaxed partitioning strategy of Mondrian [32].

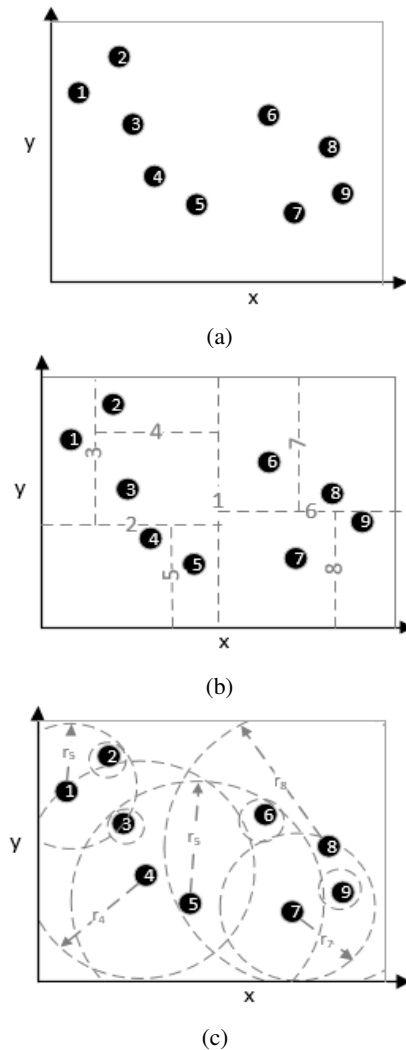


Fig. 1: a) A spatial representation of sample dataset, b) A possible KD-tree decomposition, c) A possible VP-tree decomposition

V. VP-TREE BASED GREEDY PARTITIONING ALGORITHM

In order to partition data space with a proximity based approach, we employed VP-tree in our anonymization model. VP-tree works as the following manner. Firstly, it takes a set of entire records of partition S . If there exists an allowable cut for S , a random element p (vantage-point) is then selected from S . For each $s \in S$, distance $d(p, s)$ is calculated and the median of these distances is then assigned to μ . After that, if an element is smaller than or equal to μ , it is included into the left partition, otherwise the right partition. This iteration continues until no point is left. After the execution of the algorithm, a set of multidimensional regions that the size of each is between k and $2k - 1$ are obtained. The time complexity of the partitioning algorithm is $O(n \log n)$.

Consider any multiset of points as illustrated in Figure 1a. They are labeled with some numbers and have some coordinates such as (x_i, y_i) which can be generalized for higher dimensions. If KD-tree and VP-tree partitioning are applied on these points, two potential partitions can be obtained as

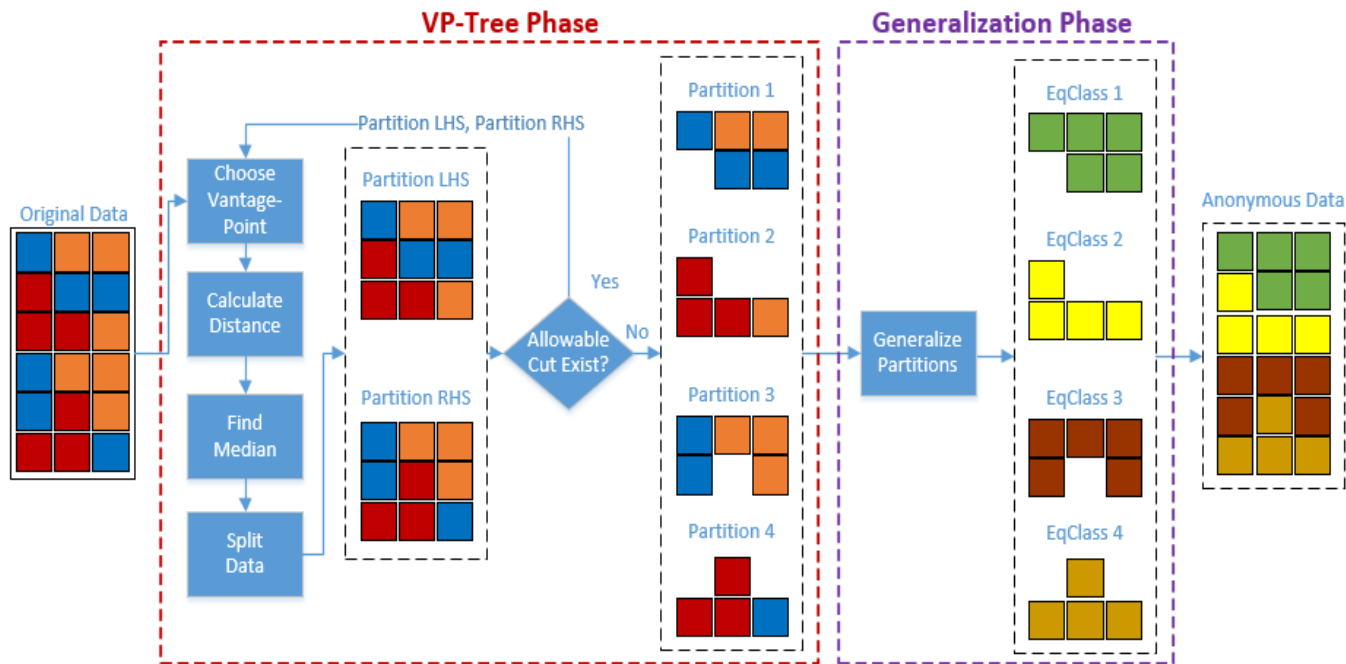


Fig. 4: The block diagram of the proposed model CANON

Adult dataset [59] is frequently used in anonymization studies and includes total 48,842 records. 18,680 records are incomplete and by removing them 30,162 records are obtained and employed in the experimental studies. Since we considered only numerical attributes through this study, we determined the quasi-identifiers as *age*, *final_weight*, *capital_loss*, *capital_gain* and *hours_per_week*.

Diabetes dataset [60] was employed to verify the proposed model. It contains 101,766 records and only numerical attributes are considered. *mean_age*, *num_medications*, *number_outpatient*, *num_lab_procedures*, *num_procedures*, *number_emergency*, *number_inpatient* and *number_diagnosis* were employed as quasi-identifiers.

B. Giving data utility metrics

In this study, we used Discernibility Metric (DM) [61], Average Equivalence Class Size (AECS) [32] and Generalized Certainty Penalty (GCP) [62] metrics to evaluate data utilities produced by the models. In addition, we evaluated the number of equivalence classes (ES) created by the models.

Consider that D is a dataset, EQ is a set of equivalence classes, qid is a quasi-identifier, k is an anonymity level, E is any equivalence class, NCP is normalized certainty penalty metric, n is the number of records in D and d is the dimension of D . DM, AECS and GCP metrics are calculated according to Eq(1), Eq(2) and Eq(3), respectively.

$$DM(D) = \sum_{qid_i} |D[(qid)_i]|^2 \quad (1)$$

$$AECS(D) = \frac{|D|}{(|EQ| * k)} \quad (2)$$

$$GCP(D) = \frac{\sum_{E \in D} |D| * NCP(E)}{(d * n)} \quad (3)$$

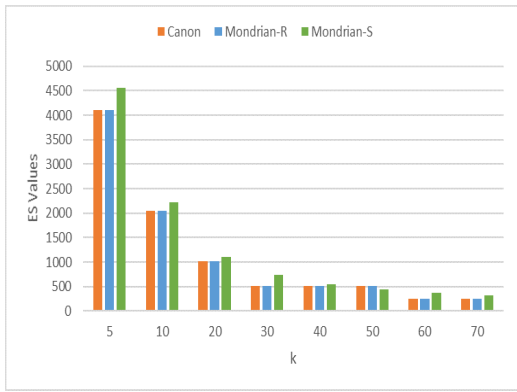
DM, AECS and GCP metrics are used to measure the size of equivalence classes, the size of average equivalence classes and the perimeter of equivalence classes, respectively. It should be emphasized that DM, AECS and GCP metrics are inversely proportional to data utility. Hence lower values of these metrics represent higher data utility and vice versa. In addition, ES values indicate the number of equivalence classes created by algorithm.

Note that CANON almost inspires from relaxed partitioning strategy of Mondrian. Hence, comparing the results of CANON and Mondrian with Relaxed partitioning (Mondrian-R) mainly reveals the success of this paper. But, it is worth to compare CANON to Mondrian with Strict partitioning (Mondrian-S).

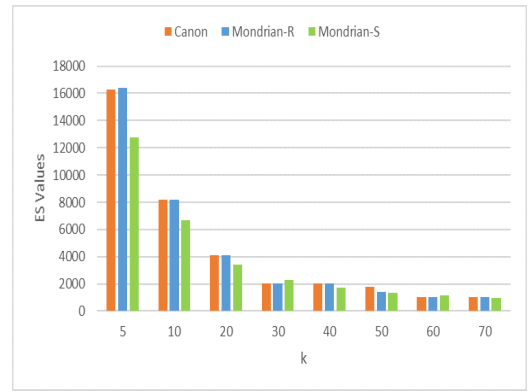
C. Conducting experiments

Two experiments were performed in this section. Experiment 1 and Experiment 2 are conducted to test and verify CANON.

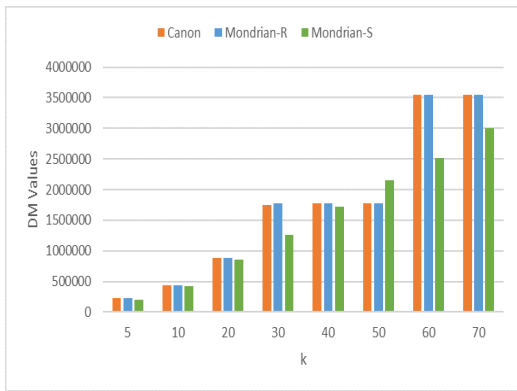
Experiment 1: Adult dataset is employed to test CANON, Mondrian-R and Mondrian-S. As presented on the left of Figure 5, DM, AECS and ES values for Mondrian-R and CANON are almost the same. The main reason is that both CANON and Mondrian-R have the same lower and upper bounds as mentioned earlier. Since DM, AECS and ES measure the size of equivalence classes, the size of average equivalence classes and the number of equivalence classes, these metrics should give almost the same results for both models. The main contribution of this paper is that CANON presents a proximity-aware anonymization and also considers data distribution with



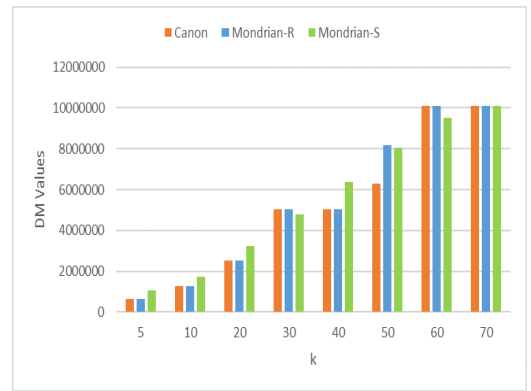
(a) ES Values for Adult Dataset



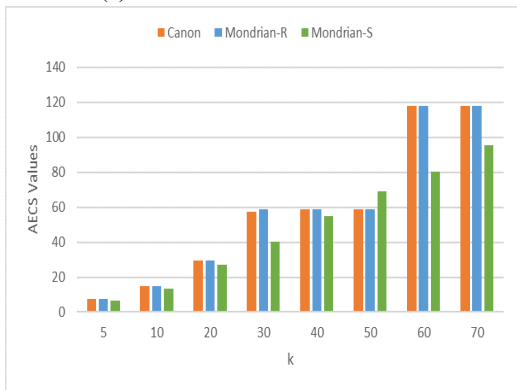
(b) ES Values for Diabetes Dataset



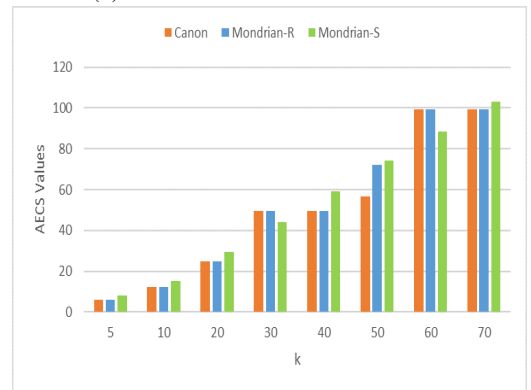
(c) DM Values for Adult Dataset



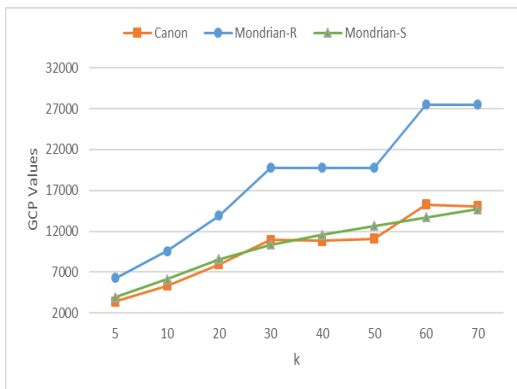
(d) DM Values for Diabetes Dataset



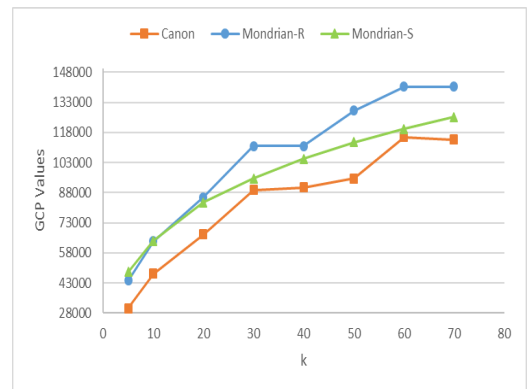
(e) AECS Values for Adult Dataset



(f) AECS Values for Diabetes Dataset



(g) GCP Values for Adult Dataset



(h) GCP Values for Diabetes Dataset

Fig. 5: Experimental Results

a distance based manner. Therefore, CANON proposes higher data utility than Mondrian-R in terms of GCP metric. However, Mondrian-S provides generally higher ES and lower DM and AECS values than other two models. If a comparison is performed for GCP metric, it can be seen that CANON also presents higher data utility than Mondrian-S, for some k values.

Experiment 2: Diabetes dataset was employed to verify the performance of the proposed model. In the right of Figure 5, the results have shown that CANON and Mondrian-R present almost the same ES, DM and AECS values, but Mondrian-S provides different values than these two models. Again, GCP metric is the key factor to evaluate the performance of CANON. The GCP results of CANON show lower values than other two models. Hence, CANON provides better data utility than Mondrian-R and Mondrian-S for different k values.

Lower GCP values represent that closer data points are located in each equivalence classes. Therefore, CANON creates equivalence classes with closer data points compared to Mondrian-R and Mondrian-S.

VIII. CONCLUSION

This paper successfully introduces a new multidimensional anonymization model called CANON. The proposed model uses VP-tree and multidimensional generalization for greedy partitioning and anonymization, respectively.

CANON is a VP-tree based anonymization model by employing a distance-based partitioning and distribution-aware splitting. As can be seen from the results of two datasets that the proposed model provides higher data utility than Mondrian for both strategies in terms of GCP metric. CANON takes data distribution into consideration and creates equivalence classes including closer data points compared to Mondrian, and provides better data utility.

For Adult dataset; CANON presents 45.47% and 13.55% higher data utility than Mondrian for both strategies in terms of GCP metric. For Diabetes dataset; CANON presents 31.01% and 37.04% higher data utility than Mondrian for both strategies in terms of GCP metric.

The proposed model introduced in this article is promising for future works, it can be accepted as a base model for anonymization and further studies may extend the proposed model to obtain higher utilities.

ACKNOWLEDGMENT

Code availability: <https://github.com/ycanbay/canon>
 Authors' contributions: Yavuz Canbay: Conceptualization, Methodology, Software, Validation, Writing - original draft, Writing - review and editing. Seref Sagiroglu: Supervision, Methodology, Writing - original draft, Writing - review and editing. Yilmaz Vural: Software, Validation, Writing - original draft, Writing - review and editing.

REFERENCES

- [1] W. Fang, X. Wen, Y. Zheng, M. Zhou, "A Survey of Big Data Security and Privacy Preserving", *IETE Technical Review*, vol. 34, pp. 544-560, 2017.
- [2] A. Hasan, Q. Jiang, "A General Framework for Privacy Preserving Sequential Data Publishing", *International Conference On Advanced Information Networking And Applications Workshops*, pp. 519-524, Taipei, Taiwan, 2017.
- [3] M. Almasi, T. Siddiqui, N. Mohammed, H. Hemmati, "The Risk-Utility Tradeoff for Data Privacy Models", *International Conference On New Technologies, Mobility And Security*, pp. 1-5, Larnaca, Cyprus, 2016.
- [4] X. Chen, V. Huang, "Privacy Preserving Data Publishing for Recommender System", *IEEE Annual Computer Software And Applications Conference Workshops*, pp. 128-133, Izmir, Turkey, 2012.
- [5] M. Akgun, "An Active Genomic Data Recovery Attack", *Balkan Journal of Electrical and Computer Engineering*, pp. 417-423, 2019.
- [6] R. Wang, Y. Zhu, C. Chang, Q. Peng, Q. "Privacy-preserving High-dimensional Data Publishing for Classification", *Computers And Security*, vol. 93, 2020.
- [7] M. Chibba, A. Cavoukian, "Privacy, Consumer Trust and Big Data: Privacy by Design and the 3 C's", *ITU Kaleidoscope: Trust In The Information Society*, pp. 1-5, 2015.
- [8] P. Jain, M. Gyanchandani, N. Khare, "Big Data Privacy: A Technological Perspective and Review", *Journal Of Big Data*, vol. 3, no. 25, 2016.
- [9] J. Nayahi, V. Kavitha, "Privacy and utility preserving data clustering for data anonymization and distribution on Hadoop". *Future Generation Computer Systems*, vol. 74, pp. 393-408, 2017.
- [10] Q. Tang, Y. Wu, S. Liao, X. Wang, "Utility-based k-Anonymization", *International Conference On Networked Computing And Advanced Information Management*, pp. 318-323, Seoul, 2010.
- [11] B. Fung, K. Wang, A. Fu, S. Philip, "Introduction to Privacy-Preserving Data Publishing: Concepts and Techniques", CRC Press, 2010.
- [12] B. Fung, K. Wang, R. Chen, P. Yu, "Privacy-Preserving Data Publishing: A Survey of Recent Developments", *Computing Surveys*, vol. 42, no. 14, 2010.
- [13] L. Sweeney, "k-Anonymity: A Model for Protecting Privacy", *International Journal Of Uncertainty, Fuzziness And Knowledge-Based Systems*, vol. 10, pp. 557-570, 2002.
- [14] A. Machanavajjhala, J. Gehrke, D. Kifer, M. Venkatasubramanian, "l-Diversity: Privacy Beyond k-Anonymity", *International Conference on Data Engineering*, Atlanta, USA, 2006.
- [15] N. Li, T. Li, S. Venkatasubramanian, "t-Closeness: Privacy Beyond k-Anonymity and l-Diversity", *IEEE International Conference On Data Engineering*, pp. 106-115, Istanbul, Turkey, 2007.
- [16] S. Abdelhameed, S. Moussa, M. Khalifa, "Privacy-Preserving Tabular Data Publishing: A Comprehensive Evaluation From Web to Cloud", *Computers And Security*, vol. 72, pp. 74-95, 2017.
- [17] C. Dwork, "Differential Privacy", *International Colloquium On Automata, Languages And Programming*, pp. 1-12, Venice, Italy, 2006.

- [18] A. Meyerson, R. Williams, "On the Complexity of Optimal k-Anonymity", ACM SIGMOD-SIGACT-SIGART Symposium On Principles Of Database Systems, pp. 223-228, Paris, France, 2004.
- [19] G. Aggarwal, T. Feder, K. Kenthapadi, R. Motwani, R. Panigrahy, D. Thomas, A. Zhu, "Approximation Algorithms for k-Anonymity", *Journal Of Privacy Technology*, pp. 1-18, 2005.
- [20] G. Aggarwal, T. Feder, K. Kenthapadi, R. Motwani, R. Panigrahy, D. Thomas, A. Zhu, "Anonymizing Tables", International Conference On Database Theory, pp. 246-258, Edinburgh, UK, 2005.
- [21] X. Sun, L. Sun, H. Wang, "Extended k-Anonymity Models Against Sensitive Attribute Disclosure", *Computer Communications*, vol. 34, pp. 526-535, 2011.
- [22] P. Bonizzoni, G. Della Vedova, R. Dondi, "Anonymizing Binary and Small Tables is Hard to Approximate", *Journal Of Combinatorial Optimization*, vol. 22, pp. 97-119, 2011.
- [23] J. Blocki, R. Williams, "Resolving the Complexity of Some Data Privacy Problems", International Colloquium On Automata, Languages, And Programming, pp. 393-404, Bordeaux, France, 2010.
- [24] A. Scott, V. Srinivasan, U. Stege, "k-Attribute-Anonymity is Hard Even for k=2", *Information Processing Letters*, vol. 115, pp. 368-370, 2015.
- [25] R. Chen, B. Fung, N. Mohammed, B. Desai, K. Wang, "Privacy-preserving trajectory data publishing by local suppression", *Information Sciences*, vol. 231, pp. 83-97, 2013.
- [26] L. Sweeney, "Achieving k-Anonymity Privacy Protection Using Generalization and Suppression", *International Journal Of Uncertainty, Fuzziness And Knowledge-Based Systems*, vol. 10, pp. 571-588, 2002.
- [27] K. LeFevre, D. DeWitt, R. Ramakrishnan, "Incognito: Efficient Full-domain k-Anonymity", ACM SIGMOD International Conference On Management Of Data, pp. 49-60, Baltimore, Maryland, 2005.
- [28] F. Kohlmayer, F. Prasser, C. Eckert, A. Kemper, K. Kuhn, "Flash: Efficient, Stable and Optimal k-Anonymity", International Conference On Privacy, Security, Risk And Trust And International Conference On Social Computing, pp. 708-717, Amsterdam, Netherlands, 2012.
- [29] L. Sweeney, "Datafly: A System for Providing Anonymity in Medical Data", *Database Security XI*, pp. 356-381, 1998.
- [30] K. Wang, P. Yu, S. Chakraborty, "Bottom-Up Generalization: A Data Mining Solution to Privacy Protection", IEEE International Conference On Data Mining, pp. 249-256, Brighton, UK, 2004.
- [31] B. Fung, K. Wang, P. Yu, "Top-Down Specialization for Information and Privacy Preservation", International Conference On Data Engineering, pp. 205-216, Tokyo, Japan, 2005.
- [32] K. LeFevre, D. DeWitt, R. Ramakrishnan, "Mondrian Multidimensional k-Anonymity", International Conference On Data Engineering, pp. 25-25, Atlanta, USA, 2006.
- [33] L. Kacha, A. Zitouni, M. Djoudi, "KAB: A new k-anonymity approach based on black hole algorithm", *Journal Of King Saud University-Computer And Information Sciences*, in press, 2021.
- [34] B. Bhati, J. Ivanchev, I. Bojic, A. Datta, D. Eckhoff, "Utility-Driven k-Anonymization of Public Transport User Data", *IEEE Access*, vol. 9, pp. 23608-23623, 2021.
- [35] W. Mahanan, W. Chaovalitwongse, J. Natwichai, "Data privacy preservation algorithm with k-anonymity", *World Wide Web*, vol. 24, pp. 1551-1561, 2021.
- [36] J. Andrew, J. Karthikeyan, "Privacy-preserving big data publication:(K, L) anonymity", *Intelligence In Big Data Technologies—Beyond The Hype*, pp. 77-88, 2021.
- [37] P. Wang, P. Huang, Y. Tsai, R. Tso, "An Enhanced Mondrian Anonymization Model based on Self-Organizing Map", Asia Joint Conference On Information Security, pp. 97-100, Taipei, Taiwan, 2020.
- [38] J. Andrew, J. Karthikeyan, J. Jebastin, "Privacy preserving big data publication on cloud using mondrian anonymization techniques and deep neural networks", International Conference On Advanced Computing And Communication Systems, pp. 722-727, Coimbatore, India, 2019.
- [39] A. Nezarat, K. Yavari, "A distributed method based on mondrian algorithm for big data anonymization", International Congress On High-Performance Computing And Big Data Analysis, pp. 84-97, Tehran, Iran, 2019.
- [40] F. Ashkouti, A. Sheikahmadi, "DI-Mondrian: Distributed improved Mondrian for satisfaction of the L-diversity privacy model using Apache Spark", *Information Sciences*, vol. 546, pp. 1-24, 2021.
- [41] Q. Tang, Y. Wu, X. Wang, "New Algorithm with Lower Upper Size Bound for k-Anonymity", International Conference On Communication Systems, Networks And Applications, pp. 421-424, Hong Kong, China, 2010.
- [42] K. Liu, C. Kuo, W. Liao, P. Wang, "Optimized Data de-Identification Using Multidimensional k-Anonymity", International Conference On Trust, Security And Privacy In Computing And Communication, International Conference On Big Data Science And Engineering, pp. 1610-1614, New York, USA, 2018.
- [43] Q. Gong, M. Yang, Z. Chen, J. Luo, "Utility Enhanced Anonymization for Incomplete Microdata", International Conference On Computer Supported Cooperative Work In Design, pp. 74-79, Nanchang, China, 2016.
- [44] M. Nergiz, M. Gök, "Hybrid k-Anonymity", *Computers And Security*, vol. 44, pp. 51-63, 2014.
- [45] K. LeFevre, D. DeWitt, R. Ramakrishnan, "Workload-aware Anonymization", ACM SIGKDD International Conference On Knowledge Discovery And Data Mining, pp. 277-286, 2006.
- [46] N. Kumar, L. Zhang, S. Nayar, "What is a Good Nearest Neighbors Algorithm for Finding Similar Patches in Images", Springer, 2008.
- [47] M. Dolatshah, A. Hadian, B. Minaei-Bidgoli, "Ball*-Tree: Efficient Spatial Indexing for Constrained Nearest-Neighbor Search in Metric Spaces", Preprint ArXiv:1511.00628, 2015.

- [48] I. Witten, E. Frank, M. Hall, C. Pal, "Data Mining: Practical Machine Learning Tools and Techniques", Morgan Kaufmann, 2016.
- [49] J. Wang, N. Wang, Y. Jia, J. Li, G. Zeng, H. Zha, X. Hua, "Trinary-Projection Trees for Approximate Nearest Neighbor Search", *IEEE Transactions On Pattern Analysis And Machine Intelligence*, vol. 36, pp. 388-403, 2014.
- [50] J. Bentley, "Multidimensional Binary Search Trees Used for Associative Searching", *Communications Of The ACM*, vol. 18, pp. 509-517, 1975.
- [51] P. Yianilos, "Data Structures and Algorithms for Nearest Neighbor Search In General Metric Spaces", Symposium On Discrete Algorithms, vol. 93, pp. 311-321, Texas, USA, 1993.
- [52] A. Fu, P. Chan, Y. Cheung, Y. Moon, "Dynamic Vp-Tree Indexing for N-Nearest Neighbor Search Given Pair-Wise Distances", *International Journal On Very Large Data Bases*, vol. 9, pp. 154-173, 2000.
- [53] T. Bozkaya, M. Ozsoyoglu, "Distance-Based Indexing for High-Dimensional Metric Spaces", ACM SIGMOD International Conference On Management Of Data, pp. 357-368, 1997.
- [54] T. DeFreitas, H. Saddiki, P. Flaherty, "GEMINI: A Computationally-Efficient Search Engine for Large Gene Expression Datasets", *BMC Bioinformatics*, vol. 17, pp. 102, 2016.
- [55] C. Böhm, S. Berchtold, D. Keim, "Searching in High-Dimensional Spaces: Index Structures for Improving the Performance of Multimedia Databases", *ACM Computing Surveys*, vol. 33, pp. 322-373 2001.
- [56] A. Kuznetsov, E. Myasnikov, "Copy-Move Detection Algorithm Efficiency Increase Using Binary Space Partitioning Trees", CEUR Workshop Proceedings, pp. 373-378, 2016.
- [57] F. Zhang, P. Di, H. Zhou, X. Liao, J. Xue, "RegTT: Accelerating Tree Traversals on GPUs by Exploiting Regularities", International Conference On Parallel Processing, pp. 562-571, Philadelphia, USA, 2016.
- [58] T. Kristensen, C. Pedersen, "Data Structures for Accelerating Tanimoto Queries on Real Valued Vectors", International Workshop On Algorithms In Bioinformatics, pp. 28-39, Liverpool, UK, 2010.
- [59] D. Dheeru, E. Taniskidou, UCI Machine Learning Repository, <http://archive.ics.uci.edu/ml>
- [60] B. Strack, J. DeShazo, C. Gennings, J. Olmo, S. Ventura, K. Cios, J. Clore, "Impact of HbA1c measurement on hospital readmission rates: analysis of 70,000 clinical database patient records", BioMed Research International, 2014.
- [61] A. Skowron, C. Rauszer, "The Discernibility Matrices and Functions in Information Systems", Intelligent Decision Support, pp. 331-362, 1992.
- [62] G. Ghinita, P. Karras, P. Kalnis, N. Mamoulis, "Fast Data Anonymization with Low Information Loss", International Conference On Very Large Databases, pp. 758-769, Vienna, Austria, 2007.



Yavuz CANBAY received his Ph.D. degree from Department of Computer Engineering, Gazi University in 2019. Currently, he is an assistant professor in Kahramanmaraş Sutcu Imam University. Data privacy, big data analytics and information security are his main research interests.



Seref Sagiroglu is a professor in the Department of Computer Engineering at Gazi University. He is the editor of International Journal of Information Security Science. His research interests include machine learning, intelligent system identification, recognition, modeling and control, artificial intelligence, software engineering, information and computer security, biometry, malware detection, big data analytics, cyber security.



Yilmaz Vural is a researcher in the Department of Computer Science at University of California, Santa Barbara. He received his Ph.D. Degree in Computer Engineering from Hacettepe University. His research interests include data privacy, information and computer security.

Reduction of Output Impedance of Buck Converter with Genetic Algorithm

Farzin Asadi

Abstract—This paper introduces a technique to reduce the output impedance in the PWM buck converters with voltage-mode control (VMC) without requiring low Equivalent Series Resistance (ESR) output capacitors. Proposed technique uses the infinity norm (H_{∞}) to convert the problem into an optimization problem. Obtained optimization problem is solved with the aid of Genetic Algorithm (GA). The proposed technique is applied to a sample buck converter operating in Continuous Conduction Mode (CCM). Simulink simulation is used to test the suggested method. Simulation results showed a considerable decrease in the low frequency region of output impedance. Such a decrease in output impedance is very desired for low voltage high current loads like computer CPU's.

Index Terms— DC-DC converters, load transients, output impedance of buck converter, State Space Averaging (SSA).


I. INTRODUCTION

MANY ATTEMPTS have been made to improve the dynamic response of PWM buck converters to a step change in load current. In [1] an averaged model of an active clamp buck converter was obtained. In this paper, output impedance of the active clamp buck converter has been modeled including the clamp capacitor effect. In [2] V2 control is used to achieve fast transient response. V2 control can take advantage of the instant feedback of the output voltage during the load transient and delay of the error amplifier is eliminated. In [3] the effects of various control techniques on the transient response of switching power supply have been discussed and compared. In this paper, transient response of switching power supply is improved by using double voltage loop control technique. In [4] the transient response of the Voltage Regulator Module (VRM) output voltage when the processor has a fast load change is analyzed. Effect of parasitic are considered in this paper. In [5] fast and efficient controller for a buck converter which supplies a microprocessor is developed and tested. Experimental data shows: tight static and dynamic regulation ($\pm 55\text{mV}$), fast transient response (within $1\mu\text{s}$), high efficiency (up to 90%), and stable operation with good noise immunity. [6] used hysteretic current-mode control in

Continuous Conduction Mode (CCM) to control a Buck converter which feeds a low voltage microprocessor load. Experimental results for a 5.0 V input, 3.1V output, 13A Buck converter are included to verify the design. In [7] analytical equations are derived for a DC-DC converter at high slew-rate load current transients. The analyzed model includes synchronous buck converter, controller, output capacitor and supply bus parasitic. This paper studies the impacts of system parameters on output voltage transient response. [8] studied the role of the inductor on dynamic response of the converter and used optimization to determine the best value to obtain a fast response for loads like microprocessor. [9] presented a method to synthesis a zero-impedance converter. This method requires both a positive current feedback and a negative voltage feedback for synthesis as well as a current sensing device. [10] presented a method to achieve near optimum dynamic regulation by combining feed-forward of the output current and input voltage with current-mode control (CMC). A common method used in industry to control the output impedance is to use many output filter capacitors placed in parallel to reduce the ESR. In this approach, the feedback compensator is designed to provide the loop gain and phase margin for stability and the peak closed-loop output impedance is achieved through proper selection of low ESR output capacitors.

Feedback control is another technique to decrease the output impedance of a converter. [11] discussed the technical motivation behind compensation, derivation of analytical and design oriented transfer functions. It provides an illustrative example as well. [12] used loop shaping to design an optimum controller for a buck converter. Loop shaping method is easy and intuitive and the controller can be designed accurately for a specific phase margin and cross over frequency. [13] is a tutorial on how to design different types of classic controllers for a Buck converter. [14] presents the internal control loop operation of the BQ 2472x/3x Buck controller IC as well as the external compensator design guideline. It contains a design example based on practical specifications is demonstrated. [15-17] contains some guidelines to select the components of a second order compensation network which controls a buck converter. [18] discussed the digital PI controller design procedure for a buck converter. Experimental results showed the effectiveness of designed controller to keep the output voltage stable despite of disturbances like input voltage variations and output load changes. [19] discussed the procedure to extract the input and output impedance of dc-dc converters. Input and output impedance of buck, buck-boost and boost converters are studied there. Extraction of small

FARZIN ASADI, is with the Department of Electrical Engineering, Maltepe University, Istanbul, Turkey, (e-mail: farzinasadi@maltepe.edu.tr).

 <https://orcid.org/0000-0002-5928-0807>

Manuscript received March 31, 2022; accepted July 18, 2022.

DOI: [10.17694/bajece.1096188](https://doi.org/10.17694/bajece.1096188)

signal model is the first step for designing a linear controller for a converter. [20] discussed the procedure to extract the small signal linear model of dc-dc converters. [21] used a synchronous buck converter with Gallium Nitride (GaN) to design a driver for a laser diode. Use of GaN based switch increased the efficiency to 96.6%.

In this paper effect of parasitic elements on the operation of the converter is studied. This paper introduces a technique to reduce the output impedance in the PWM buck converters. The technique translates the reduction of output impedance into an optimization problem. Obtained optimization problem is solved with the aid of GA.

This paper is organized as follows: Dynamics of buck converter is studied in the second section. Small signal transfer functions of buck converter are extracted with the aid of State Space Averaging (SSA). Third section introduces the suggested technique. Simulink simulations are done in the fourth section. Finally, suitable conclusions are drawn.

II. DYNAMICS OF THE BUCK CONVERTER

The circuit diagram of an open-loop buck converter is shown in Fig. 1. The buck converter composed of two switches, a MOSFET switch and a diode. In this schematic, V_g , r_g , L , r_L , C , r_C and R shows input DC source, input DC source internal resistance, inductor, inductor ESR, capacitor, capacitor ESR and load, respectively. i_o is a fictitious current source added to the schematic in order to calculate the output impedance of converter. In this section we assume that converter works in CCM. MOSFET switch is controlled with the aid of a Pulse Width Modulator (PWM) controller. MOSFET switch keeps closed for $D.T$ seconds and $(1 - D).T$ seconds open. D and T show duty ratio and switching period, respectively.

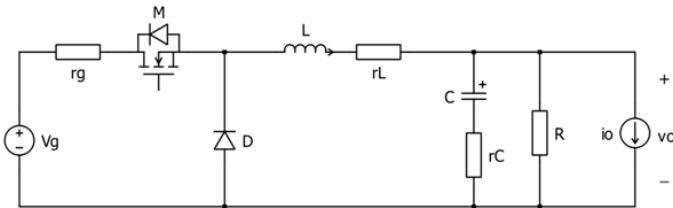


Fig. 1. Schematic of a buck converter.

When MOSFET is closed, the diode is opened (Fig. 2).

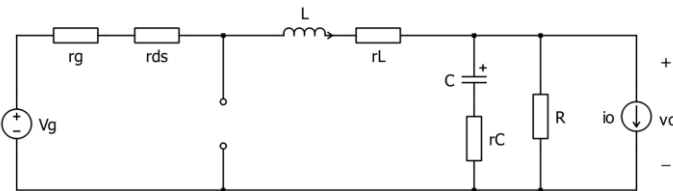


Fig. 2. Equivalent circuit of a buck converter for closed MOSFET.

The circuit differential equations can be written as:

$$\left(\frac{di_L(t)}{dt} = \frac{1}{L} \left(- \left(r_g + r_{ds} + r_L + \frac{R \times r_C}{R + r_C} \right) i_L - \frac{R}{R + r_C} v_C + \frac{R \times r_C}{R + r_C} i_o + v_g \right) \right) \quad (1)$$

$$\left(\frac{dv_C(t)}{dt} = \frac{1}{C} \left(\frac{R}{R + r_C} i_L - \frac{1}{R + r_C} v_C - \frac{R}{R + r_C} i_o \right) \right) \quad (2)$$

$$v_o = r_C C \frac{dv_C}{dt} + v_C = \frac{R \times r_C}{R + r_C} i_L + \frac{R}{R + r_C} v_C - \frac{R \times r_C}{R + r_C} i_o \quad (3)$$

When MOSFET is opened, the diode is closed (Fig. 3).

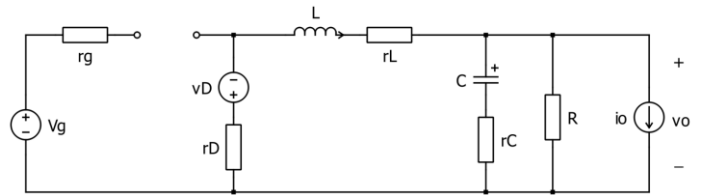


Fig. 3. Equivalent circuit of a buck converter for opened MOSFET.

The circuit differential equations can be written as:

$$\left(\frac{di_L(t)}{dt} = \frac{1}{L} \left(- \left(r_D + r_L + \frac{R \times r_C}{R + r_C} \right) i_L - \frac{R}{R + r_C} v_C + \frac{R \times r_C}{R + r_C} i_o - v_D \right) \right) \quad (4)$$

$$\left(\frac{dv_C(t)}{dt} = \frac{1}{C} \left(\frac{R}{R + r_C} i_L - \frac{1}{R + r_C} v_C - \frac{R}{R + r_C} i_o \right) \right) \quad (5)$$

$$v_o = r_C C \frac{dv_C}{dt} + v_C = \frac{R \times r_C}{R + r_C} i_L + \frac{R}{R + r_C} v_C - \frac{R \times r_C}{R + r_C} i_o \quad (6)$$

State Space Averaging (SSA) can be used to extract the small signal transfer functions of the DC-DC converter. The procedure of state space averaging is explained in detail in [22] and [23].

The MATLAB program applies the SSA procedure to a buck converter with component values as given in Table 1. Component (the capacitor, the inductor and the load resistance) values shown in Table 1 are calculated with the aid of the procedure given in [22]. In this table, for the sake of simplicity, typical values are assumed for diode and MOSFET parameters.

TABLE I
THE BUCK CONVERTER PARAMETERS

	Nominal Value
Output voltage, V_o	1.5 V
Duty ratio, D	0.156
Input DC source voltage, V_g	12 V
Input DC source internal resistance, r_g	0 Ω
MOSFET Drain-Source resistance, r_{ds}	15 m Ω
Capacitor, C	3290 μ F
Capacitor ESR, r_C	1.4 m Ω
Inductor, L	13 μ H
Inductor ESR, r_L	9 m Ω
Diode voltage drop, v_D	0.39 V
Diode forward resistance, r_D	15 m Ω
Load resistor, R	0.2-3 Ω
Nominal load resistor	1 Ω
Switching Frequency, F_{sw}	200 kHz

Following results are obtained after running the MATLAB code:

$$\frac{\tilde{v}_o}{\tilde{i}_o} = 0.0476 \frac{(s + 6079)(s + 1846)}{s^2 + 5799s + 2.28 \times 10^7} \quad (7)$$

$$\frac{\tilde{v}_o}{\tilde{v}_g} = 571.43 \frac{(s + 6079)}{s^2 + 5799s + 2.28 \times 10^7} \quad (8)$$

$$\frac{\tilde{v}_o}{\tilde{d}} = 45385 \frac{(s + 6079)}{s^2 + 5799s + 2.28 \times 10^7} \quad (9)$$

Bode diagram of these transfer functions are shown in Figs. 4, 5 and 6. All of the obtained transfer functions are stable and minimum phase since they don't have any pole or zero in the Right Half Plane (RHP). Based on the obtained results, the dynamic model shown in Fig. 7 can be suggested to the studied converter.

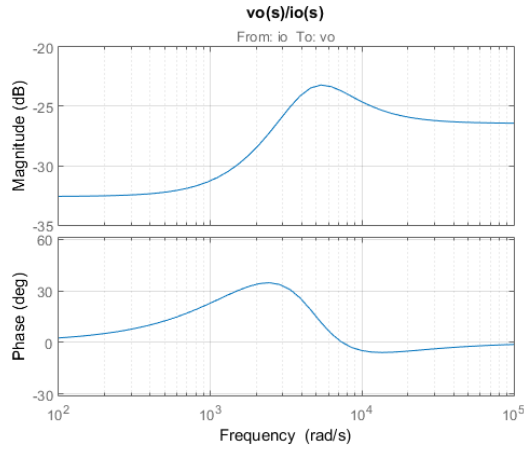


Fig. 4. Bode diagram of $\frac{\tilde{v}_o}{i_o} = 0.0476 \frac{(s+6079)(s+1846)}{s^2+5799s+2.28 \times 10^7}$.

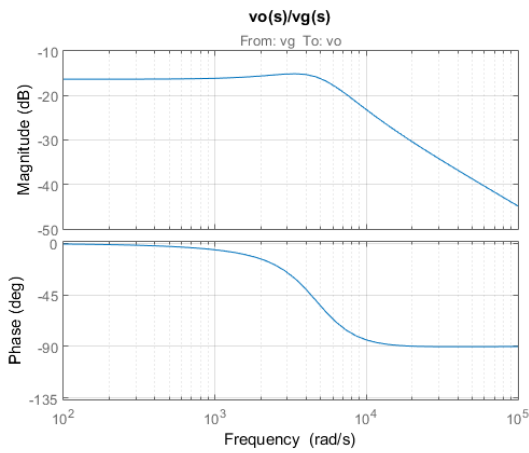


Fig. 5. Bode diagram of $\frac{\tilde{v}_o}{v_g} = 571.43 \frac{(s+6079)}{s^2+5799s+2.28 \times 10^7}$.

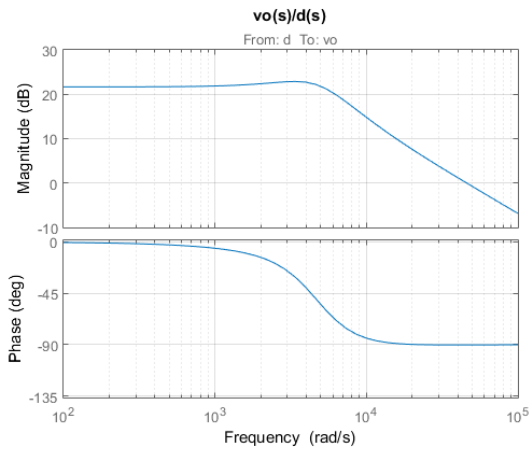


Fig. 6. Bode diagram of $\frac{\tilde{v}_o}{d} = 45385 \frac{(s+6079)}{s^2+5799s+2.28 \times 10^7}$.

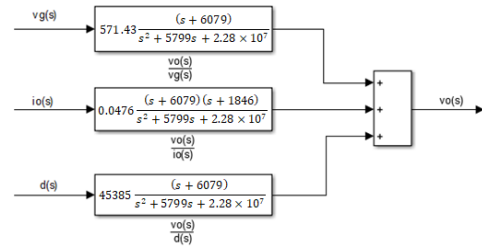


Fig. 7. Dynamical model of the studied buck converter.

III. SUGGESTED METHOD

This section introduces the suggested method. Infinity norm and genetic algorithm make the foundations of the proposed method and are studied briefly.

A. H_∞ Norm

The steady-state output of the SISO stable system,

$$\begin{aligned} \dot{x} &= Ax + Bu \\ y &= Cx \end{aligned} \tag{10}$$

to the input function,

$$u(t) = a \sin(\omega t + \varphi) \tag{11}$$

with unknown $a \neq 0, \omega, \varphi \in \mathbb{R}^1$ is

$$y_{ss}(t) = |G(j\omega)| \cdot a \cdot \sin(\omega t + \varphi + \arg(G(j\omega))) \tag{12}$$

where $G(s) = C(sI - A)^{-1}B$. The H_∞ -norm is the maximal possible amplification, i.e.

$$\begin{aligned} \|G(\cdot)\|_\infty &= \frac{\sup_{\omega \in \mathbb{R}^1} |G(j\omega)|}{\sup_{\omega \in \mathbb{R}^1} |U(s)|} = \sup_{\omega \in \mathbb{R}^1} |G(j\omega)| \end{aligned} \tag{13}$$

The H_∞ -norm of a MIMO system is its maximum singular value [24].

$$\|G(\cdot)\|_\infty = \sup_{\omega} \bar{\sigma} \{G(j\omega)\} \tag{14}$$

The H_∞ -norm of a scalar transfer function can be obtained graphically. Infinity norm of a SISO system G is the distance in the complex plane from the origin to the farthest point on the Nyquist plot of G , and it also appears as the peak value on the Bode magnitude plot of $|G(j\omega)|$. For instance, the infinity norm of $G(s) = \frac{10}{s+0.4s+10}$ is 7.91. Bode plot of $G(s) = \frac{10}{s+0.4s+10}$ is shown in Fig. 8. Maximum of the Bode graph occurs around $\frac{\text{Rad}}{s}$ and its value is 17.96 dB which is equivalent to gain of $10^{\frac{17.96}{20}} = 7.91$.

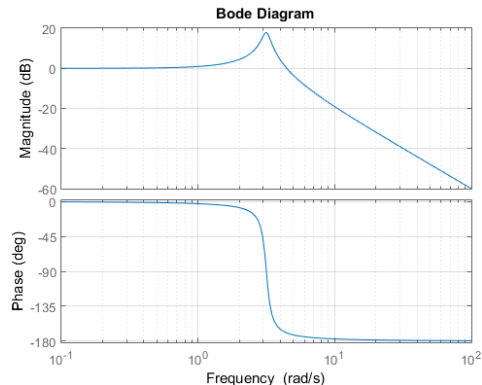


Fig. 8. Bode plot of $G(s) = \frac{10}{s+0.4s+10}$. Peak occurs at 3.15 Rad/s with magnitude of 17.96 dB.

B. Genetic Algorithm (GA)

In 1960 the first serious investigation into Genetic Algorithm (GA) was undertaken by John Holland. These search techniques are based on the process of biological evolution and are used to provide useful solutions to optimization and search problems. As they are based on biological evolution, they use techniques that are emulated from the concepts of inheritance, mutation and selection.

Genetic algorithms explore a parameter space while optimizing a function. The problem is broken down into a *population* of candidate solutions which are refined over a number of *generations*.

A single candidate solution is represented by a *chromosome* which essentially encodes all of the optimizable parameters into a single entity. Each candidate is ranked using a fitness function and those with the best fitness score are selected for further refinement. The refinement stage then operates on each of the chromosomes by

I) *breeding* - a process where a new population of improved candidates are generated using the present generation's population, and

II) *mutation* - in which chromosomes are modified in some way. Both of these operations permit the parameter space to be more effectively explored. The whole process is iterated for many generations where the candidate solutions can be seen to evolve and, hopefully, converge towards a single solution. The use of natural evolution method for the optimization of control system has been of interest for the researchers since a long time. Fig. 9, illustrate the GA flowchart.

Ready to use implementation of GA is available in the MATLAB's optimization toolbox.

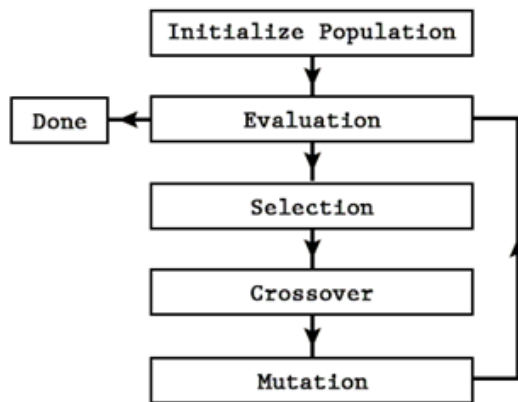


Fig. 9. GA flowchart.

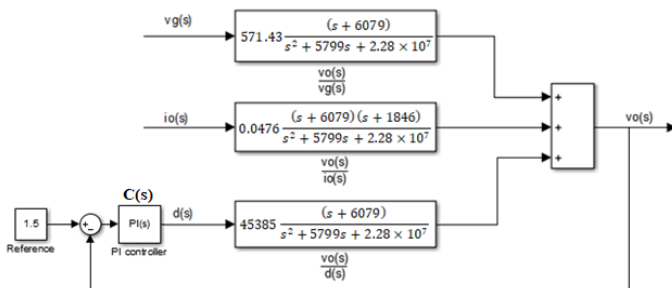


Fig. 10. Block diagram of the system with PI controller.

C. Suggested Method

The open loop dynamical model of the studied buck converter is shown in Fig. 7. Block diagram of closed loop voltage mode control of the converter is shown in Fig. 10. The controller is PI (Proportional Integral) since it is good enough for most of applications.

The PI controller must stabilize the loop. According to the Routh-Hurwitz stability test, following inequality must be satisfied in order to stabilize the loop.

$$(a.d + b.c)k_p + (a.c - b)k_i + c.d + a^2.k_p.k_i + a.b.k_p^2 > 0 \tag{15}$$

where, $a = 45385, b = 45385 \times 6079 = 275895415, c = 5799, d = 2.28 \times 10^7$. $k_p > 0$ and $k_i > 0$ are unknown proportional gain and integral gain, respectively.

According to the Fig. 10, the closed loop output impedance is:

$$Z_{CL} = \frac{\tilde{v}_o(s)}{\tilde{i}_o(s)} \times \frac{1}{1 + C(s) \cdot \frac{\tilde{v}_o(s)}{\tilde{d}(s)}} \tag{16}$$

or,

$$Z_{CL} = 0.0476 \frac{(s + 6079)(s + 1846)}{s^2 + 5799s + 2.28 \times 10^7} \times \frac{1}{1 + (k_p + \frac{k_i}{s}) \cdot 45385 \frac{(s + 6079)}{s^2 + 5799s + 2.28 \times 10^7}} \tag{17}$$

Equation (17) has two unknowns: k_p and k_i . We want to solve the following optimization problem:

$$\begin{aligned} \min \|Z_{CL}\|_{\infty} \\ \text{s.t. } C(s) \text{ stabilize } \frac{\tilde{v}_o(s)}{\tilde{d}(s)} \end{aligned} \tag{18}$$

Ideally Z_{CL} would be zero across the frequency bandwidth. However, since that would require an infinite loop gain, it is not practical for synthesis. The GA can be used to solve this optimization problem. GA changes the k_p and k_i values until the minimum is obtained. After running the GA, $k_p = 3.733$ and $k_i = 1.108 \times 10^4$ and $\min \|Z_{CL}\|_{\infty} = 0.476 \Omega$.

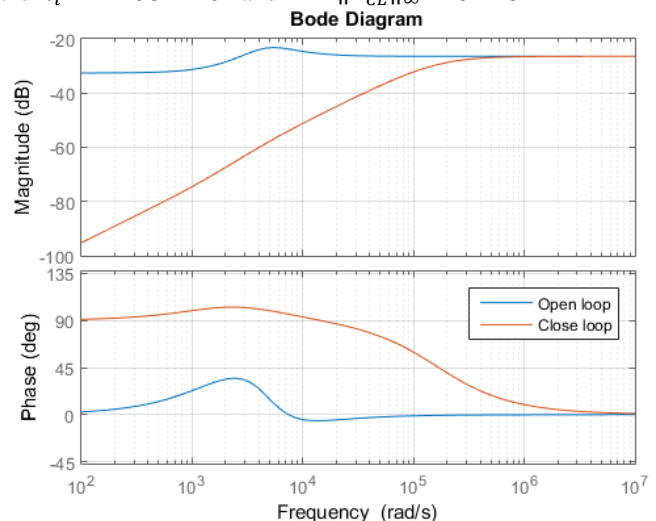


Fig. 11. Comparison between open loop output impedance and closed loop output impedance.

Fig. 11 compares the closed loop output impedance (Equation (17) with $k_p = 3.733$ and $k_i = 1.108 \times 10^4$) and open loop output impedance (Equation (7)). Closed loop output impedance decreased considerably in the low frequency range of the graph. Reduction of output impedance means a smaller voltage drop for larger currents. Such a property is highly desired in low voltage high current applications.

IV. SIMULATION RESULTS

Designed PI controller is tested in the Simulink environment. Simulation diagram is shown in Fig. 12. In the first test scenario, reference (Fig. 10) changes from 1.5 V to 3 V at $t= 25$ ms. Simulation result is shown in Figs. 13 and 14. Simulation result shows that the designed controller is able to follow the reference signal with zero steady state error. This is expected since the controller contains an integrator term.

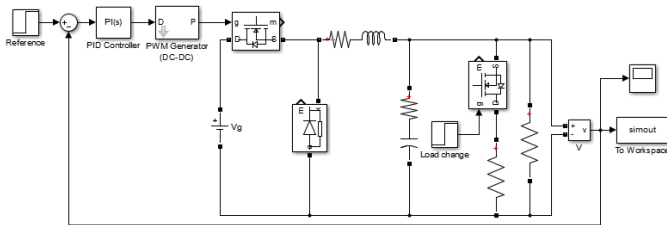


Fig. 12. Simulation diagram of the system.

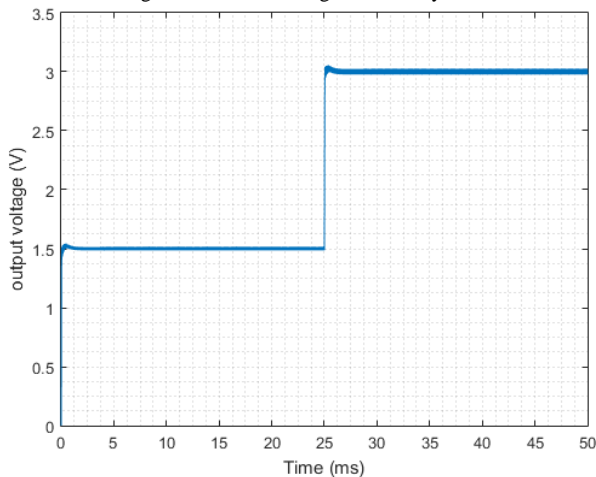


Fig. 13. Effect of change in reference.

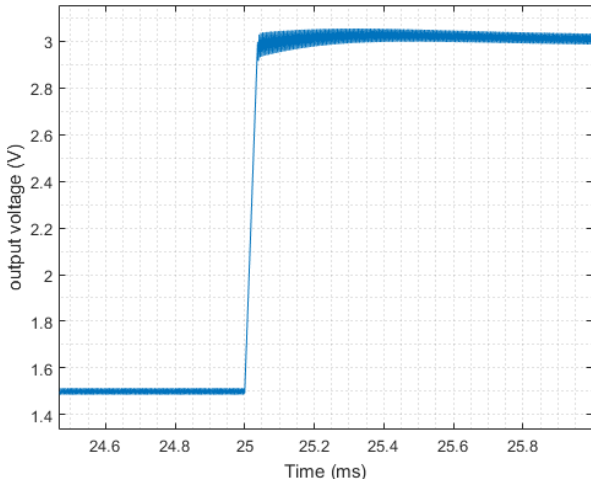


Fig. 14. Close-up Fig. 13 around $t= 25$ ms.

In the second scenario load current changes from 0.5 A to 5 A at $t=25$ ms. Simulation result is shown in Figs. 15 and 16. Simulation result shows that the controller is able to keep tracking the reference voltage even when the load changes. According to Fig. 16, maximum deviation for change of output current from 0.5 A to 5 A is $1.5-1.4=0.1$ V. In other words, change of 900% in output current generates change of 6.66% in output voltage. Such a small change in output voltage shows that the designed controller decreased the output impedance considerably.

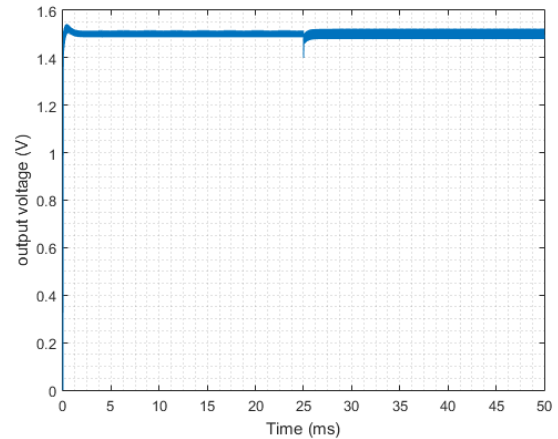


Fig. 15. Effect of change in output load.

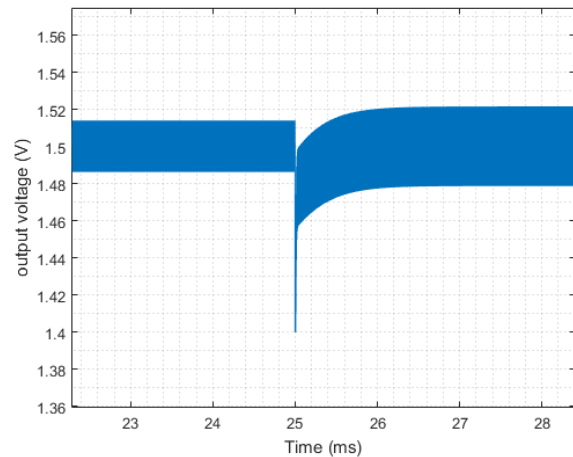


Fig. 16. Close-up Fig. 15 around $t= 25$ ms.

V. CONCLUSIONS

A technique to reduce the output impedance in the PWM buck converters with voltage-mode control is studied in this paper. Proposed technique used the H_∞ infinity norm to convert the impedance reduction problem into an optimization problem and use genetic algorithm to solve the obtained optimization problem. Suggested technique is cheaper in comparison to the classical solution of using the capacitors with lower ESR. Because in the suggested technique, there is no need to use any expensive low ESR capacitor and reduction of output impedance is done by good selection of controller parameters. Suggested technique is tested with the aid of Simulink simulations. According to the simulation results, the output impedance of converter decreased considerably in the low frequency region. Such a decrease is very desired for low

voltage high current loads like computer CPU's. Suggested method can be applied to other types of converters as well.

REFERENCES

- [1] E. Joard, J. Villarejo, F. Soto, and J. Muro, "Effect of the Output Impedance in Multiphase Active Clamp Buck Converters," *IEEE Trans. Ind. Electron.*, vol. 55, no. 9, pp. 3231-3238, Sep. 2008.
- [2] D. Goder and W. R. Pelletier, "V2 architecture provides ultra-fast transient response in switch mode power supplies," in *Proc. HFPC*, 1996, pp. 16-23.
- [3] J. Xu, X. Cao, and Q. Luo, "The effects of control techniques on the transient response of switching DC-DC converters," in *Proc. IEEE PEDS*, 1999, pp. 794-796
- [4] P. Wong, F. C. Lee, X. Zhou, and J. Chen, "VRM transient study and output filter design for future processors," in *Proc. IEEE IECON*, 1998, pp. 410-415.
- [5] D. Briggs, R. Martinez, R. Miftakhutdinov, and D. Skelton, "A fast, efficient synchronous buck controller for microprocessor power supplies," in *Proc. HFPC*, 1998, pp. 170-176.
- [6] B. Arbetter and D. Maksimovic, "DC-DC converter with fast transient response and high efficiency for low-voltage microprocessor loads," in *Proc. IEEE APEC*, 1998, pp. 156-162.
- [7] R. Miftakhutdinov, "Analysis of synchronous buck converter with hysteretic controller at high slew-rate load current transients," in *Proc. HFPC*, 1999, pp. 55-69.
- [8] C. J. Mehas, K. D. Coonley, and C. R. Sullivan, "Converter and inductor design for fast response microprocessor power delivery," in *Proc. IEEE PESC*, 2000, pp. 1621-1626.
- [9] L. D. Varga and N. A. Losic, "Synthesis of zero-impedance converter," *IEEE Trans. on Power Electronics*, vol. 7, no. 1, Jan 1992.
- [10] R. Redl and N. Sokal, "Near-optimum dynamic regulation of dc-dc converters using feed forward of output current and input voltage with current-mode control," *IEEE Trans. on Power Electronics*, vol. PE-1, no. 3, Jul 1986
- [11] J. Steenis, "Details on compensating voltage mode buck regulators," *Power Management Design Line*, September 4, 2006.
- [12] W.H. Lei, T.K. Man, "A general approach for optimizing dynamic response for buck converter," *ON Semiconductor*, Apr. 2004.
- [13] T. Hagerty, "Voltage-mode control and compensation: Intricacies for buck regulators," *Electronics Design, Strategy, News (edn.com)*, June 30, 2008.
- [14] L. Zhao, "Closed-loop compensation design of a synchronous switching charger using bq2472x/3x," *Texas Instruments application report*, September 2006.
- [15] "A handy method to obtain satisfactory response of buck converter," *Analog Integrations Corporation*, application note AN021, October 2001.
- [16] "Loop compensation of voltage-mode buck converters," *Sipex Corporation Technical Note*, October 11, 2006.
- [17] D. Mattingly, "Designing stable compensation networks for single phase voltage mode buck regulators," *Intersil Corporation Technical Note TB417.1*, December 2003.
- [18] H. Sucu, T. Goktas, M. Arkan, "Design, simulation and application of buck converter with digital PI converter", *Balkan Journal of Electrical and Computer Engineering*, April 2021.
- [19] F. Asadi, K. Eguchi, "On the Extraction of Input and Output Impedance of PWM DC-DC Converters", *Balkan Journal of Electrical and Computer Engineering*, April 2019.
- [20] F. Asadi, "On the Extraction of Input and Output Impedance of PWM DC-DC Converters", *Balkan Journal of Electrical and Computer Engineering*, April 2019.
- [21] H. Sucu, T. Goktas M. Arkan, "A GaN-Based Synchronous Buck Converter for High Power Laser Diode Drive Applications", *Balkan Journal of Electrical and Computer Engineering*, Nisan 2021.
- [22] Asadi F, Eguchi K. Dynamics and control of DC-DC converters, *San Rafael: Morgan and Claypool*; 2018. p. 89-145.
- [23] Suntio T. Dynamic profile of switched mode converter: modeling, analysis and control. *New Jersey: John Wiley & Sons*. 2009. p. 17-37.
- [24] Gu D, Petkov P, Konstantinov M. Robust Control Design with MATLAB. *Springer*. 2013. p. 9-11.



BIOGRAPHY

FARZIN ASADI received his B.Sc. in Electronics Engineering, M.Sc. in Control Engineering, and Ph.D. in Mechatronics Engineering. Currently, he is with the Department of Electrical and Electronics Engineering at Maltepe University, Istanbul, Turkey.

Dr. Asadi has published more than 40 international papers and 15 books. He is on the editorial board of 7 scientific journals as well. His research interests include switching converters, control theory, robust control of power electronics converters, and robotics.

Wavelength Tune Of InGaN Based Blue LEDs By Changing Indium Percentage And Operational Voltage Variables

Bekir Gecer and Ismail Kiyak

Abstract—Blue light-emitting-diodes (LEDs) are special and different than the other LEDs due to their high-efficient lighting. They have large bandgap energy. So gallium nitrides are mostly used during designing blue light. This application focus on the emission properties of a InGaN LED. The emission intensity, energy diagram, spectrum, and efficiency are calculated for an applied voltage. The indium and GaN composition in the blue LED can be varied to control the emission wavelength. In here, composition of InGaN and operational voltage values were varied to control wavelength. Effects of the different InGaN composition and different voltage values were given in results as comprehensively for 5 different simulations. This study is dissimilar than other wavelength studies due to used original parameter values and wavelength compared methods for many situations about InGaN percentage and voltage values. As a result of simulations, we can infer that the high In_x percentage in composition and high voltage makes wavelength gap of an blue LED larger.

Index Terms—Comsol, InGaN, LED, Voltage, Wavelengths.


I. INTRODUCTION

BLUE LEDs are so popular and different than other LEDs due to their use in modern high-efficiency lighting. Because of the large bandgap energy of blue LEDs, gallium nitrides are mostly used for designing blue light [1]. In the first years of LED technology, blue devices were lacking. The firstly, red LED was generated in the 1950s and by the 1960s the pursuit of shorter emission wavelengths had already yielded green LEDs. The GaN-based blue LED researches were started by Akasaki in 1973. In there, The MBE and HVPE methods were used for growth [2]. The solid-state lighting performs such as reality thanks to improvements in high


power light emitting diode technology. It is 100 + lumens per LED chip. The sensor position, driver of LED, design, control and stability with temperature are practical issues for implementation of combining red, green and blue (RGB) LEDs. [3]. In another paper, the structure optimization of the multi-quantum well based Light Emitting Diode (LED) was presented. In there, the electrical and optical properties of the device on several factors such as well width, barrier width, the number of quantum wells were investigated by authors [4]. Room-temperature photoluminescence (PL) measurements are performed on the GaInN/GaN multiple-quantum-well heterostructures grown on GaN-on-sapphire templates with different threading-dislocation densities in [5]. A photoluminescence technique measures the Auger recombination coefficient in quasi-bulk In_xGa_{1-x}N/In_xGa_{1-x}N (x~9%–15%)(x~9%–5%) layers grown on GaN (0001). The samples vary in InN composition, thickness, and threading dislocation density in [6]. In [7], a blue-red LED wavelength-shifting system (B-R system) was studied to improve the photosynthetic betacarotene productivity of *Dunaliella salina*. The characteristics of the GaN-based blue LED on Si substrate is studied in [8]. In [9], effect of the different wavelength Blue LED on human optical biorhythm was given with used methods. The performance of 2QWs LED is studied by comparing device behavior according to available one quantum well (1QW) light emitting diode model in Comsol simulation program in [10]. In another comsol work, to design high bandwidth GaN-based blue LEDs, Ag-grating and diamond heat sink were used [11]. In reference [12], the performance improvement of GaN-based LEDs using different methods. In another study, InGaN based blue LEDs are worked to enhance their efficiency in bottom tunnel junction [13]. The optical model of red, green and yellow phosphors with a Blue LED is constructed using the optical simulation software LightTools in [14].

In this study, 5 different InGaN-based blue LED designed and simulated using Comsol programs. Effects of the different input voltages of different In_x percentages on the wavelength were compared after simulations. There is no another study that compare wavelength of blue LED at different voltage and In_x percentages for 5 simulations in the literature. Here, many applications were studied at varied In_x percentages and voltage parameters. Also, optimizations were worked for maximum LED performance about wavelength. So, this paper is novel and original. Section II is about LED

BEKIR GECER, is with Machine and Metal Technology Department of Atasehir Adiguzel Myo, Istanbul, Turkey, (e-mail: bekirgecer@adiguzel.edu.tr).

 <https://orcid.org/0000-0002-7803-3844>

ISMAL KIYAK, is with Department of Electrical Engineering University of Marmara University, Istanbul, Turkey, (e-mail: imkiyak@marmara.edu.tr).

 <https://orcid.org/0000-0002-5061-6378>

Manuscript received July 1, 2021; accepted July 19, 2022.

DOI: [10.17694/bajece.960918](https://doi.org/10.17694/bajece.960918)

design and simulation parameters, III is about results and conclusion is in stage IV.

II. LED DESIGN AND SIMULATION PARAMETERS

The Comsol software program offers many facilities in simulation-LED design. Here, you can control and test designs in the virtual environment and determine the optimal configuration. The Fig.1 shows basic geometry of AlGaIn based LED device with layers. In this stage, we focus on the InGaIn based blue LED, its design and simulation parameters were given for the simulation programs in tables. The following input parameters were used to specify the composition of the optically active InGaIn material and device lateral dimensions. There are 5 different LEDs design and simulation parameters for comparison of wavelength.

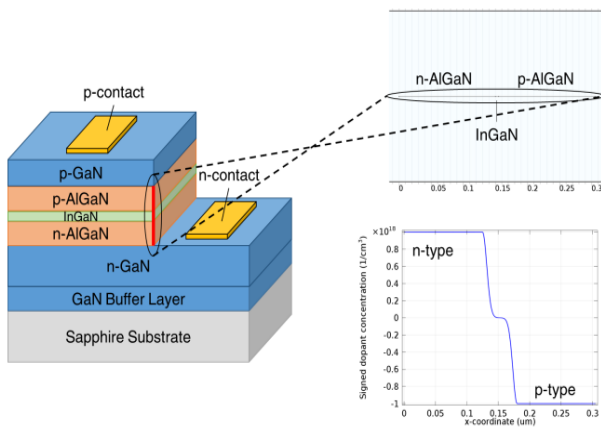


Fig.1. Geometry of LED device.

First simulation parameters were given in Table I. Here, In_x percentage is %12. The operational voltage value is 3.2 V and efficiency is 18.03 for simulation 1.

TABLE I
REFERENCE PARAMETERS 1

Name	Expression	Value	Description
In _x	0.127	0.127	Indium fraction
InGaIn _{bg}	$In_x * InN_{bg} + (1 - In_x) * GaN_{bg} - b * In_x * (1 - In_x)$	2.753 V	Bandgap energy of InGaIn layer
A _{cross}	200[um]*200[um]	4E-8 m ²	Cross sectional area

Second simulation parameters were given in Table II. Here, In_x percentage is %9. The operational voltage value is 3.2 V and efficiency is 18.78 for simulation 2.

TABLE II
REFERENCE PARAMETERS 2
(LOWER IN_x PERCENTAGE THAN REFERENCE)

Name	Expression	Value	Description
In _x	0.09	0.09	Indium fraction
InGaIn _{bg}	$In_x * InN_{bg} + (1 - In_x) * GaN_{bg} - b * In_x * (1 - In_x)$	2.9322 V	Bandgap energy of InGaIn layer
A _{cross}	200[um]*200[um]	4E-8 m ²	Cross sectional area

TABLE III
REFERENCE PARAMETERS 3
(HIGHER IN_x PERCENTAGE THAN REFERENCE)

Name	Expression	Value	Description
In _x	0.134	0.134	Indium fraction
InGaIn _{bg}	$In_x * InN_{bg} + (1 - In_x) * GaN_{bg} - b * In_x * (1 - In_x)$	2.72 V	Bandgap energy of InGaIn layer
A _{cross}	200[um]*200[um]	4E-8 m ²	Cross sectional area

Third simulation parameters were given in Table III. Here, In_x percentage is 13.4 %. The operational voltage value is 3.2 V and efficiency is 18.04 for simulation 3.

TABLE IV
REFERENCE PARAMETERS 4 (SAME VALUES AS REFERENCE)

Name	Expression	Value	Description
In _x	0.127	0.127	Indium fraction
InGaIn _{bg}	$In_x * InN_{bg} + (1 - In_x) * GaN_{bg} - b * In_x * (1 - In_x)$	2.753 V	Bandgap energy of InGaIn layer
A _{cross}	200[um]*200[um]	4E-8 m ²	Cross sectional area

Fourth simulation parameters were given in Table IV. Here, In_x percentage is 12 %. The operational voltage value is 3.1 V and efficiency is 25.46 for simulation 4.

TABLE IV
REFERENCE PARAMETERS 4 (SAME VALUES AS REFERENCE)

Name	Expression	Value	Description
In _x	0.127	0.127	Indium fraction
InGaIn _{bg}	$In_x * InN_{bg} + (1 - In_x) * GaN_{bg} - b * In_x * (1 - In_x)$	2.753 V	Bandgap energy of InGaIn layer
A _{cross}	200[um]*200[um]	4E-8 m ²	Cross sectional area

Fifth simulation parameters were given in Table V. Here, In_x percentage is 12 %. The operational voltage value is 3.3 V and efficiency is 9.073 for simulation 5.

TABLE V
REFERENCE PARAMETERS 5 (SAME VALUES AS REFERENCE)

Name	Expression	Value	Description
In _x	0.127	0.127	Indium fraction
InGaIn _{bg}	$In_x * InN_{bg} + (1 - In_x) * GaN_{bg} - b * In_x * (1 - In_x)$	2.753 V	Bandgap energy of InGaIn layer
A _{cross}	200[um]*200[um]	4E-8 m ²	Cross sectional area

According to tables, the efficiency was found depend on voltage and percentages of materials.

III. SIMULATION RESULTS

In this section, results of 5 different LED design simulations which input parameters were given in section II, were shown in figures as below. The electroluminescence spectrum is shown below for simulation 1. Effects of wavelength gap and

arc length on the relative intensity were seen in Fig 2 and Fig 3.

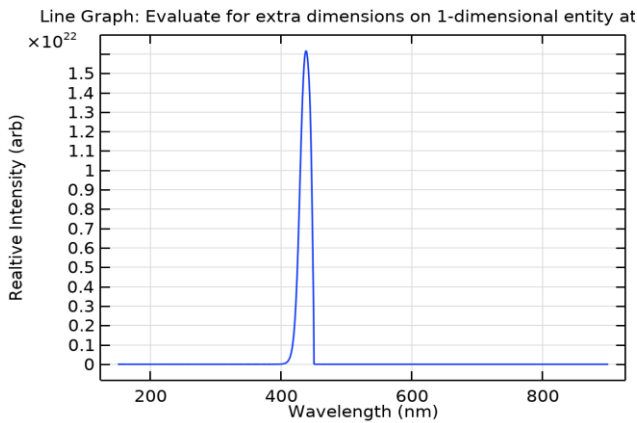


Fig.2. Line Graph: Evaluate for extra dimensions on 1-dimensional entity at coordinates in Geometry 1 (kg/(m2*s3))(371-449nm)

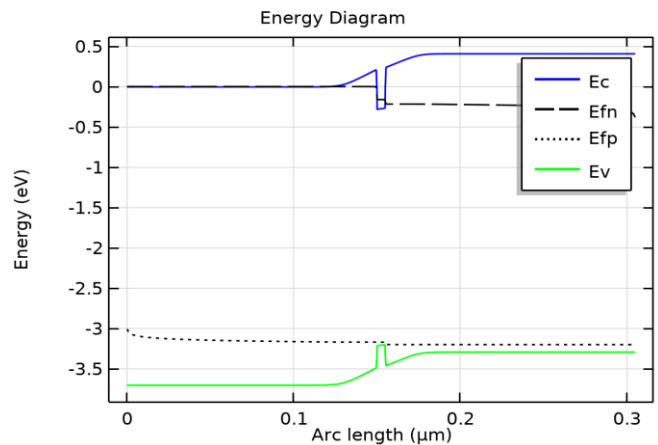


Fig.5. Energy Diagram

The electroluminescence spectrum is shown below for simulation 3. Effects of wavelength gap and arc length on the relative intensity were seen in Fig 6 and Fig 7.

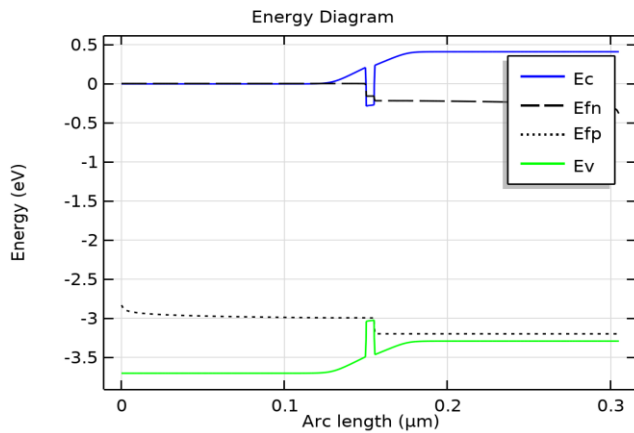


Fig.3. Energy Diagram

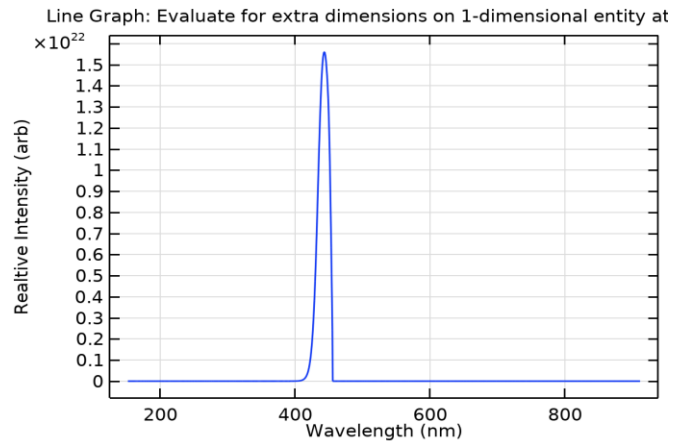


Fig.6. Line Graph: Evaluate for extra dimensions on 1-dimensional entity at coordinates in Geometry 1 (kg/(m2*s3))(374-456nm)

The electroluminescence spectrum is shown below for simulation 2. Effects of wavelength gap and arc length on the relative intensity were seen in Fig 4 and Fig 5.

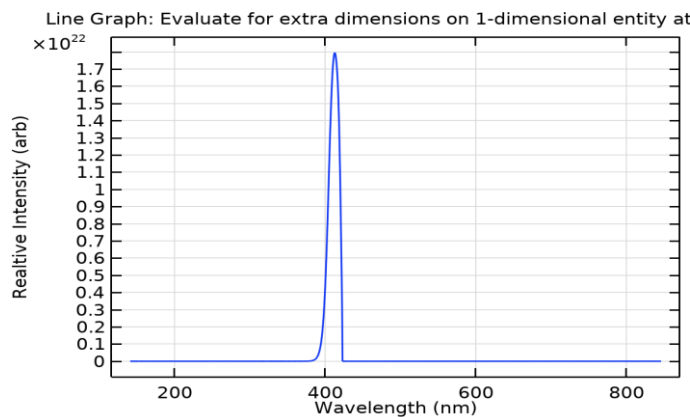


Fig.4. Line Graph: Evaluate for extra dimensions on 1-dimensional entity at coordinates in Geometry 1 (kg/(m2*s3))(380-423nm)

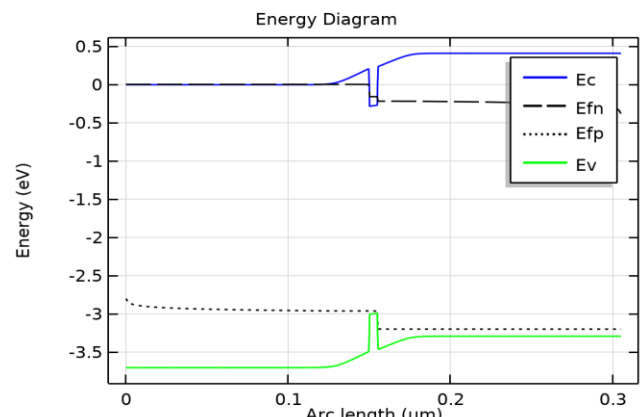


Fig.7. Energy Diagram

The electroluminescence spectrum is shown below for simulation 4. Effects of wavelength gap and arc length on the relative intensity were seen in Fig 8 and Fig 9.

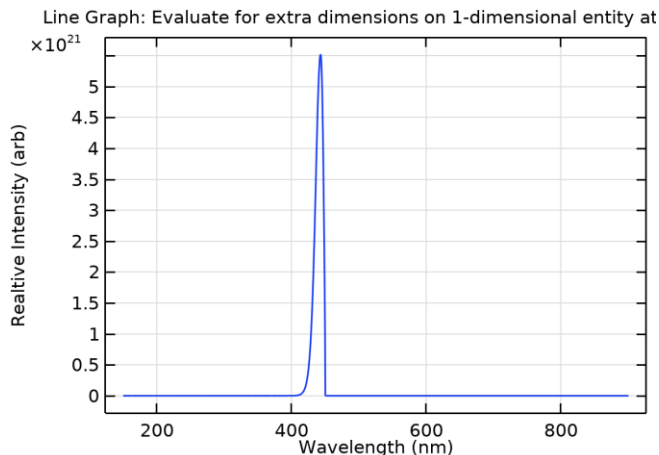


Fig.8. Line Graph: Evaluate for extra dimensions on 1-dimensional entity at coordinates in Geometry 1 (kg/(m2*s3))(374-451nm)

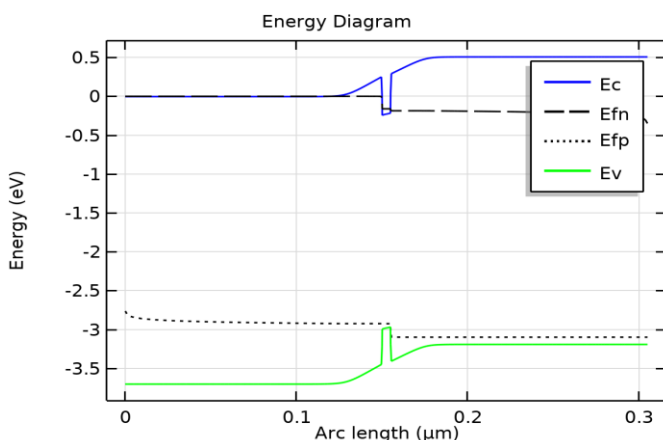


Fig.9. Energy Diagram

The electroluminescence spectrum is shown below for simulation 5. Effects of wavelength gap and arc length on the realtive intensity were seen in Fig 10 and Fig 11.

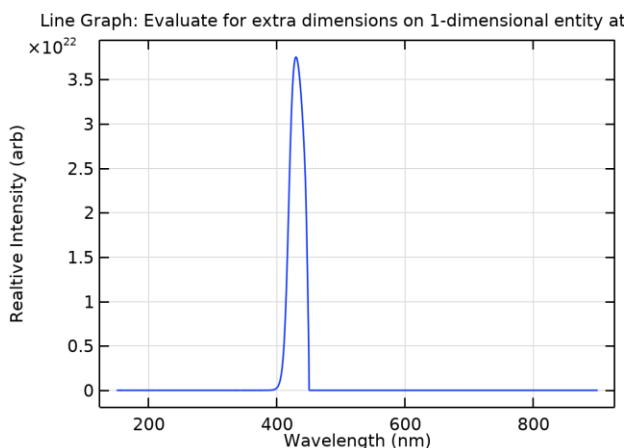


Fig.10. Line Graph: Evaluate for extra dimensions on 1-dimensional entity at coordinates in Geometry 1 (kg/(m2*s3))(367-452nm)

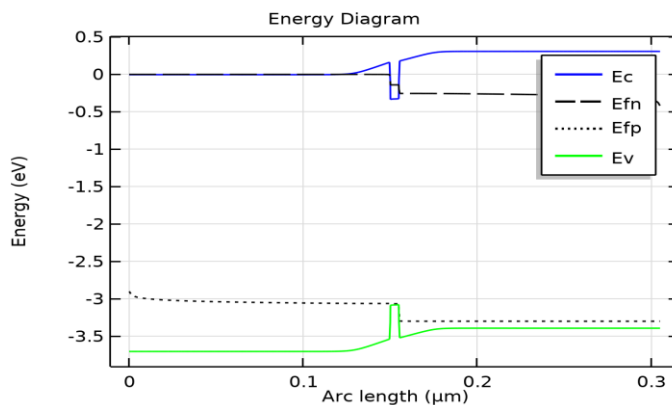


Fig.11. Energy Diagram

IV. CONCLUSION

The electroluminescence spectrum is calculated and plotted as a line graph for the case of a single input voltage, and as a height plot as a function of voltage and wavelength for the case of a voltage range input. The current that flows through the device is calculated for each applied voltage. High voltage input values show larger wavelength gap behaviour than low voltage input values. Also, effects of the different percentegas of InN and GaN on the wavelength were seen in the results. High In_x percentaged LED gives larger wavelength gap behavior than low In_x percentaged LED in the results.

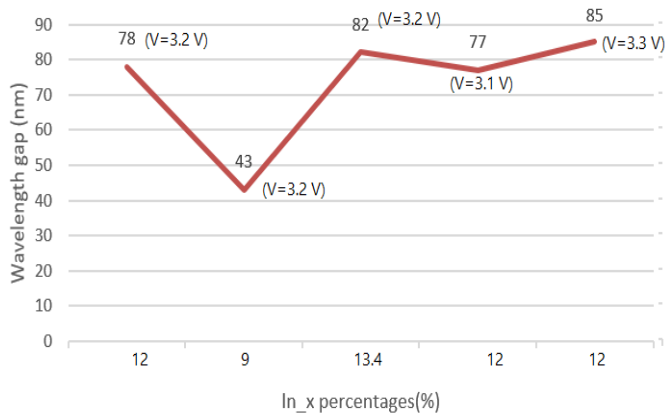


Fig.12. Relationship between In_x percentage and Wavelength gap.

Wavelength gap-In_x graph is seen in Fig.12 and simulation results are given in Table VI.

TABLE VI. SIMULATION RESULTS	
Simulation No	Wavelength
Simulation 1	371-449 nm
Simulation 2	380-423 nm
Simulation 3	374-456 nm
Simulation 4	374-451 nm
Simulation 5	367-452 nm

According to Table VI, the maximum wavelength gap with 85 nm was obtained at V=3.3 V voltage value and %12 In composition percentage. The minimum wavelength gap is 43 nm at V=3.2 V voltage and %9 percentage.

REFERENCES

- [1] S. Nakamura, T. Mukia, and M. Senoh, "Candela-class high-brightness InGaN/AlGaIn double-heterostructure blue-light-emitting diodes," *Appl. Phys. Lett.*, vol. 64, p. 1687, 1994.
- [2] S.Muthu, J.Gaines, "Red, green and blue LED-based white light source:implementation challenges and control design", 38th Conference Record of the IEEE Industry Applications Society Annual Meeting (IAS), 2003.
- [3] Y. Nanishi, "The birth of the blue LED", *Nature Photonics*, 2014.
- [4] K.Meel, P. Mahala, "Design and Fabrication of Multi Quantum well based GaN/InGaIn Blue LED", *IOP Conference Series Materials Science and Engineering*, 2018.
- [5] Q. Dai et. al., "Internal quantum efficiency and nonradiative recombination coefficient of GaInN/GaN multiple quantum wells with different dislocation densities," *Appl. Phys. Lett.*, vol. 94, 111109, 2009.
- [6] Y.C. Shen, G.O. Mueller, S. Watanabe, N.F. Gardner, A. Munkholm, and M.R. Krames, "Auger recombination in InGaIn measured by photoluminescence," *Appl. Phys. Lett.*, vol. 91, 141101, 2007.
- [7] S. Han, S. Kim, C. Lee, "Blue-Red LED wavelength shifting strategy for enhancing beta-carotene production from halotolerant microalga, *Dunaliella Salina*", *Microbial Systematics and Evolutionary Microbiology*, 2018.
- [8] C. Xiong, F. Jiang, W. Fang, "The characteristics of GaN-based LED on Si substrate", *Journal of Luminescence*, pages 185-187, 2007.
- [9] L. Yu-hong, W. Yo-rong, J. Shang, "Influence of Different Wavelength Blue LED on Human Optical Biorhythm Effect", 2013 .
- [10] P. Kumar, M. Saheed, Z. Burhanudin, "Performance comparison of one&two quantum wells light emitting diodes simulated with COMSOL Multiphysics", 6th International Conference on Intelligent and Advanced Systems(ICIAS), 2016.
- [11] R. Xie, Z. Li, S. Guo, "High Bandwidth GaN-based blue LEDs using Ag-grating and diamond heat sink", *Photonics and Nanostructures-Fundamentals and Applications*, 2020.
- [12] C. Jia, T. Yu, C. Zhong, "Performance improvement of GaN-based LEDs with step stage InGaIn/GaN strain relief layers in GaN-based blue LEDs", *Optics Express*, volume 21, 2013.
- [13] L. Deurzen, S. Bharadwaj, "Enhanced efficiency in bottom tunnel junction InGaIn blue LEDs", *Light-Emitting Devices, Materials and Applications XXV. Conference*, 2021.
- [14] W. Sun, C. Tien, "Optical Modeling Analysis of Red, Green, and Yellow Phosphors with a Blue LED", *Advances in Condensed Matter Physics*, volume 2018.

research areas are lighting technique, semi conductive light sources and renewable energy resource.

BIOGRAPHIES



B. GECER Maltepe, Istanbul, in 2022. He received the B.S. degree in electrical-electronic engineering from the Fatih University, Istanbul, in 2014 and M.S. degree in electrical-electronic engineering from the Marmara University, Istanbul, in 2019. He is a PHD

student at Marmara University. He works as a Lecturer at Atasehir Adiguzel MYO for the Machine and Metal Technology Department, Istanbul. His research areas are electric motors, power electronic and control systems.



I. KIYAK Maltepe, Istanbul, in 2022. He received the B.Sc, M.Sc. and Ph.D. degrees in the Department of Electrical Education from the Marmara University (MU) in 2000, 2005 and 2010 respectively. He is working as a Professor at the Department of Electrical and

Electronics Engineering, Marmara University, Turkey. His

Efficient Task Scheduling in Cloud Systems with Adaptive Discrete Chimp Algorithm

Emrullah Gunduzalp, Gungor Yildirim and Yetkin Tatar


Abstract— Successful task scheduling is one of the priority actions to increase energy efficiency, commercial earnings, and customer satisfaction in cloud computing. On the other hand, since task scheduling processes are NP-hard problems, it is difficult to talk about an absolute solution, especially in scenarios with large task numbers. For this reason, metaheuristic algorithms are frequently used in solving these problems. This study focuses on the metaheuristic-based solution of optimization of makespan, which is one of the important scheduling problems of cloud computing. The adapted Chimp Optimization Algorithm, with enhanced exploration and exploitation phases, is proposed for the first time to solve these problems. The success of the proposed method has been tested for different simulation scenarios. According to the simulation results, the proposed method achieved a makespan improvement of approximately 30% compared to the standard task scheduling algorithms.

Index Terms— Chimp algorithm, Cloud computing, Makespan, Metaheuristic, Optimization.


I. INTRODUCTION

CLOUD COMPUTING is technology that allows the dynamic delivery of flexible, scalable, and distributed computing resources to end-users [1]. Many reasons such as the development of the Internet infrastructure, the use of the Internet of Things (IoT) technology, the increase in the need for big data technologies, and the developments in artificial intelligence have led to the widespread use of cloud computing. Cloud computing enables users to access various services and resources (CPU, RAM, storage) wherever there is internet access. Thanks to the pay-as-you-go system, both software and hardware resources become more economical, while installation and maintenance costs are significantly reduced for customer companies [2]. Cloud service providers must provide resources and services to their customers without violating the


EMRULLAH GUNDUZALP, is with Department of Computer Engineering University of Firat University, Elazig, Turkey, (e-mail: emrullahg@dsi.gov.tr).

 <https://orcid.org/0000-0001-6418-5663>

GUNGOR YILDIRIM, is with Department of Computer Engineering University of Firat University, Elazig, Turkey, (e-mail: gungor.yildirim@firat.edu.tr).

 <https://orcid.org/0000-0002-4096-4838>

YETKIN TATAR is with Department of Computer Engineering University of Firat University, Elazig, Turkey, (e-mail: ytatar@firat.edu.tr).

 <https://orcid.org/0000-0002-7181-0014>

Manuscript received September 1, 2021; accepted July 29, 2022.

DOI: [10.17694/bajece.989467](https://doi.org/10.17694/bajece.989467)

Service Level Agreement (SLA) and guaranteeing Quality of Service (QoS). Therefore, cloud performance is very important for both service providers and users. One of the main issues affecting the performance of cloud systems is task scheduling. Task scheduling is one of the important problems in cloud computing. Especially, inefficient task scheduling could cause loss of performance and revenue, and SLA violation. Efficient scheduling algorithms can optimize important measurements such as makespan, traffic volume, computational time, communication cost, system efficiency and utilization [3].

Virtual machines (VMs), one of the basic mechanisms of cloud technologies, contribute to the efficient use of infrastructure resources. The task scheduling process in this study can be briefly described as follows: Distributing tasks of different sizes reaching the cloud system to the most suitable VMs in a way that provides the shortest response time. Inappropriate scheduling can reveal underloaded (underutilized) or overloaded (overutilized) of resources, referred to as resource dilemma. These situations ultimately lead to wasted cloud resources or reduced service performance [4]. In solving NP-hard problems, using metaheuristic methods instead of standard deterministic algorithms may yield more successful results [1]. In this study, the task scheduling process in cloud systems is studied with an adapted version of Chimp Optimization Algorithm (ChOA), which is a metaheuristic algorithm in use. The makespan problem was taken as a basis in the scheduling process. To the best of the authors' knowledge, the proposed adaptive version of the ChOA and its application on this type of cloud system problem has never been studied before in the literature. Simulations were carried out by integrating ChOA into the CloudSim 3.0.3 simulator. First of all, the Adaptive Chimpanzee Optimization Algorithm (AChOA) has been proposed, which makes the exploration and exploitation stages of ChOA more adaptive and uses different mathematical functions for this purpose. In AChOA, exploration and exploitation mechanisms, apart from the standard method, Sigmoid Decreasing Weight (SDW), Oscillating Weight (OscW) and V-shaped Family (VSW) functions have been included in the calculation process. Thus, by using different functions, the efficiency and performance of the optimization process are comparatively observable. Finally, the space shared task scheduling approach of both AChOA and CloudSim, which use different functions, are tested for different experimental scenarios and the results are discussed. The original aspects of this study can be briefly summarized as follows:

- The study uses the metaheuristic Chimp algorithm for the first time for task scheduling problems in cloud systems.
- Unlike other swarm-based metaheuristic algorithms, the proposed method adaptively manages exploration and exploitation processes.
- Compared to the standard scheduling method, the proposed method achieves an improvement of approximately 30% in makespan.
- The adaptive method used can be easily adapted to other optimization problems that uses the Chimp algorithm.

The adaptive method used can be easily adapted to other problems using the Chimp algorithm. The continuation of this article is organized as follows. In section II, the task scheduling problem in cloud systems is introduced and literature studies on this subject are explained. In Section III, the methodology of the method applied in this study is presented. Experiments and analyzes are shared in section IV, and conclusions are given in section V.

II. TASK SCHEDULING AND LITERATURE SUMMARY IN CLOUD COMPUTING

Task scheduling is a process that affects the performance and efficiency of cloud systems. In short, task scheduling can be expressed as the optimal assignment of n tasks $\{T_1, T_2, \dots, T_n\}$ to m machines $\{M_1, M_2, \dots, M_m\}$, taking into account one or more predefined optimization targets [4]. With virtualization techniques, which have provided significant advantages in recent years, physical servers used in cloud systems can be divided into more than one virtual machine (VM), and each virtual machine can be used to allocate different tasks. Cloud service providers (CSPs) may have multiple observer and control infrastructure services. One of the most important of these is the broker services that optimally distribute the incoming tasks to the resources in the system according to their types and characteristics. While doing this, they use algorithms that take into account both the task and the resource characteristics offered.

Task scheduling in cloud computing generally includes three main operations [6]. These are Strategy Phase, Planning Phase, and Deployment Phase. In Strategy Phase, all shared resources in the data center and their properties are made discoverable and questionable. In Planning Phase, a suitable resource is determined according to the task requirements, while in Deployment Phase, the selected resources are allocated to the relevant tasks. It is a complex process to carry out the planning process in a heterogeneous and dynamic environment such as the cloud environment. It is quite difficult to find the optimum planning method in this process, which is carried out with different optimization objectives such as cost, energy consumption, makespan, and execution variability. Target strategies are generally grouped under four headings [7]. The scheduling strategies are dynamism, target architecture and scheduling algorithms.

Task scheduling can be single-objective or multi-objective. While optimizing the scheduling process, one of the goals, such as makespan, computational cost, or several goals contradicting with each other can be taken as a basis. At the same time, the

scheduling process can be done statically or dynamically. Static algorithms can produce successful and fast results in small-scale cloud systems where prior knowledge of incoming tasks and available resources is not required, the workload does not change frequently. However, cloud systems are by nature dynamic systems with a lot of variation in workload. Therefore, dynamic algorithms give more successful results. [4,7]. The target system architecture is another issue that affects the scheduling strategy to be implemented. Suitable solution algorithms are determined during the planning process. The algorithms used here are classified as heuristics, metaheuristics, and hybrid algorithms. Heuristic algorithms are often preferred for static scheduling. These algorithms are fast, but they are insufficient for cloud systems with a large-scale dynamic environment. Metaheuristic strategies are effective methods for solving NP-hard optimization problems with high efficiency. These algorithms can be expressed as Heuristic + Randomization. Another approach to solving the task scheduling problem is the use of hybrid methods in which two or more algorithms are combined. [4]. In Fig.1, examples of algorithms used in the literature according to this classification are given.

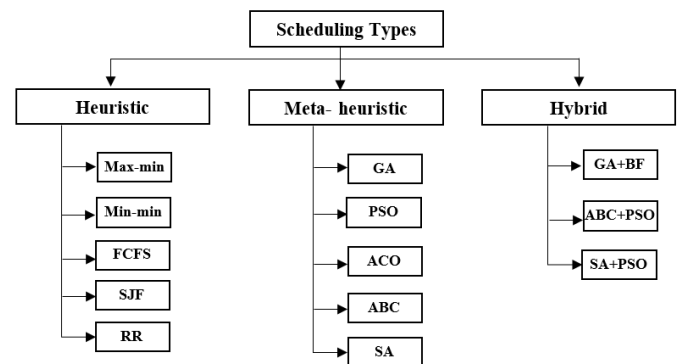


Fig.1. Various scheduling algorithms [6,7]

Many problems encountered in cloud computing attract the attention of researchers. One of these problems is the task scheduling process. In [8], the authors use the minimum completion time (MCT) and longest job to fastest processor (LJFP) to initialize the PSO. The goal is to minimize the makespan, total energy consumption and execution time, and degree of imbalance. In [9], the authors used the gray wolf optimization (GWO) technique to solve the task scheduling problem. The authors aimed to minimize the makespan. In [10], the authors proposed a whale optimization algorithm (WOA)-based method that aims to improve the performance of task scheduling with a multi-objective optimization model. They claimed that they improved their ability to search for optimum solutions with the approach they named IWC. In another study [11], the authors proposed Electromagnetism Metaheuristic Algorithm (EMA) in order to improve QoS in the cloud. They achieved this by scheduling tasks on virtual machines (VMs) to optimize completion time. In [12], the authors analyzed it with a time-shared and space-shared genetic algorithm. The study states that the method used outperforms competitive scheduling methods in terms of completion time and cost. In [13], the authors introduced an ACO-based load balancing algorithm for

task scheduling. A hybrid approach using PSO and ACO algorithms for task scheduling was proposed in [14]. Another hybrid approach is introduced in [15]. In this approach, the genetic algorithm uses its global search capability to minimize the task execution time and then transforms the obtained results into the initial pheromone of ACO to achieve more successful optimization. In [16], the authors proposed a hyper-heuristic scheduling algorithm using a framework including GA, PSO, and ACO in order to optimize makespan.

III. PROBLEM DEFINITION AND METHODOLOGY

Tasks submitted by users are put into task queues before they are assigned to the respective virtual machines. Tasks waiting in the queue are sent to the task planner by the VMM (Virtual Machine Manager). The task planner determines the resources that will execute the tasks and makes the assignments to the relevant VMs in a way as to use the resources efficiently [9]. In this study, this process was performed with AChOA. Task assignment is usually performed using the tabulation technique. In this technique, first of all, it is determined which task will be executed by which virtual machines. For example, mapping 10 tasks, belonging to a specific application and not dependent on each other's output, to $m=5$ virtual machines can be shown in Fig.2.

VM ₅	VM ₁	VM ₂	VM ₃	VM ₁	VM ₅	VM ₅	VM ₁	VM ₂	VM ₂
T ₁	T ₂	T ₃	T ₄	T ₅	T ₆	T ₇	T ₈	T ₉	T ₁₀

Fig.2. An example task scheduling representation

The task group, $T=\{T_1, T_3, \dots, T_n\}$, is a set of independent tasks containing million instructions (MI). Each task in T is limited to the specified number of commands. The V cluster contains VMs, and each VM has a metric that shows how many million instructions per second (MIPS) it can process. The MIPS value of the VMs is also within the specified limits. Assume that the sizes of the tasks are as in Fig.3 after they have been assigned to the VMs as in the example in Fig.2. In this case, the makespan value for this task set will be the total time that VM₅ will spend executing T₁, T₆, and T₇.

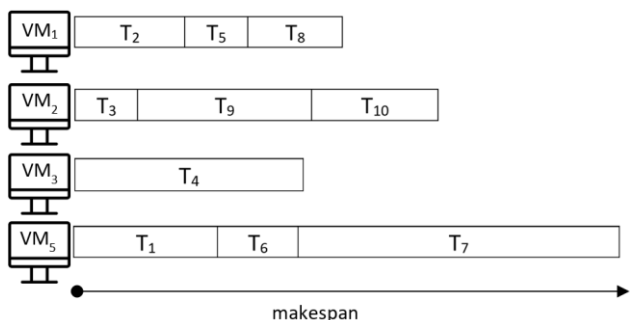


Fig.3. Tasks assigned to VMs according to an example scenario

In the representation of these types of assignment operations, it is usually used a matrix model that shows which task is assigned to which VM. This matrix is as follows.

$$X = \begin{bmatrix} x_{11} & \dots & x_{1m} \\ \vdots & \ddots & \vdots \\ x_{n1} & \dots & x_{nm} \end{bmatrix}, \text{ here} \quad (1)$$

$$\begin{cases} x_{ij} = 1, & \text{if } T_i \text{ is assigned to } VM_j \\ x_{ij} = 0, & \text{if } T_i \text{ is not assigned to } VM_j \end{cases}$$

As a result of the above-mentioned definitions, the estimated execution time for the i th task in j th VM is calculated by Eq.2. Since the tasks are independent, each task can only be assigned to one VM, and this is expressed as $\sum_{j=0}^m x_{ij} = 1, (1 \leq i \leq n)$. Thus, the execution time (VET_j) of j th VM is calculated as in Eq.3.

$$ECT_{ij} = \frac{T_i}{VM_j} \quad (2)$$

$$VET_j = \sum_{i=0}^n x_{ij} * ECT_{ij}, \quad 1 \leq j \leq m \quad (3)$$

The makespan (MS) in a task group is equal to the maximum execution time, expressed by Eq 4. The objective function used is to minimize the MS for all candidate solutions.

$$MS = \text{Max} \{VET_j\}, \quad 1 \leq j \leq m \quad (4)$$

A. Chimp Optimization Algorithm

Metaheuristic algorithms are among the techniques that are frequently used in solving optimization problems in the literature. Population-based algorithms using swarm intelligence constitute an important part of metaheuristic algorithms. In this section, one of the current algorithms, Chimp Optimization Algorithm-ChOA will be explained and the application of this algorithm to task scheduling in cloud systems will be shown. In the study in [5], the authors developed the ChOA inspired by the hunting strategies of chimps. According to this algorithm, there are four different chimps in a colony. Those are the attacker, chaser, barrier, driver. They all have different abilities and they use these abilities while hunting. In this algorithm, the attacker is the candidate that holds the best result. The hunting mechanism in ChOA consists of two phases. The first is the exploration phase, in which driving and chasing the prey are performed. This is expressed by Eq. 5-6.

$$d = |\vec{c} \circ \vec{X}_{prey}(t) - \vec{m} \circ \vec{X}_{chimp}(t)| \quad (5)$$

$$\vec{X}_{chimp}(t+1) = \vec{X}_{prey}(t) - \vec{a} \cdot d \quad (6)$$

In this equation, t represents the current iteration, while c , m , and a are the coefficient vectors calculated by Eq. 7-9. \vec{X}_{prey} and \vec{X}_{chimp} are the position vectors of prey and predator, respectively. \circ represents the Hadamard product.

$$\vec{a} = 2 \cdot f \cdot \vec{r}_1 - f \quad (7)$$

$$\vec{c} = 2 \cdot \vec{r}_2 \quad (8)$$

$$\vec{m}_{(t+1)} = \begin{cases} 1, & \text{if } \vec{m}_t = 0 \\ \frac{1}{\text{mod}(\vec{m}_t, 1)}, & \text{if } \vec{m}_t \neq 0 \end{cases} \quad (9)$$

Here, the f value is reduced from 2.5 to 0 depending on the current iteration value. r_1 and r_2 are random uniform vectors. The m value is a chaotic vector representing the effect of chimps' intuitive motivation in the hunting process. In this study, this chaotic value was calculated by Gauss/Mouse map [23] method as in Eq. 9. The elements of the vector \vec{c} change randomly in the interval [0,2]. It also improves the ChOA's stochastic behaviour and reduce the likelihood of being caught at the local minimum.

The second stage is the exploitation (attack stage). In this stage, the attacker, chaser, barrier, and driver have information about the prey location. Thus, the four best solutions obtained so far are retained, and the other chimpanzees update their positions to these four best chimp positions, as in Eqs. 10-12.

$$\begin{aligned}
 d_{attacker} &= |\vec{c}_1 \circ \vec{X}_{attacker} - \vec{m}_1 \circ \vec{X}_{candidate}|, \\
 d_{barrier} &= |\vec{c}_2 \circ \vec{X}_{barrier} - \vec{m}_2 \circ \vec{X}_{candidate}|, \\
 d_{chaser} &= |\vec{c}_3 \circ \vec{X}_{chaser} - \vec{m}_3 \circ \vec{X}_{candidate}|, \\
 d_{driver} &= |\vec{c}_4 \circ \vec{X}_{driver} - \vec{m}_4 \circ \vec{X}_{candidate}|
 \end{aligned}
 \tag{10}$$

$$\begin{aligned}
 x_1 &= x_{attacker} - a_1(d_{attacker}), \\
 x_2 &= x_{barrier} - a_2(d_{barrier}), \\
 x_3 &= x_{chaser} - a_3(d_{chaser}), \\
 x_4 &= x_{driver} - a_4(d_{driver})
 \end{aligned}
 \tag{11}$$

$$x(t + 1) = \frac{x_1 + x_2 + x_3 + x_4}{4}
 \tag{12}$$

It is assumed that the chimps have a 50% probability of choosing between the normal update or the chaotic update mechanism. This is modeled is by Eq.13. In this equation, μ is a random variable between 0 and 1.

$$\vec{X}_{chimp}(t + 1) = \begin{cases} \vec{X}_{prey}(t) - \vec{a} \cdot d, & \text{if } \mu < 0.5 \\ m(t), & \text{if } \mu > 0.5 \end{cases}
 \tag{13}$$

On the other hand, there is an approach in the literature where the next positions of the candidate solutions are calculated with the weighted values of x_1, x_2, x_3 and x_4 [24]. However, the method proposed in this study uses traditional Eq.11 and 12.

B. Adapted Chimp Optimization Algorithm (AChOA)

In the ChOA algorithm, $|a| < 1$ forces chimpanzees to attack the prey (exploitation), while $|a| > 1$ causes chimpanzees to scatter in search of better prey (exploration). This parameter depends on two important sub-parameter changes. The first of these is the random vector r_1 and can be obtained as either uniform or chaotic. The second subparameter is f. In classical ChOA, this value is calculated to be reduced from 2.5 to 0. The adaptability of this sub-parameter will also make the exploration and exploitation mechanisms adaptive. Adaptive metaheuristic approaches are known to provide performance improvement in solving different problems in the literature [17,22]. This study has implemented the ChOA and its adaptive version to the task scheduling problem in cloud systems. As far as is known, this is the first time in the literature. In AChOA, three different

mathematical methods given below are suggested for calculating the f coefficient.

Sigmoid Decreasing Weight (SDW): SDW uses a sigmoid function [17, 22]. This function, given in Eq.14, generates a value by using the upper(U), lower(L) bounds, and the current iteration (t). In this equation, T_{max} is the maximum iteration, and u is calculated with $u = 10^{(\log(T_{max})-2)}$.

$$f(t) = U + \frac{U - L}{(1 + e^{u(0.5 \cdot T_{max} - t)})}
 \tag{14}$$

Oscillating weight (OscW): In this method, a waveform is used for discovery and exploitation processes [18, 22]. The main function of this waveform is given in Eq.15. In this equation, S depends on S_1 and they are calculated by $S_1 = \frac{3T_{max}}{4}$ and $S = \frac{2S_1}{3+2k}$. The k value is a predefined constant.

$$f(t) = \frac{U + L}{2} + \frac{U - L}{2} \text{Cos}\left(\frac{2\pi t}{S}\right)
 \tag{15}$$

V-shape Family (VSW): This function, which is frequently used in binary search PSO algorithms, is not compelling in the displacement of solutions [19, 22]. The general expression of VSW is as in Eq.16.

$$f(t) = \left| \frac{\pi}{2} \text{Tan}^{-1}\left(\frac{\pi}{2} t\right) \right|
 \tag{16}$$

C. Discrete AChOA and Application to the Task Scheduling Problem

The fact that virtual machine representation formats are generally integer encoded allows task scheduling algorithms to be discrete. This study also uses discrete value representation for the task scheduling optimization algorithm. For this reason, the discrete form of AChOA was developed. The discrete optimization of AChOA, which will assign tasks to m VMs with certain MIPS values that have been created before, consists of five steps.

T_1	T_2	...	T_{n-1}	T_n
VM ₇	VM ₃	...	VM ₂	VM ₉
⋮	⋮	⋮	⋮	⋮
VM ₄	VM ₂	...	VM ₃	VM ₈

$p \times n$

Fig.4. A sample initial population

Step 1: Initializing the population: In the first step, n tasks are generated with random MI values within the maximum and minimum limits. According to the number of tasks (n), the initial positions of all chimpanzees are determined. In this step, random candidate solutions are generated for the population of size p as in Figure 4.

Step 2: Creation of Expected Time to Compute Matrix (ETC): In a candidate solution as in Figure 4, the ETC matrix is created by checking the assignment of the tasks to the VM as in Eq.2.

Step 3: Calculation of the maximum makespan: In a population where an ETC matrix is created as in Eq. 2, the maximum makespans are calculated for each candidate solution, as in Eqs. 1-4. The best value found is compared with the scores of the attacker, chaser, barrier, and driver chimps. Then the information of the chimp with the best position and the best score are updated accordingly.

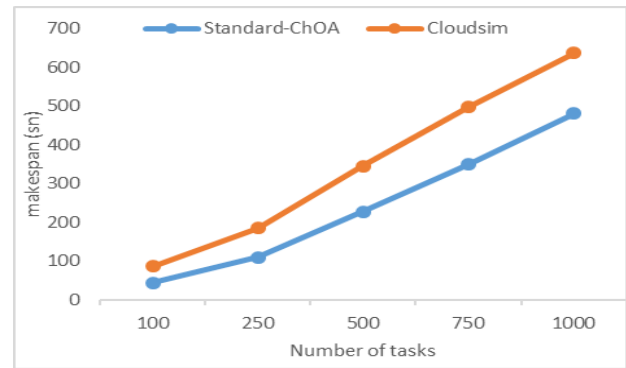
Step 4: Updating the search location: The coefficients are recalculated according to Eqs.7-9. The search position of the entire population is updated by Eqs.10-12. As a result, a new candidate solution is created, replacing the tasks initially randomly assigned to the VMs.

Step 5: Finalizing the iteration: Steps 2-4 are repeated until the maximum number of iterations is reached. As a result of iterations, the attacker score gives the optimum makespan and VM-task match according to AChOA. The pseudocode of the task scheduling process with AChOA is given in Fig. 5.

```

Task scheduling with AChOA
Set initial positions of attacker, barrier, chaser and driver chimps
Generate random tasks and virtual machines based on MI and MIPS
boundaries
Start chimp population  $x_i$  ( $i=1,2, \dots,n$ )
Set  $f, m, a$  and  $c$  values
 $X_{attacker}$  = best search agent,  $P_{attacker}$  = best position
 $X_{barrier}$  = second best search agent,  $P_{barrier}$  = second best position
 $X_{chaser}$  = third best search agent,  $P_{chaser}$  = third best position
 $X_{driver}$  = fourth best search agent,  $P_{driver}$  = fourth best position
while ( $t <$  maximum number of iterations)
  for each chimp:
    Check boundary conditions
    Calculate the fit (max makespan) value based on the objective
function
    if(fitness value  $<$   $X_{attacker}$ )
       $X_{attacker}$  = fitness value
       $P_{attacker}$  = chimp position
    if( $X_{barrier} <$  fitness value  $>$   $X_{attacker}$ )
       $X_{barrier}$  = fitness value
       $P_{barrier}$  = chimp position
    if( $X_{chaser} <$  fitness value  $>$   $X_{attacker}, X_{barrier}$ )
       $X_{chaser}$  = fitness value
       $P_{chaser}$  = chimp position
    if( $X_{driver} <$  fitness value  $>$   $X_{attacker}, X_{barrier}, X_{chaser}$ )
       $X_{driver}$  = fitness value
       $P_{driver}$  = chimp position
  end for
  update  $f$  value
  for each chimp:
    calculate  $a, c, m$  values (Eqs. 7-9)
    update chimp position (Eqs. 10-14)
  end for
   $t=t+1$ 
end while
return  $X_{attacker}$ 
    
```

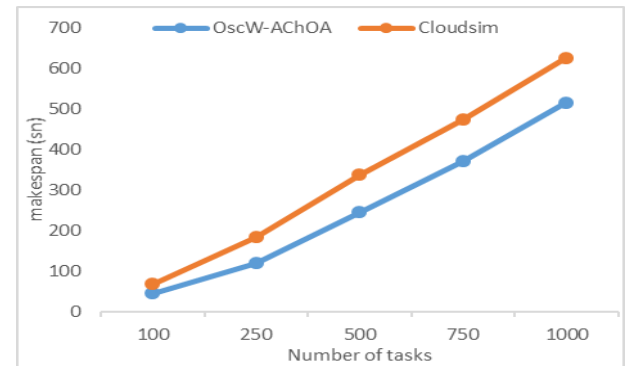
Fig.5. Task scheduling pseudocode with AChOA



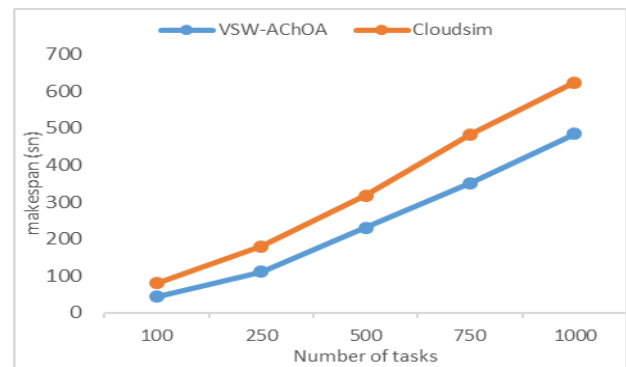
a) Standard-ChOA and Cloudsim makespan results



b) SDW-AChOA and Cloudsim makespan results



c) OscW-AChOA and Cloudsim makespan results



d) VSW-AChOA and Cloudsim makespan results

Figure 6. Comparison of time to completion (makespan) performance for all methods

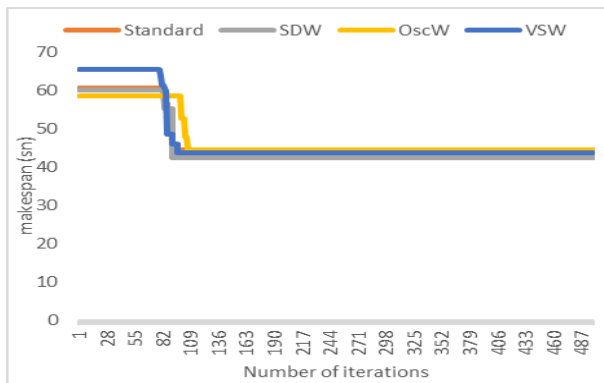
IV. EXPERIMENTS

CloudSim simulator was used to prove the accuracy of AChOA proposed for task scheduling. To be compatible with this simulator, ChOA is written in the JAVA programming language. CloudSim is one of the most popular simulators used by researchers and engineers doing research and development for cloud-related problems [20]. CloudSim simulator enables the modelling of main cloud system entities (data centers, mainframes, VMs, and cloudlets, etc.). Besides, it allows examining scheduling approaches in the created cloud environment [21]. The parameters used for the simulations are as in Table I.

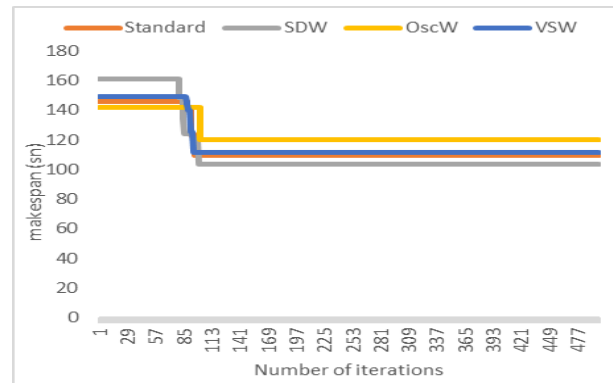
TABLE I
SIMULATION PARAMETERS

Parameter	Value
Population size	50
Number of iterations	500
Number of tasks	100-1000
Number of VMs	20
Task command length (MI)	5000-20000
VM processing capacity (MIPS)	1000-5000

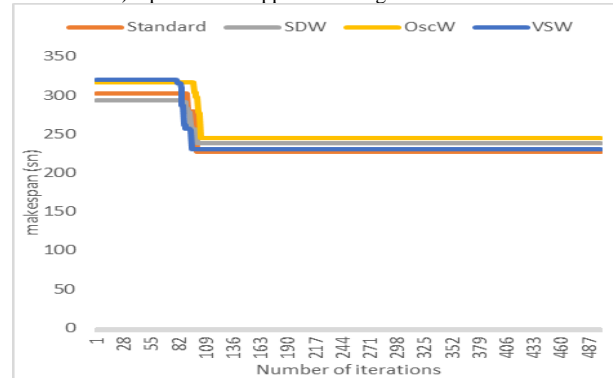
Task scheduling can be made on a real cloud system or a user-defined virtual system on the cloud. In the simulations, custom task scheduling scenarios, which are carried out on a user-defined virtual system on the cloud, have been considered. In the experiments, simulations of AChOA using both standard ChOA and different adaptive functions were run separately. Obtained results are also discussed in comparison with CloudSim default task scheduling results. To observe the performances in different scenarios, each experiment carried out with 100, 250, 500, 750, and 1000 tasks were run 20 times, and statistical results were obtained. In practice, customers generally purchase cloud services on a pay-as-you-go model. Although nowadays there are more advanced server and virtual machine specifications (>100000MIPS, >2GB memory), economical specifications of the hardware used have been chosen according to realistic customer scenario suitability. For this reason, MIPS values are generally close to the standard Intel Pentium specifications, and the task sizes are chosen according to the traditional image processing and scientific algorithms (>5000) used in the literature. First, two main physical servers for VMs were created in the data center, each with 16 GB ram, 10 TB storage, 1 GB/s bandwidth, and time-sharing VM scheduling.



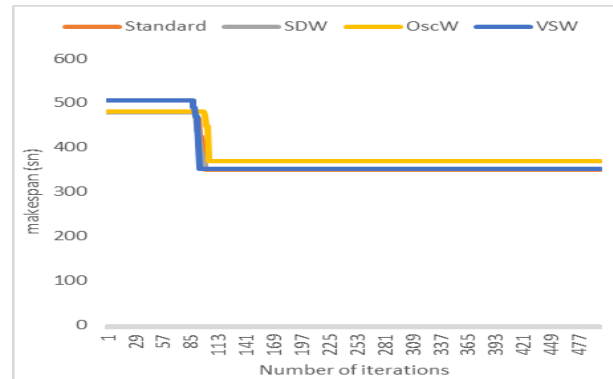
a) Optimization approach changes for 100 tasks



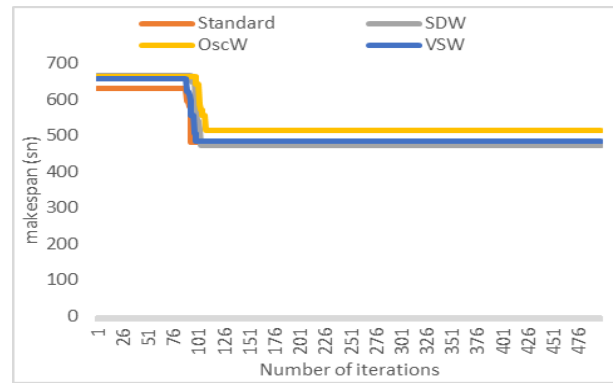
b) Optimization approach changes for 250 tasks



c) Optimization approach changes for 500 tasks



d) Optimization approach changes for 750 tasks



e) Optimization approach changes for 1000 tasks

Figure 7. Optimization approach changes for different number of tasks

All VMs are equally shared on these physical servers. The first of these computers has a 4-core CPU and the second has a dual-core X86 architecture CPU. The processing capacity of each processor core is 10000 MIPS. The computers have Linux operating system and Xen VMM. VMs have 512 MB ram, 10 GB storage, 10 MB/s bandwidth, and time-sharing task scheduling configuration. The processing capacity of the VMs varies between 1000 and 5000 MIPS. The instruction length of the tasks varies between 5000 and 20000 MI. On the CloudSim side, the standard task scheduling method *CloudletSchedulerSpaceShared* is used.

In the experimental results, firstly, the makespan performances of all methods were examined. In Fig.6, the simulation results of each method and CloudSim are given.

Since the task MI values are randomly determined in each simulation, the same task types in that simulation are also given to CloudSim for accurate evaluation. Therefore, comparisons are given separately in the four charts. The general evaluation between the methods is presented with the statistical data in Table II.

While SDW found the minimum makespan in three of the experiments performed with five different task numbers, the Standard method caught the best makespan in the other two. In all experiments performed, it was seen that all variations of ChOA achieved better results than CloudSim's standard space shared task scheduling process. Considering all the task numbers in this experiment, SDW-AChOA 35.5%, standard ChOA 32.2%, OscW-AChOA 27.2%, and VSW-AChOA 32% improved according to CloudSim. According to these average values, it is seen that the SDW method is better in performance. Another result obtained from the experiments is that as the number of tasks increases, the percentage of success decreases according to *CloudletSchedulerSpaceShared*. For example, while the improvement rate compared to CloudSim was 48.5% in the experiment with 100 tasks, which is performed by the Standard method, this rate decreased to 24.5% in the

experiment with 1000 tasks. As the number of tasks increased, the most significant decrease was 50.6% in VSW-AChOA compared to CloudSim. While the improvement rate was 45.44% in the 100-tasks experiment performed with VSW-AChOA, this value decreased to 22.43% in the 1000-tasks simulation. However, while the decrease in the amount of improvement due to the increase in the number of tasks in OscW-AChOA was 49.18%, this rate was 49.46% in the Standard method. When this example was evaluated on an experimental basis, the consistent method was seen as SDW-AChOA.

To discuss the performance of all methods relative to each other in a more general perspective, it is necessary to evaluate the statistical results obtained from 20 independent experiments and given in Table II. Considering all the experimental results, it is seen that SDW -AChOA is more successful in reaching the minimum value for 100 and 250 tasks. However, this performance decreased as the number of tasks increased, even worse than other results. It has been seen that Standard ChOA is more successful at high task numbers. Another method that performs well for high tasks is VSW-AChOA. In fact, based on the mean value criterion, VSW-AChOA can be seen as relatively more successful than standard ChOA. As a result, SDW-AChOA performed better on small task numbers, while Standard and VSW-AChOA were successful in experiments involving a high number of tasks. Although all methods were more successful than standard scheduling methods, OscW had the lowest performance among AChOA variations.

A large number of candidate variables (tasks) also affected the approach to the best solution in the optimization process. The convergence to the best in all methods generally occurred after the 70th iteration, and the best was reached after about 50 iterations. Fig.7 shows the Standard, SDW, OscW, and VSW approximation results in the experiments.

TABLE II
STATISTICAL RESULTS OF EXPERIMENTS FOR DIFFERENT TASK NUMBERS

Method		100 tasks	250 tasks	500 tasks	750 tasks	1000 tasks
Standard-ChOA	Minimum	44,4	109,5	227,4	349,3	480,1
	Maximum	74,1	177,7	346,8	511,8	680,4
	Mean	51,2	126,4	261,5	399,4	538,8
	Standard Deviation	2,3	4,2	5,5	8,3	9,0
	Median	47,8	119,3	249,2	381,9	515,6
SDW-AChOA	Minimum	42,5	103,7	238,2	350,8	471,9
	Maximum	71,8	172,4	349,7	530,1	705,8
	Mean	50,9	128,1	262,1	397,8	534,9
	Standard Deviation	1,8	4,0	4,9	10,1	12,0
	Median	48,5	121,2	247,3	377,8	507,6
OscW-AChOA	Minimum	44,4	120,0	245,4	370,0	515,0
	Maximum	73,6	178,8	347,9	518,5	683,5
	Mean	52,5	136,1	280,1	420,7	571,8
	Standard Deviation	1,6	4,4	7,2	8,0	13,5
	Median	49,0	131,6	268,3	406,6	554,8
VSW-AChOA	Minimum	43,6	111,9	230,9	351,1	483,5
	Maximum	72,7	181,4	353,8	520,1	695,2
	Mean	50,8	127,3	263,5	393,8	538,8
	Standard Deviation	2,5	4,3	5,6	9,6	9,2
	Median	47,8	122,2	248,7	375,4	514,7

V. CONCLUSION

This study focused on task scheduling optimization, which is one of the most important problems in cloud computing. The solution to this problem was realized with AChOA, which is an improved adapted version of ChOA that is a new metaheuristic algorithm. AChOA uses functions that could make the exploration and exploitation mechanisms of the standard ChOA more flexible. Thus, this adapted metaheuristic algorithm can run in three different states: SDW-AChOA, OscW-AChOA, VSW-AChOA. The success of all methods was also proven in CloudSim, a well-known and successful cloud simulator in the literature. All metaheuristic methods achieved about 30% improvement over the standard scheduling method. Although the improvement decreased with the increase in the number of tasks, this rate was over 20%. In comparisons among themselves, it was observed that the results of SDW-AChOA in low task numbers and standard ChOA and VSW-AChOA in scheduling with high task numbers were more successful. Cloud systems will continue to be a research area where different problems will arise for a long time. For this reason, the authors will focus on the solutions of different optimization methods in cloud systems in their future studies.

REFERENCES

- [1] Strumberger, I., Tuba, E., Bacanin, N., & Tuba, M. (2019, June). Dynamic tree growth algorithm for load scheduling in cloud environments. In 2019 IEEE Congress on Evolutionary Computation (CEC) (pp. 65-72). IEEE.
- [2] Avram, M. G. (2014). Advantages and challenges of adopting cloud computing from an enterprise perspective. *Procedia Technology*, 12, 529-534.
- [3] Abdullahi, M., & Ngadi, M. A. (2016). Symbiotic organism search optimization based task scheduling in cloud computing environment. *Future Generation Computer Systems*, 56, 640-650.
- [4] Houssein, E. H., Gad, A. G., Wazery, Y. M., & Suganthan, P. N. (2021). Task scheduling in cloud computing based on meta-heuristics: review, taxonomy, open challenges, and future trends. *Swarm and Evolutionary Computation*, 62, 100841.
- [5] Khishe, M., & Mosavi, M. R. (2020). Chimp optimization algorithm. *Expert systems with applications*, 149, 113338. <https://doi.org/10.1016/j.eswa.2020.113338>
- [6] Pradhan, A., Bisoy, S. K., & Das, A. (2021). A survey on pso based meta-heuristic scheduling mechanism in cloud computing environment. *Journal of King Saud University-Computer and Information Sciences*.
- [7] Saurav, S. K., & Benedict, S. (2021, January). A Taxonomy and Survey on Energy-Aware Scientific Workflows Scheduling in Large-Scale Heterogeneous Architecture. In 2021 6th International Conference on Inventive Computation Technologies (ICICT) (pp. 820-826). IEEE.
- [8] Alsaidy, S. A., Abbood, A. D., & Sahib, M. A. (2020). Heuristic initialization of PSO task scheduling algorithm in cloud computing. *Journal of King Saud University-Computer and Information Sciences*.
- [9] Bacanin, N., Bezdán, T., Tuba, E., Strumberger, I., Tuba, M., & Zivkovic, M. (2019, November). Task scheduling in cloud computing environment by grey wolf optimizer. In 2019 27th Telecommunications Forum (TELFOR) (pp. 1-4). IEEE.
- [10] [Chen, X., Cheng, L., Liu, C., Liu, J., Mao, Y., & Murphy, J. (2020). A woa-based optimization approach for task scheduling in cloud computing systems. *IEEE Systems Journal*, 14(3), 3117-3128.
- [11] Belgacem, A., Beghdad-Bey, K., & Nacer, H. (2018, October). Task scheduling optimization in cloud based on electromagnetism metaheuristic algorithm. In 2018 3rd International Conference on Pattern Analysis and Intelligent Systems (PAIS) (pp. 1-7). IEEE.
- [12] Aziza, H., & Krichen, S. (2018). Bi-objective decision support system for task-scheduling based on genetic algorithm in cloud computing. *Computing*, 100(2), 65-91.
- [13] Li, K., Xu, G., Zhao, G., Dong, Y., & Wang, D. (2011, August). Cloud task scheduling based on load balancing ant colony optimization. In 2011 sixth annual ChinaGrid conference (pp. 3-9). IEEE.

- [14] Chen, X., & Long, D. (2019). Task scheduling of cloud computing using integrated particle swarm algorithm and ant colony algorithm. *Cluster Computing*, 22(2), 2761-2769.
- [15] Liu, C. Y., Zou, C. M., & Wu, P. (2014, November). A task scheduling algorithm based on genetic algorithm and ant colony optimization in cloud computing. In 2014 13th International Symposium on Distributed Computing and Applications to Business, Engineering and Science (pp. 68-72). IEEE.
- [16] Tsai, C. W., Huang, W. C., Chiang, M. H., Chiang, M. C., & Yang, C. S. (2014). A hyper-heuristic scheduling algorithm for cloud. *IEEE Transactions on Cloud Computing*, 2(2), 236-250.
- [17] Malik, R. F., Rahman, T. A., Hashim, S. Z. M., & Ngah, R. (2007). New particle swarm optimizer with sigmoid increasing inertia weight. *International Journal of Computer Science and Security*, 1(2), 35-44.
- [18] Bansal, J. C., Singh, P. K., Saraswat, M., Verma, A., Jadon, S. S., & Abraham, A. (2011, October). Inertia weight strategies in particle swarm optimization. In 2011 Third world congress on nature and biologically inspired computing (pp. 633-640). IEEE.
- [19] Mafarja, M., Aljarah, I., Faris, H., Hammouri, A. I., Ala'M, A. Z., & Mirjalili, S. (2019). Binary grasshopper optimisation algorithm approaches for feature selection problems. *Expert Systems with Applications*, 117, 267-286.
- [20] Roth, G., & Dicke, U. (2005). Evolution of the brain and intelligence. *Trends in cognitive sciences*, 9(5), 250-257.
- [21] Calheiros, R. N., Ranjan, R., Beloglazov, A., De Rose, C. A., & Buyya, R. (2011). CloudSim: a toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms. *Software: Practice and experience*, 41(1), 23-50.
- [22] Yildirim, G., Alatas, B. New adaptive intelligent grey wolf optimizer based multi-objective quantitative classification rules mining approaches. *J Ambient Intell Human Comput* (2021). <https://doi.org/10.1007/s12652-020-02701-9>
- [23] Koyuncu, H. GM-CPSO: A New Viewpoint to Chaotic Particle Swarm Optimization via Gauss Map. *Neural Process Lett* 52, 241–266 (2020). <https://doi.org/10.1007/s11063-020-10247-2>Electromagnetic Fields (300 Hz to 300 GHz), *Environmental Health Criteria* 137, World Health Organization, Geneva, Switzerland, 1993.
- [24] Khishe M, Nezhadshahbodaghi M., Mosavi M.R., Martín D., "A Weighted Chimp Optimization Algorithm," in *IEEE Access*, vol. 9, pp. 158508-158539, 2021, doi: 10.1109/ACCESS.2021.3130933.

BIOGRAPHIES



EMRULLAH GUNDUZALP graduated from Firat University, Department of Computer Engineering in 2007 with a bachelor's degree. In 2018, he completed his master's degree in Firat University. He worked at the General Directorate of Land Registry and Cadastre between 2007-2009. Since 2009, he has been working as a computer engineer at the 9th Regional Directorate of DSI, IT department. His research interests are the propagation of electromagnetic waves, machine learning, and optimization.



GUNGOR YILDIRIM received the B.Sc. degree from Firat University, Turkey, in electrical and electronic engineering. He received the M.Sc. and the Ph.D. degrees from computer engineering of Firat University, in 2012, and 2017, respectively. Currently, he is an assistant professor at the Department of Computer Engineering at Firat University. He worked as a project and control engineer in private sectors and a public institution (DSI). He served as the head of the electric program at Tunceli MYO for three years. In 2014, he received the Award

of Appreciation in DSI. His research interests include wireless systems, wireless sensor networks and IoT systems.



YETKIN TATAR received the B.Sc. degree from EDMMA in 1974 and the M.Sc. and D.Sc. degrees in electrical and electronic engineering from Firat University, Turkey, in 1984 and 1994, respectively.

He worked as a Professor with the Department of Computer Engineering, Firat University. His research areas are wireless sensor networks, computer networks, and network security.

Ear semantic segmentation in natural images with Tversky loss function supported DeepLabv3+ convolutional neural network

Tolga Inan, Umit Kacar

Abstract—Semantic segmentation is a fundamental problem for computer vision. On the other hand, for studies in the field of biometrics, semantic segmentation is gaining more importance. Many successful biometric recognition systems require a high-performance semantic segmentation algorithm. This study presents an effective ear segmentation technique in natural images. A convolutional neural network is trained for pixel-based ear segmentation. DeepLab v3+ network structure, with ResNet-18 as the backbone and Tversky lost function layer as the last layer, has been trained with natural and uncontrolled images. The proposed network training is performed using only the 750 images in the Annotated Web Ears (AWE) training set. The corresponding tests are performed on the AWE Test Set, University of Ljubljana Test Set, and the Collection A of In-The-Wild dataset. For the Annotated Web Ears (AWE) dataset, intersection over union (IoU) is measured as 86.3% for the AWE database. To the best of our knowledge, this is the highest performance achieved among the algorithms tested on the AWE test set.

Index Terms—Semantic Segmentation, Ear Segmentation, Biometrics, Convolutional Neural Networks, Tversky Loss Function.

I. INTRODUCTION

BIOMETRIC systems have become an integral part of our lives for many years, and are now widely used by forensic, security, and law enforcement agencies. To use biometric systems more easily in daily life and achieve higher performance, researchers have focused on end-to-end and fully automated solutions. In particular, facial and ear biometrics come to the fore in the public sphere because the face and ear biometrics can be recorded in a non-cooperative manner.

As with the recognition studies in many biometric modalities, studies in ear recognition were first conducted with data recorded in a controlled environment. There are many ear recognition studies using ear datasets recorded in controlled environments[1], [2]. However, given the difficulties in practical applications, these studies lack applicability to real-world scenarios. In 2017, to overcome this drawback, the Unconstrained Ear Recognition Challenge (UERC) competition was organized by [3]. The UERC competition was held

once again in 2019 [4], and the competitors were given an unconstrained ear dataset. This database is open to the public and available to the entire research community. Since the UERC competition's main focus was ear recognition, the ear images were manually cropped and given to the participants and the research community. With this aspect, the competition database was unsuitable for ear detection studies.

It is widely accepted that ear detection is the first and most crucial step in the ear biometric recognition line[5]. With the semantic segmentation method, the localization accuracy at the ear detection stage can be increased. This article provides an effective method for ear segmentation. Our contribution to this problem is twofold. The first is to choose a suitable model for training supported by the augmentation of the data and the selection of the model parameters. Our second contribution is to choose an appropriate loss function for the ear segmentation problem.


The rest of the paper is as follows: Chapter 2 summarizes relevant studies, Chapter 3 introduces the Proposed Method, Chapter 4 and Chapter 5 describe Experimental Setup and Experimental Results, respectively. The article ends with the Conclusion section in Chapter 6.


II. RELATED WORK

Object detection is a fundamental task in computer vision, and the primary purpose of object detection is: "Which object is where?"[6]. In the object recognition problem, the solution algorithm tries to find the rectangle that will best encircle the object. In this type of research, marked data sets are required for the algorithms' training and testing phases. In these data sets, the positions of the objects are given by the rectangles surrounding them. On the other hand, semantic segmentation studies aim to classify the object-class at the pixel level. The semantic segmentation algorithm classifies pixels and labels pixels as the object-class or the background-class. In this study, the object-class is chosen as the human ear, and the scope is limited to ear semantic segmentation.

A. Convolution Neural Network based approaches for Semantic Segmentation

In semantic segmentation studies, convolutional neural network-based approaches stand out. In this context, the studies trying to solve the semantic segmentation problem with the convolutional neural network approach are briefly mentioned in this section. The study by [7] focuses on designing a deep

 **Tolga Inan** is with the Department of Electrical - Electronics Engineering, Cankaya University, Ankara, 06790 TURKEY e-mail: tolga.inan@cankaya.edu.tr

 **Umit Kacar** is with the Department of Electrical - Electronics Engineering, Cankaya University, Ankara, 06790 TURKEY e-mail: umitkacar@itu.edu.tr

Manuscript received Nov 15, 2021; accepted July 18, 2022.
DOI: [10.17694/bajece.1024073](https://doi.org/10.17694/bajece.1024073)

neural network architecture with low latency operation for the semantic segmentation problem. They build a solution applicable to real-time applications. They report more than 10 fps image segmentation speed with an input image resolution of 640x360 for the practical road scene parsing datasets.

One of the earlier attempts to use convolutional neural networks (CNN) for semantic segmentation problem is the study of [8]. This study represents a fully deep CNN with the support of a probabilistic graphical model. The location-invariant nature of the features generated by convolutional neural networks may result in poor localization for semantic segmentation tasks. To overcome this bottleneck of poor localization, [8] employs a fully-connected conditional random field (CRF) as a probabilistic graphical model to extract spatial dependence in the semantic segmentation problem. The aforementioned study reports 71.6 % Intersection over Union (IoU) accuracy as the state-of-art, for the PASCAL VOC-2012 [9] semantic image segmentation test set.

[10] propose a solution called RefineNet, a multi-path refinement network, and they report an intersection-over-union score of 83.4 on the PASCAL VOC-2012 dataset. [11] propose a semantic segmentation framework with the capability of handling zero-labeled and few-labeled object classes, improving their approach which was reported in [12]. Their method employs indirect information acquired from semantic space via the semantic projection network. They report their results for the zero-label and few-label learning semantic segmentation experiments conducted on COCO-Stuff [13] and PASCAL VOC12 [9] datasets.

[14] propose a convolutional network for semantic segmentation with attention support. Their study facilitates a criss-cross network structure to capture horizontal and vertical contextual information around a particular pixel. The harvesting of contextual information is repeated so that each pixel can finally capture the dependencies from all pixels. They report the mean intersection over union (mIoU) 81.4 and 45.22 as scores on Cityscapes [15] test set and ADE20K [16] validation set, respectively. The proposed method is claimed to be both memory and computation effective for GPU implementation with the state of the art performance. In the study by Chen et al. [17], they use the dilated (atrous) convolution approach with an atrous spatial pyramid pooling module to extract multi-scale features. The proposed system, also known as, DeepLabv3 demonstrates competitive performance on PASCAL VOC12 [9] dataset.

The study by [18] proposes a task called panoptic segmentation to integrate the distinct tasks of semantic segmentation and instance segmentation. Hence, their study attempts to assign pixel-level labels and to detect each object instance simultaneously. To report panoptic segmentation performance, the authors represent a new metric named as panoptic segmentation metric. Panoptic segmentation metric is a hybrid metric that unifies the effects of the segmentation quality and the recognition quality.

The recent attempt by [19] focuses on the neural architecture search. The search for neural architecture is hierarchical and twofold: cell level search and network-level search. They avoid the hand-designing of the higher-level network struc-

ture. The problem formulation is continuous and allows the gradient-based architecture search. They report state-of-the-art performance on Cityscapes, PASCAL VOC12, and ADE20K datasets without any pretraining on ImageNet [20]. The study by [21] uses an adversarial training to learn from unlabeled data to achieve pixel-level classification. They report state-of-the-art performance for semi-supervised learning.

For further information on semantic segmentation, the reader is referred to detailed survey papers [22], [23], [6], [24], [25] on semantic segmentation using deep learning techniques.

B. Datasets for Ear Segmentation

Semantic segmentation approaches are usually supervised learning approaches; therefore, they require a sufficient amount of labeled data for the training stage and the test stage's performance measurement. In this section, four databases that are commonly used for ear segmentation studies are pointed-out.

[26] shared the Annotated Web Ears (AWE) dataset. This database consists of 1000 images (750 training images and 250 test images) of 100 subjects. They had collected the images from the web by a semi-automatic procedure. The largest image's size in the database is 473-by-1022 pixels, whereas that of the smallest image is 15-by-82 pixels. The average image size is reported to be 83-by-160 pixels. The binary masks indicating ear and non-ear classes are also distributed with the database.

[27] used 12500 images from the web, collected by the researchers and students from University of Ljubljana. The images are unconstrained, and they have different resolutions. They generate a semi-automatic procedure that obtains a pixel-wise class label (ear and non-ear) masks. They use the RefineNet [10] method to generate pixel-wise class label masks and manually examine the results for valid masks. Therefore they can conveniently extend the database size reported in [27]. At the date, the database have been obtained, there were 16765 pixel-wise class label masks available for the researchers.

[28] shared the UBEAR database. This database consists of 4412 gray-scale images and their pixel-wise class label masks. The images are recorded in dynamic lighting conditions with on-the-move subjects. The subjects did not pay attention to ear occlusions and their poses during the recording sessions.

[29] shares "In-the-Wild" database with images from collected Google Images. The images are gathered with ear-related tags, and the identities are not known. In this study, the Collection A set, in which the ear is manually annotated, is used. Fifty-five anatomically distinct landmark points around the ear regions (ascending helix, descending helix, helix, ear lobe, tragus, canal, antitragus, concha, etc.) are manually annotated. Collection A set includes 605 images, and the images are randomly divided into two disjoint sets of Collection A-training (500 images) and Collection A-testing (105 images). For the images shared by [29], 55 points are marked manually, but the masks for ear semantic segmentation are not available. These masks were obtained with the following procedure. A convex-hull is created from the 55 marked points. The points that determined the convex-hull are the outermost ones. The

outermost points helped to determine the outer border of the ear. The outer points were connected with the line segments, and the outer border was utterly determined. With the help of this determined closed exterior, the masks for semantic segmentation have been created.

C. Convolution Neural Network based approaches for Pixel-wise Ear Semantic Segmentation

In this section of the paper, four studies devoted specifically to the pixel-wise ear segmentation will be briefly outlined. This first one is by [30]. They propose a pixel-wise ear detection approach based on their convolution encoder-decoder network. Encoder-decoder architectures consist of two main stages. In the first stage, the input image is encoded to an abstract representation with the help of convolutional and pooling layers. Whereas in the second stage, the abstract representation is decoded into the desired output format. In [30], the detection pipeline has the assumption that there is a single face in the input image, and the aim is to detect at most two ears. The convolutional encoder-decoder is reported to perform well on image inputs that are recorded totally in unconstrained environments. They test the performance of their pixel-wise ear segmentation network on AWE dataset [26] and report the average accuracy, the average IoU, the average precision, the average recall as 99.4 %, 55.7%, 67.7% and 77.7%, respectively.

The second work is a recent paper by [27]. They propose a method called Mask R-CNN for the pixel-wise ear segmentation. The Mask R-CNN method consists of five stages: Convolutional backbone architecture, region proposal network, region of interest classifier, bounding box regressor, and detection pixel-wise masks. The performance of the method is tested on AWE dataset [26]. The performance indicators on the AWE dataset are as follows. The average IoU is 79.24%, the average precision is 92.04%, and the average recall is 84.14%.

Most of the biometric recognition studies include the detection and/or segmentation stage. The third study summarized as a pixel-wise ear segmentation approach given in [5] presents a complete ear recognition pipeline. [5] includes the pixel-wise ear segmentation stage and the ear segmentation is based on RefineNet [10]. RefineNet is an effective semantic segmentation network using residual convolution units, multi-resolution fusion, and chained residual pooling to achieve high-performance semantic segmentation. The ear segmentation performance is reported on the AWE dataset [26] and the corresponding performance figures for the average accuracy, the average IoU, the average precision, the average recall are 99.8 %, 84.8%, 91.7% and 91.6%, respectively.

The most recent study related to ear semantic segmentation is Context-aware Ear Detection Network (ContextedNet) [31]. ContextedNet has two stages. The first stage is a context-provider, and it generates the probability maps for facial regions. The second stage is the semantic segmentation stage, supported by the probability maps obtained in the first stage. The corresponding performance figures on the AWE dataset are 99.74 %, 81.46%, 89.07%, and 87.47%, for the average accuracy, the average IoU, the average precision, the average

recall, respectively.

Although this study is about ear segmentation in pixel resolution, two other types of ear detection approaches are briefly mentioned in the following two sections. Firstly, the examples for the studies that detect curved boundaries or landmarks of the ear are given. Moreover, finally, the studies in the literature that detect the ear as a rectangular region are mentioned.

D. Studies those Detect Curved Boundaries or Landmarks of the Ear

[29] represents holistic and patch-based statistical deformable models to localize the ear landmarks. This holistic model is based on a shape, appearance, and deformation models. Inversion compositional algorithm, an efficient variant of gradient descent, is used for incremental wrapping. They also offer a patch-based active appearance model. Both holistic and patch-based models are tested on the 55-point ear landmark database. The 55-point ear landmark database has been prepared as a part of this study. Fitting accuracy for the landmark points is reported for the aforementioned database. [32] trains a small-sized CNN to localize 45 landmarks on the ear. They claim that the proposed method is robust to occlusion to some extent. [33] propose a method for finding physiological curves for the ear. The main focus of this study is finding interval curves of the ear (helix, antihelix, concha auricularae).

E. Studies those Detect the Rectangular Boundary for the Ear

The study of [34] offers an approach to detects ears from 2D images. Their method is based on the ensemble of convolutional neural networks (CNN). Three different CNNs had been trained for this purpose. In the next step, the weighted average of the outputs from three different CNNs was used, and ear detection was performed in this way. It has been reported that the performance achieved in this way is higher than the performance achieved with a single model.

In [35], a method based on Faster R-CNN is proposed for detecting ears in 2D images. They claim that they have improved the original R-CNN algorithm using multiple scales, resulting in faster and more accurate system performance.

Paper by [36] addresses the critical role of ear detectors that can work in unconstrained environments to have reliable biometric recognition systems. To tackle with ear detection problem in the wild, they offer two context-aware detection models based on CNNs. In this study, an accuracy of 99 % is reported at IOU 0.5 for the majority of the datasets.

The study proposed by [37] is an ear detection system based on CNN. Instead of employing a single CNN, they prefer to use three CNN, which are trained for different scales of ear images. The scales are obtained by cropping the ear region from the image with the small, medium, and large windows. The proposes are displayed to have higher performance than the Haar Cascade classifier. No other comparisons with recent methods are reported.

III. PROPOSED METHOD

A. Problem Definition

Semantic segmentation is basically the classification of images at the pixel level. A system that successfully performs semantic segmentation needs to assign the class label to each pixel correctly. There are many proven, cutting-edge models of semantic segmentation in the literature. Most of these models offer solutions for semantic segmentation problems for objects (trees, automobiles, animals, buildings, roads, etc.) that are frequently encountered in daily life. However, special adjustments and improvements are required to solve certain problems such as ear detection. These improvements can be expressed as follows:

- Obtaining the appropriate model for transfer learning
- Loss function selection

In this study, an efficient ear segmentation model was created by considering the items summarized above. Since it has been confirmed in the literature that transfer learning increases model accuracy in a shorter training period with fewer data, the transfer learning is used in model training. The following sections describe the improvements at each step.

B. Obtaining the appropriate model for transfer learning

DeepLab v3+ [38] is one of the state-of-art deep learning models for semantic image segmentation, where the goal is to assign semantic labels to every pixel in the input image. This study uses the DeepLab v3+ encoder-decoder network, and pre-trained model ResNet-18 [39] to make an effective start to the training process. ResNet-18 architecture requires less computing resources compared to ResNet-50. This advantage of the ResNet-18 architecture played an important role in the choice of this architecture and makes the overall system more efficient.

C. Loss function selection

The selection of the loss function is a fundamental issue having a significant effect on CNNs' performances. Many studies [40], [41], [42], [43], [44] have examined the importance of loss functions for semantic segmentation. Especially in Jadon's study [41], different error functions are discussed. They point out that the Tversky loss function generally generates optimal results. In this study, the Tversky loss function proposed by [45] is used in the last layer of our network. Tversky loss function is given Equation 1. In Equation 1, P and G are the sets of predicted and ground truth binary labels. On the other hand, α and β are the control parameters of the penalties for false positives and false negatives, respectively.

$$S(P, G; \alpha, \beta) = \frac{|PG|}{|PG| + \alpha|P \setminus G| + \beta|G \setminus P|} \quad (1)$$

Tversky loss function [45] is proposed for tackling the negative effects of the data imbalance in medical applications like lesion segmentation. In ear semantic segmentation, a similar problem arises. Only a small number of pixels belong to the ear class in the training and testing images. Considering this fact, the Tversky loss function is used to increase the ear segmentation network's overall performance.

IV. EXPERIMENTAL SETUP

A. Data Augmentation

The performance of supervised training algorithms increases with the variety of training data. Data augmentation is used to increase the variety of training data to support our training process. Two ways of data augmentation are used. The first one is flipping the training image vertically at random (with 0.5 probability). Correspondingly, so that balanced training is carried out for human faces randomly oriented to the right and the left. The second way for data augmentation is changing the input image's scale randomly. The training images are scaled with a factor chosen randomly from [0.8, 1.2] closed interval. These two augmentation methods contribute to the increase of data diversity in the training set.

B. Evaluation Metric

Ear semantic segmentation is a pixel-based ear detection problem. Therefore, definitions for detection performance can be used in reporting semantic segmentation performance. True Positive (TP) refers to a pixel correctly labeled as the ear-class, while True Negative (TN) is used for pixels that are correctly labeled as the background. False Positive (FP) indicates a background pixel labeled as an ear. False Negative (FN) indicates an ear pixel labeled as the background. The other performance parameters are calculated in Equations 2-9 and the performance reporting is carried out with these calculated parameters based on these definitions of the TP, TN, FP, and FN.

$$Accuracy = 100 \times \frac{(TP + TN)}{(TP + TN + FN + FP)} \quad (2)$$

$$Precision = 100 \times \frac{TP}{(TP + FP)} \quad (3)$$

$$Recall = 100 \times \frac{TP}{(TP + FN)} \quad (4)$$

$$IoU = 100 \times \frac{TP}{(TP + FN + FP)} \quad (5)$$

$$F1 \text{ Score} = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (6)$$

$$FP \text{ Rate (FPR)} = 100 \times \frac{FP}{(FP + TN)} \quad (7)$$

$$FN \text{ Rate (FNR)} = 100 \times \frac{FN}{(FN + TP)} \quad (8)$$

$$E2 = \frac{FPR + FNR}{2} \quad (9)$$

C. Training Set

Increasing data diversity in semantic segmentation works helps to increase performance. However, to make a fair comparison with other ear semantic segmentation studies in the literature, the training set is limited to the training set (750 images) in the AWE dataset [26]. Throughout this study, only the training set in the AWE database was used for training purposes.

A label mask is also used for each training image. The label mask is a binary image showing labels (ear and background) for each pixel and has the same number of rows and columns as the training image.

D. Preliminary Experiments for Training Parameters Selection

The training performance of the CNNs is related to the initial conditions of the training process. It is not always possible to select all parameters optimally. In this study, some parameters are kept constant while some other parameters are chosen in a way that will increase the performance. In the following, the fixed parameters and the parameters selected to achieve high-performance objective are briefly described.

1) *Fixed Parameters*: The first fixed parameter is the size of the input image. The input image is set to 480 x 640 pixels. Another fixed parameter used is the mini-batch size. Due to the limit of the memory capacity of the GPU, the mini-batch value is set to 32.

2) *Parameters selected to achieve high-performance objective*: In this section, the process for selecting the parameters in order to design a high-performance ear semantic segmentation system is described.

Three parameters are defined as variables. The first of these parameters is the control parameters of the penalties for false positives parameter, that is α . Seven different values are considered for the α value. These values are 0.35, 0.40, 0.45, 0.50, 0.55, 0.60, 0.65, respectively.

The second parameter to be selected is the initial learn rate parameter. Three different values have been considered for the initial learn rate parameter. These values are 0.001, 0.0005 and 0.0001, respectively.

The third and last parameter to choose is the weight decay parameter. Two different values have been considered for the weight decay parameter. These values are 0.0005 and 0.0001, respectively.

3) *Preliminary experiments to select parameters to achieve high-performance objective*: Using the aforementioned parameter values, 42 (7x3x2) different parameter sets are determined.

For preliminary experiments, the preliminary-experiment-training-set and preliminary-experiment-test-set are defined. To make a fair comparison with the studies in the literature, only the AWE training set can be used at this step. So, first 500 images in the AWE training set are labeled as preliminary-experiment-training-set. The last 250 images in the AWE training set are labeled as preliminary-experiment-test-set. Therefore, only 750 images of AWE training set are used in preliminary experiments for parameter selection.

For preliminary experiments, the epoch number was kept

constant as 2. Separate experiments were performed for 42 (7x3x2) parameter sets. Precision and accuracy graphs of these experiments are given in Figure 1. In Figure 1, axes represent α , initial learn rate and weight decay, respectively. Precision and accuracy values are expressed in the color codes given in the legend.

After evaluating the experimental results given in Figure 1, α was chosen as 0.55, initial learn rate as 0.001, and weight decay as 0.0001. Since α is set to 0.55; $\beta = (1 - \alpha)$, the control parameter of the penalties for false negatives, is calculated as 0.45. In this way, it is aimed to increase the system performance.

This approach has been used to make a more reasonable choice among the possible parameters. Therefore; we do not claim that the parameters chosen in this way are optimal. Care has been taken to use only the images in the AWE Training Set while making the parameter selection.

E. Test Sets

1) *AWE Test Set*: The test data were prepared by combining 250 images from the AWE data set [26]. Label masks given for these images were also used for performance calculations. Since this test set has been widely used in other studies in the field of ear semantic segmentation, using this test set helps us compare our ear semantic segmentation solution's performance realistically with other studies.

2) *University of Ljubljana (UL) Test Set*: This test set consists of 16765 images from the University of Ljubljana [27] dataset.

3) *UBEAR Test Set*: This test set is prepared by 4412 images from the UBEAR dataset [28].

4) *Wild-1 Test Set*: This test set consists of 500 training images of Collection A of "In-the-Wild" database gathered by [29].

5) *Wild-2 Test Set*: This test set consists of 105 testing images of Collection A of "In-the-Wild" database gathered by [29].

V. EXPERIMENTAL RESULTS

A. Hardware and Software

The experiments were carried out on a personal computer with an AMD Ryzen 7 2700x processor and 32 GB of RAM. There is an Nvidia RTX 2080Ti GPU card in the setup. The operating system of the personal computer is Ubuntu 20.04 LTS. Training and tests are carried out in Matlab 2021a environment. The ADAM algorithm [46] is used as an optimizer.

B. Training with the Selected Parameters

The semantic segmentation network is trained with the parameters defined in Section IV-D. Training is performed for ten epochs with 750 images in the AWE training dataset [26]. The calculated loss values for ten epochs are shown in Figure 6.

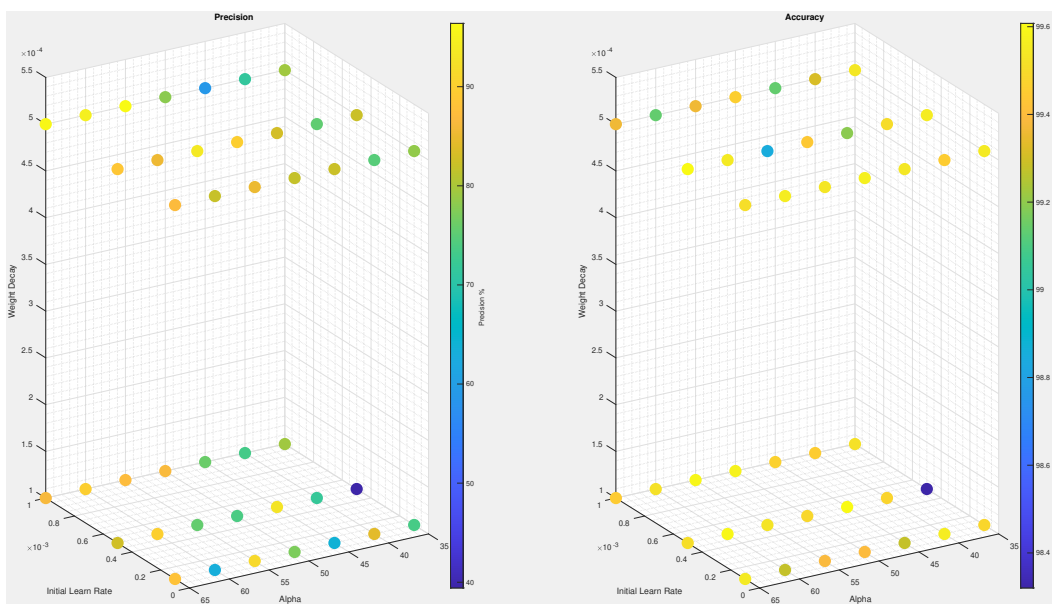


Fig. 1. Precision and Accuracy for the preliminary experiments on 250 images from AWE Training Dataset

C. Test Results for Ear Semantic Segmentation

1) *AWE Test Set*: The performance test has been carried on 250 test images on AWE dataset [26]. The performance results are summarised in Table I. The accuracy results are very high in all studies, including ours. This is basically due to the fact that the background-class labeled pixels are dominating in number in most of the images. A major performance metric in the semantic segmentation studies is the IoU. We report the IoU as 86.3%. For the AWE test set, the average precision is 93.5 %, and the average recall is 91.8 %. In Table I, high performance of our proposed algorithm on AWE dataset [26] is clearly observed. The average processing time for the images in the test set is 0.037 seconds.

The ear segmentation results are displayed for thirty-six test images having the highest IoU scores in the AWE test set Figure 2. Images are ordered with respect to IoU scores with row-major ordering. Thirty-six test images with the lowest IoU scores in the AWE test set are shown in Figure 3. The IoU scores are also represented as a histogram and are shown in Figure 4. It is observed from Figure 3 and Figure 4, only five images are having an IoU score lower than 0.8. Accuracy values for the images are given in Figure 5 as a histogram. We reported high-performance results for the AWE dataset. We also performed tests on other datasets to document the performance of our ear segmentation solution in different datasets. The results we obtained from these tests, together with the number of images in the datasets, are given in Table II. In the sections followed, we will briefly evaluate the results obtained in these datasets.

2) *UL Test Set*: The UL dataset [27] consists of images on the web collected by researchers and students from the University of Ljubljana. This dataset with 16765 images is quite large and suitable for performance measurement. The IoU value is reported on this dataset as 82.6 % and the precision value as 93.5 % (Table II). Performance values in

this extensive database of images recorded in an uncontrolled environment support the high performance of the proposed ear segmentation method.

3) *UBEAR Test Set*: The UBEAR dataset [28] consists of moving subjects, gray-scale images collected under varying lighting conditions. In addition, the subjects did not pay attention to their pose and ear occlusion. Due to the movement of the subjects and the variable lighting, degradation in sharpness in some images are present. With all these aspects, the UBEAR dataset has the most challenging conditions among the datasets used in this study. The IoU value is reported on this dataset as 55.9 % and the precision value as 85.3 % (Table II). Performance values obtained in this dataset are lower compared to the performance figures in other datasets. The decrease in performance is considered to be due to grey-scale images, moving and non-cooperative subjects, and variable lighting conditions.

4) *Wild-1 Test Set*: Wild-1 Test Set [29] consists of 500 images. The IoU value is reported on this dataset as 77.1 % and the precision value as 80.2. % (Table II). Images of the "In-the-Wild" dataset are collected from Google Images, and the subjects are not cooperative. This independent test set confirms the high values obtained for the performance metrics of the proposed ear segmentation network.

5) *Wild-2 Test Set*: Wild-2 Test Set [29] consists of 105 images. The IoU value is reported on this dataset as 79.7 % and the precision value as 81.8. % (Table II).

D. Ablation Study

In this study, the DeepLabv3+ convolutional neural network's last layer is updated using Tversky Loss Layer. An ablation study has been conducted to demonstrate the positive contribution of this update. For this purpose, the training is performed with the AWE Training Set using the Cross-Entropy Loss Layer as the last layer of the DeepLabv3+ convolutional



Fig. 2. Our ear segmentation results for the test images in the AWE dataset (with highest IoU scores).

neural network. The cross-entropy loss layer is the original loss layer in the DeepLabv3+ structure. Tests are performed on test sets using the convolutional neural network trained using the cross-entropy loss layer. All other hyper-parameters of the CNN are kept constant; therefore, the effect of the loss layer on the overall performance is displayed. The results with cross-entropy loss are given in Table III. These results confirm the positive impact of the Tversky Loss Layer on the overall performance.

TABLE I
COMPARISON OF TEST RESULTS FOR EAR SEGMENTATION ON AWE DATASET

Method	Accuracy	IoU	Precision	Recall	E2
Ped-Ced-Alt[30]	99.2	50.8	62.5	78.5	24.6
Ped-Ced[30]	99.4	55.7	67.7	77.7	22.2
Mask R-CNN[27]	—	79.24	92.04	84.14	—
RefineNet[5]	99.8	84.8	91.7	91.6	7.6
ContexedNet[31]	99.74	81.46	89.07	87.47	—
Our Study	99.8	86.3	93.5	91.8	4.1

TABLE II
OUR TEST RESULTS WITH TVERSKY LOSS FOR EAR SEGMENTATION ON DIFFERENT TEST SETS

Dataset	Images	Accuracy	IoU	Precision	Recall	E2
AWE	250	99.8	86.3	93.5	91.8	4.1
UL	16765	99.8	82.6	93.5	87.6	6.2
UBEAR	4412	99.3	55.9	85.3	61.8	19.2
WILD-1	500	99.8	77.1	80.2	95.2	2.5
WILD-2	105	99.8	79.7	81.8	96.8	1.7

TABLE III
OUR TEST RESULTS WITH CROSS ENTROPY LOSS FOR EAR SEGMENTATION ON DIFFERENT TEST SETS

Dataset	Images	Accuracy	IoU	Precision	Recall	E2
AWE	250	99.76	79.13	93.89	83.42	8.32
UL	16765	99.77	73.99	94.39	77.40	11.32
UBEAR	4412	99.24	45.92	91.07	48.09	25.99
WILD-1	500	99.75	73.62	83.83	85.81	7.16
WILD-2	105	99.76	77.48	83.94	90.97	4.60

E. Limitations of the Study

This study uses challenging databases for ear semantic segmentation, and high-performance results are reported for ear segmentation. The subjects in these databases are not cooperative and look in various directions. Again, in these databases, there are ear images with different scales. With these aspects, the results we obtained for ear-segmentation are enlightening about the performance of the proposed method. The study's main limitation is that the segmentation data used is specific to ear-segmentation only. Therefore, it is impossible to comment on the validity and performance of the proposed method in general semantic segmentation problems. Examining the proposed method in different segmentation problems and data types will help make more general evaluations.

VI. CONCLUSION

This article proposed a semantic segmentation method to segment the ears in the natural images recorded in uncontrolled environments. Our method is supported by selecting



Fig. 3. Our ear segmentation results for the test images in the AWE dataset (with lowest IoU scores).

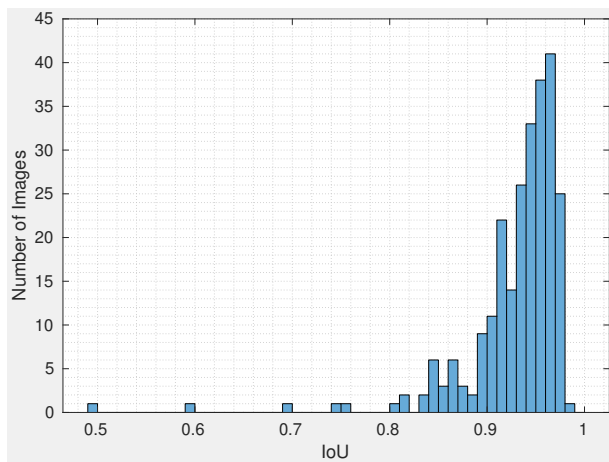


Fig. 4. Histogram for Intersection-over-Union(IoU) for AWE test set [26].

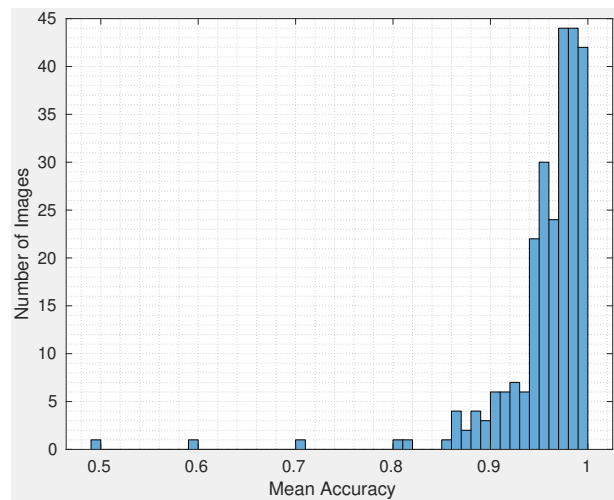


Fig. 5. Histogram for Accuracy for AWE test set [26].

the appropriate model for transfer learning, and the appropriate loss function. AWE training set is used as the training data. DeepLab v3+ [38] encoder-decoder network and pre-trained model ResNet-18 [39] are used in this study to implement an accurate and effective ear semantic segmentation solution. The Tversky loss function [45] has been used in the final stage of our network to overcome the negative effects of the unbalanced distribution of the ear and background classes. The proposed algorithm's performance has been tested, and to the best of our knowledge, we have reported the highest scores related to performance on the AWE dataset [26]. We have also tested

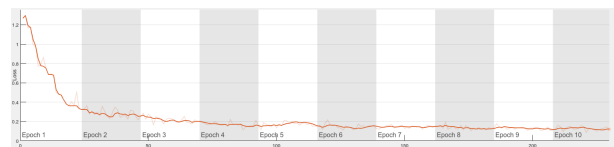


Fig. 6. Loss vs Iterations for Training on AWE Training Set

ear segmentation system performance on different challenging datasets and obtained test results displaying high accuracy and high precision of the proposed ear segmentation system.

One of our authors' recent study [47] is about high-performance ear recognition systems. In our future work, we will use the high performance ear semantic segmentation solution we developed in this publication, together with our knowledge of ear recognition to implement a very high performance, end-to-end, fully automatic ear recognition system. In our future studies, we plan to achieve successful ear recognition performance on the images recorded in natural and uncontrolled environments.

As another future work we plan to explore the few-label semantic segmentation. In this study, all of the images in the training set were completely labeled. Recently, there have been efforts to achieve semantic segmentation with few labels [11]. We also aim to extend our solution within this direction so that it will be possible to complete the training process with less labeled data.

ACKNOWLEDGMENT

Authors would like to thank Dr. Ziga Emeršič, Dr. Vitomir Struc, and their research team for sharing the ear databases, which made this study possible.

REFERENCES

- [1] A. Abaza, A. Ross, C. Hebert, M. A. F. Harrison, and M. S. Nixon, "A survey on ear biometrics," *ACM Computing Surveys*, vol. 45, no. 2, pp. 1–35, Feb. 2013, number: 2 Reporter: ACM Computing Surveys. [Online]. Available: <http://dl.acm.org/citation.cfm?doi=2431211.2431221>
- [2] A. Pflug and C. Busch, "Ear biometrics: a survey of detection, feature extraction and recognition methods," *IET Biometrics*, vol. 1, no. 2, pp. 114–129, Jun. 2012, number: 2 Reporter: IET Biometrics. [Online]. Available: <https://digital-library.theiet.org/content/journals/10.1049/iet-bmt.2011.0003>
- [3] Z. Emeršic, D. Stepec, V. Struc, P. Peer, A. George, A. Ahmad, E. Omar, T. E. Boulter, R. Safdaii, Y. Zhou, S. Zafeiriou, D. Yaman, F. I. Eyiokur, and H. K. Ekenel, "The unconstrained ear recognition challenge," in *2017 IEEE International Joint Conference on Biometrics (IJCB)*, Oct. 2017, pp. 715–724, meeting Name: 2017 IEEE International Joint Conference on Biometrics (IJCB) Reporter: 2017 IEEE International Joint Conference on Biometrics (IJCB) ISSN: 2474-9699.
- [4] Z. Emeršic, A. K. S. V. B. S. Harish, W. Gutfeter, J. N. Khirak, A. Pacut, E. Hansley, M. P. Segundo, S. Sarkar, H. J. Park, G. P. Nam, L.-J. Kim, S. G. Sangodkar, U. Kacar, M. Kirci, L. Yuan, J. Yuan, H. Zhao, F. Lu, J. Mao, X. Zhang, D. Yaman, F. I. Eyiokur, K. B. Özler, H. K. Ekenel, D. P. Chowdhury, S. Bakshi, P. K. Sa, B. Majhi, P. Peer, and V. Struc, "The Unconstrained Ear Recognition Challenge 2019," in *2019 International Conference on Biometrics (ICB)*, 2019, pp. 1–15.
- [5] Z. Emeršic, J. Krizaj, V. Struc, and P. Peer, "Deep Ear Recognition Pipeline," in *Recent Advances in Computer Vision: Theories and Applications*, ser. Studies in Computational Intelligence, M. Hassaballah and K. M. Hosny, Eds. Cham: Springer International Publishing, 2019, pp. 333–362, reporter: Recent Advances in Computer Vision: Theories and Applications. [Online]. Available: https://doi.org/10.1007/978-3-030-03000-1_14
- [6] Z. Zou, Z. Shi, Y. Guo, and J. Ye, "Object Detection in 20 Years: A Survey," *arXiv:1905.05055 [cs]*, May 2019, reporter: arXiv:1905.05055 [cs] arXiv: 1905.05055. [Online]. Available: <http://arxiv.org/abs/1905.05055>
- [7] A. Paszke, A. Chaurasia, S. Kim, and E. Culurciello, "ENet: A Deep Neural Network Architecture for Real-Time Semantic Segmentation," *arXiv:1606.02147 [cs]*, Jun. 2016, reporter: arXiv:1606.02147 [cs] arXiv: 1606.02147. [Online]. Available: <http://arxiv.org/abs/1606.02147>
- [8] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs," *arXiv:1412.7062 [cs]*, Jun. 2016, reporter: arXiv:1412.7062 [cs] arXiv: 1412.7062. [Online]. Available: <http://arxiv.org/abs/1412.7062>
- [9] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes (voc) challenge," *International Journal of Computer Vision*, vol. 88, no. 2, pp. 303–338, Jun. 2010.
- [10] G. Lin, A. Milan, C. Shen, and I. Reid, "RefineNet: Multi-path Refinement Networks for High-Resolution Semantic Segmentation," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Honolulu, HI: IEEE, Jul. 2017, pp. 5168–5177, meeting Name: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Reporter: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). [Online]. Available: <http://ieeexplore.ieee.org/document/8100032/>
- [11] Y. Xian, S. Choudhury, Y. He, B. Schiele, and Z. Akata, "Semantic Projection Network for Zero- and Few-Label Semantic Segmentation," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Long Beach, CA, USA: IEEE, Jun. 2019, pp. 8248–8257, meeting Name: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Reporter: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). [Online]. Available: <https://ieeexplore.ieee.org/document/8953827/>
- [12] Y. Xian, C. H. Lampert, B. Schiele, and Z. Akata, "Zero-Shot Learning - A Comprehensive Evaluation of the Good, the Bad and the Ugly," *arXiv:1707.00600 [cs]*, Aug. 2018, arXiv: 1707.00600. [Online]. Available: <http://arxiv.org/abs/1707.00600>
- [13] H. Caesar, J. Uijlings, and V. Ferrari, "Coco-stuff: Thing and stuff classes in context," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 1209–1218.
- [14] Z. Huang, X. Wang, L. Huang, C. Huang, Y. Wei, and W. Liu, "CCNet: Criss-Cross Attention for Semantic Segmentation," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. Seoul, Korea (South): IEEE, Oct. 2019, pp. 603–612, meeting Name: 2019 IEEE/CVF International Conference on Computer Vision (ICCV) Reporter: 2019 IEEE/CVF International Conference on Computer Vision (ICCV). [Online]. Available: <https://ieeexplore.ieee.org/document/9009011/>
- [15] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, "The cityscapes dataset for semantic urban scene understanding," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 3213–3223.
- [16] B. Zhou, H. Zhao, X. Puig, S. Fidler, A. Barriuso, and A. Torralba, "Scene parsing through ade20k dataset," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 633–641.
- [17] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking Atrous Convolution for Semantic Image Segmentation," *arXiv:1706.05587 [cs]*, Dec. 2017, reporter: arXiv:1706.05587 [cs] arXiv: 1706.05587. [Online]. Available: <http://arxiv.org/abs/1706.05587>
- [18] A. Kirillov, K. He, R. Girshick, C. Rother, and P. Dollar, "Panoptic Segmentation," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Long Beach, CA, USA: IEEE, Jun. 2019, pp. 9396–9405, meeting Name: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Reporter: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). [Online]. Available: <https://ieeexplore.ieee.org/document/8953237/>
- [19] C. Liu, L.-C. Chen, F. Schroff, H. Adam, W. Hua, A. L. Yuille, and L. Fei-Fei, "Auto-DeepLab: Hierarchical Neural Architecture Search for Semantic Image Segmentation," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Long Beach, CA, USA: IEEE, Jun. 2019, pp. 82–92, meeting Name: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Reporter: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). [Online]. Available: <https://ieeexplore.ieee.org/document/8954247/>
- [20] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 2009, pp. 248–255.
- [21] S. Mittal, M. Tatarchenko, and T. Brox, "Semi-Supervised Semantic Segmentation with High- and Low-level Consistency," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–1, 2019, reporter: IEEE Transactions on Pattern Analysis and Machine Intelligence. [Online]. Available: <https://ieeexplore.ieee.org/document/8935407/>
- [22] A. Garcia-Garcia, S. Orts-Escolano, S. Oprea, V. Villena-Martinez, P. Martinez-Gonzalez, and J. Garcia-Rodriguez, "A survey on deep learning techniques for image and video semantic segmentation," *Applied Soft Computing*, vol. 70, pp. 41–65, Sep. 2018, reporter: Applied Soft Computing. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S1568494618302813>

- [23] S. Minaee, Y. Boykov, F. Porikli, A. Plaza, N. Kehtarnavaz, and D. Terzopoulos, "Image Segmentation Using Deep Learning: A Survey," *arXiv:2001.05566 [cs]*, Jan. 2020, reporter: arXiv:2001.05566 [cs] arXiv: 2001.05566. [Online]. Available: <http://arxiv.org/abs/2001.05566>
- [24] F. Lateef and Y. Ruichek, "Survey on semantic segmentation using deep learning techniques," *Neurocomputing*, vol. 338, pp. 321–348, Apr. 2019, reporter: Neurocomputing. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S092523121930181X>
- [25] I. Ulku and E. Akagunduz, "A Survey on Deep Learning-based Architectures for Semantic Segmentation on 2D images," *IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING*, p. 14, 2019, reporter: IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING.
- [26] Z. Emersic, V. Struc, and P. Peer, "Ear recognition: More than a survey," *Neurocomputing*, vol. 255, pp. 26–39, Sep. 2017, reporter: Neurocomputing. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S092523121730543X>
- [27] M. Bizjak, P. Peer, and Z. Emersic, "Mask R-CNN for Ear Detection," in *2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*. Opatija, Croatia: IEEE, May 2019, pp. 1624–1628, meeting Name: 2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO) Reporter: 2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO). [Online]. Available: <https://ieeexplore.ieee.org/document/8756760/>
- [28] R. Raposo, E. Hoyle, A. Peixinho, and H. Proença, "UBEAR: A dataset of ear images captured on-the-move in uncontrolled conditions," *2011 IEEE Workshop on Computational Intelligence in Biometrics and Identity Management (CIBIM)*, pp. 84–90, 2011.
- [29] Y. Zhou and S. Zaferiou, "Deformable Models of Ears in-the-Wild for Alignment and Recognition," in *2017 12th IEEE International Conference on Automatic Face Gesture Recognition (FG 2017)*, 2017, pp. 626–633.
- [30] Z. Emersic, L. L. Gabriel, V. Struc, and P. Peer, "Convolutional encoder-decoder networks for pixel-wise ear detection and segmentation," *IET Biometrics*, vol. 7, no. 3, pp. 175–184, May 2018, number: 3 Reporter: IET Biometrics. [Online]. Available: <https://digital-library.theiet.org/content/journals/10.1049/iet-bmt.2017.0240>
- [31] Z. Emersic, D. Susanj, B. Meden, P. Peer, and V. Struc, "Contextednet: Context-aware ear detection in unconstrained settings," *IEEE Access*, vol. 9, pp. 145 175–145 190, 2021.
- [32] C. Cintas, C. Delrieux, P. Navarro, M. Quinto-Sánchez, B. Pazos, and R. Gonzalez-José, "Automatic Ear Detection and Segmentation over Partially Occluded Profile Face Images," *Journal of Computer Science and Technology*, vol. 19, no. 01, p. e08, Apr. 2019, number: 01 Reporter: Journal of Computer Science and Technology. [Online]. Available: <http://journal.info.unlp.edu.ar/JCST/article/view/1097>
- [33] X. Zhang, L. Yuan, and J. Huang, "Physiological Curves Extraction of Human Ear Based on Improved YOLACT," in *2020 IEEE 2nd International Conference on Civil Aviation Safety and Information Technology (ICCASIT)*, 2020, pp. 390–394.
- [34] I. I. Ganapathi, S. Prakash, I. R. Dave, and S. Bakshi, "Unconstrained ear detection using ensemble-based convolutional neural network model: Unconstrained ear detection using ensemble-based convolutional neural network model," *Concurrency and Computation: Practice and Experience*, p. e5197, Feb. 2019, reporter: Concurrency and Computation: Practice and Experience. [Online]. Available: <http://doi.wiley.com/10.1002/cpe.5197>
- [35] Y. Zhang and Z. Mu, "Ear Detection under Uncontrolled Conditions with Multiple Scale Faster Region-Based Convolutional Neural Networks," *Symmetry*, vol. 9, no. 4, p. 53, Apr. 2017, number: 4 Reporter: Symmetry. [Online]. Available: <http://www.mdpi.com/2073-8994/9/4/53>
- [36] A. Kamboj, R. Rani, A. Nigam, and R. Jha, "CED-Net: context-aware ear detection network for unconstrained images," *Pattern Analysis and Applications*, 2020.
- [37] W. Raveane, P. L. Galdámez, and M. A. González Arrieta, "Ear Detection and Localization with Convolutional Neural Networks in Natural Images and Videos," *Processes*, vol. 7, no. 7, p. 457, Jul. 2019, number: 7 Reporter: Processes. [Online]. Available: <https://www.mdpi.com/2227-9717/7/7/457>
- [38] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation," in *Computer Vision – ECCV 2018*, V. Ferrari, M. Hebert, C. Sminchisescu, and Y. Weiss, Eds. Cham: Springer International Publishing, 2018, vol. 11211, pp. 833–851, reporter: Computer Vision – ECCV 2018. [Online]. Available: http://link.springer.com/10.1007/978-3-030-01234-2_49
- [39] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.
- [40] N. Abraham and N. M. Khan, "A novel focal tversky loss function with improved attention u-net for lesion segmentation," in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, 2019, pp. 683–687.
- [41] S. Jadon, "A survey of loss functions for semantic segmentation," in *2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*. IEEE, 2020, pp. 1–7.
- [42] W. Yuan and W. Xu, "Neighborloss: a loss function considering spatial correlation for semantic segmentation of remote sensing image," *IEEE Access*, vol. 9, pp. 75 641–75 649, 2021.
- [43] D. Duque-Arias, S. Velasco-Forero, J.-E. Deschaut, F. Goulette, A. Serna, E. Decencièrre, and B. Marcotegui, "On power jaccard losses for semantic segmentation," in *VISAPP 2021: 16th International Conference on Computer Vision Theory and Applications*, 2021.
- [44] H. Harkat, J. M. P. Nascimento, A. Bernardino, and H. F. Thariq Ahmed, "Assessing the impact of the loss function and encoder architecture for fire aerial images segmentation using deeplabv3+," *Remote Sensing*, vol. 14, no. 9, 2022. [Online]. Available: <https://www.mdpi.com/2072-4292/14/9/2023>
- [45] S. S. M. Salehi, D. Erdogmus, and A. Gholipour, "Tversky loss function for image segmentation using 3D fully convolutional deep networks," *arXiv:1706.05721 [cs]*, Jun. 2017, reporter: arXiv:1706.05721 [cs] arXiv: 1706.05721. [Online]. Available: <http://arxiv.org/abs/1706.05721>
- [46] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," *arXiv:1412.6980 [cs]*, Jan. 2017, arXiv: 1412.6980. [Online]. Available: <http://arxiv.org/abs/1412.6980>
- [47] U. Kacar and M. Kirci, "ScoreNet: Deep cascade score level fusion for unconstrained ear recognition," *IET Biometrics*, vol. 8, no. 2, pp. 109–120, 2018, number: 2 Publisher: IET.

BIOGRAPHIES

Tolga Inan received the B.Sc., M.Sc., and Ph.D. degrees in electrical and electronics engineering from the Middle East Technical University, Ankara, Turkey, in 2000, 2003, and 2011, respectively. He is currently with the Department of Electrical-Electronics Engineering, Cankaya University, Ankara. His areas of research include machine learning, biometric recognition, computer vision, and power quality analysis.



Umit Kacar completed his Master's degree in Electronics Engineering from Istanbul Technical University in 2013 and his PhD in Electronics Engineering from Istanbul Technical University in 2019. He is currently working as a Senior Machine Learning Engineer in a private company. His research interests are biometrics, fusion, ensemble, artificial intelligence, machine learning and computer vision.



Comparison of Deep Learning and Traditional Machine Learning Classification Performance in a SSVEP Based Brain Computer Interface

Zafer İçsan


Abstract—Brain-computer interfaces (BCIs) offer a very high potential to help those who cannot use their organs properly. In the literature, many electroencephalogram (EEG) based BCIs exist. Steady state visual evoked potential (SSVEP) based BCIs provide relatively higher accuracy values which make them very popular in BCI research. Recently, deep learning (DL) based methods have been used in EEG classification problems and they had superior performance over traditional machine learning (ML) methods, which require a feature extraction step. This study aimed at comparing the performance of DL and traditional ML-based classification performance in terms of stimuli duration, number of channels, and number of trials in an SSVEP based BCI experiment. In the traditional approach, canonical correlation analysis method was used for the feature extraction and then three well-known classifiers were used for classification. In DL-based classification, spatio-spectral decomposition (SSD) method was integrated as a preprocessing step to extract oscillatory signals in the frequency band of interest with a convolutional neural network structure. Obtained offline classification results show that proposed DL approach could generate better accuracy values than traditional ML-based methods for short time segments (< 1 s). Besides, use of SSD as a preprocessing step increased the accuracy of DL classification. Superior performance of proposed SSD based DL approach over the traditional ML methods in short trials shows the feasibility of this approach in future BCI designs. A similar approach can be used in other fields where there are oscillatory activities in the recorded signals.

Index Terms—Brain computer interface, classification, convolutional neural network, deep learning, spatio-spectral decomposition, steady state visual evoked potential.

I. INTRODUCTION

BRAIN COMPUTER INTERFACES (BCIs) have the potential to improve the quality of disabled people's lives by providing an additional communication channel [1].

ZAFER İÇSAN, is with Department of Electrical and Electronics Engineering, Bahçeşehir University, Istanbul, Turkey, (e-mail: zaferiscan@yahoo.com).

 <https://orcid.org/0000-0001-9832-6591>

Manuscript received March 15, 2022; accepted July 18, 2022.
DOI: [10.17694/bajece.1088353](https://doi.org/10.17694/bajece.1088353)

Therefore, people who lost their ability to use their organs can still benefit from their neuronal signals in the brain to execute the desired action.

Steady state visual evoked potential (SSVEP) is one of the most popular evoked potentials due to its robustness and high signal-to-noise ratio (SNR) [2]. SSVEP based BCIs offer high information transfer rate (ITR) - a common measure to calculate the performance of a BCI [3].

In the SSVEP based BCIs, one common feature extraction method is the canonical correlation analysis (CCA) that has been shown to be superior to the other existing methods like the power spectral density [4] and minimum energy combination [5]. In the problems where two sets of data are expected to have some correlations CCA method can be preferred.

Deep learning (DL) is a relatively new approach in neuroscience. However, it has already shown its superiority over traditional feature based classification algorithms in several electroencephalogram (EEG) classification problems [6]–[8]. Convolutional neural network (CNN) is a class of deep neural network (DNN), which is commonly used in DL based approaches. Kwak et al. [9] proposed a CNN for classifying SSVEP under a static and an ambulatory environment. The proposed CNN generated better classification rates than the standard neural network, CCA classifier, CCA combined with k-nearest neighbor (KNN), and a multivariate synchronization index. Ravi et al. [10] compared user-dependent and user-independent training of CNN for SSVEP based BCI classification in two datasets using magnitude and complex spectrum features. They used the CCA method as the baseline method and applied task-related component analysis and filter-bank CCA. Their results suggested that user-independent complex CNN method provides a good trade-off between performance and training cost. Ikeda and Washizawa [11] proposed a complex valued CNN to overcome the limitation of available frequencies in the SSVEP based BCIs. Their method outperformed the CCA based classification methods. Guney et al. [12] proposed a new CNN based DNN and reached an ITR of 265.23 bits/min and 196.59 bits/min on two different SSVEP datasets for a very short (0.4 s) stimuli duration. By the time of the

publication, these ITR rates were reported to be the highest performance results obtained on these datasets. Recently, Zhao et al. [13] verified the feasibility of using CNN for augmented reality based SSVEP classification.

Although many studies run the DL algorithms on the raw data, DL based classification algorithms were shown to be sensitive to the preprocessing [14]. In EEG classification problems, SNR is a critical value. It is important to obtain the oscillatory sources reliably for an accurate classification. Spatio-spectral decomposition (SSD) is a method to find linear filters that maximizes the power in the frequency band of studied neuronal oscillations while minimizing the power at the neighboring frequencies [15]. In this study, SSD method was used to reliably extract oscillatory signals in the frequency band of interest in the DL-based classification.

Channel selection is an important step in finding the relevant EEG channels that are used in feature extraction process [16]. Therefore, this step is directly affecting the accuracy of the classification. Several channel selection methods have been proposed in the literature. For example, common spatial pattern (CSP) is one of the popular methods in motor imagery based BCIs [17]. Arvaneh et al. [18] proposed a decision tree based channel selection method in a motor imagery based BCI dataset and showed that this method outperformed the other methods based on Fisher Criterion, Mutual Information, Support Vector Machine and CSP coefficients especially when there are few channels. They later proposed a sparse CSP algorithm for channel selection in two motor imagery based BCI datasets and showed its superiority over the regularized CSP in addition to the other methods [19]. Recently Feng et al. [20] proposed a multi frequency CSP-Rank method for channel selection in a motor imagery-based BCI dataset and showed that it improved the classification accuracy compared with the CSP-rank channel selection method.

In SSVEP based BCIs there are other approaches for channel selection. For example, Zhang et al. proposed spatial temporal correlation method to select the best channels and showed that it increased the average accuracy [21]. In another SSVEP based BCI, Meng et al. used sequential floating forward selection, discrete particle swarm optimization and F-score to select the optimal EEG channels and showed that this generates higher classification accuracies than using traditional O1, O2 and Oz channels [22]. Webster et al. used an unsupervised channel selection method in an SSVEP based BCI with the majority voting of classification outputs obtained from each subset of channels and showed that this method could be better from a priori channel selection method when CCA is used in feature extraction [23]. Carvalho et al. compared many methods for feature extraction and classification in an SSVEP-based two-class BCI, and noticed that the best feature selection method was incremental wrapper method that performed feature selection using the performance of the classifier as in the case of genetic algorithms (GAs) [24]. In fact, GAs have been used in many EEG classification problems. Yang et al. applied the genetic

neural mathematic method to two EEG channel selection and classification problems, and showed that it improved the generalization ability [25]. In a slow cortical potential based BCI task, Schroder et al. used a GA for feature selection and showed an increase in classification accuracy [26]. Peterson et al. also used GA for feature selection in a two-class visual BCI task and showed that this method found better features than using all features or random feature subsets [27].

Stimuli duration is another important factor for SSVEP based BCIs which directly influences the ITR. Although the accuracy of the SSVEP increases with the increasing duration of the stimuli, ITR starts to decay after some time [28].

In the study, DL and traditional ML-based classification accuracies were compared in an SSVEP based BCI experiment in terms of stimuli length, number of channels, and number of trials. Furthermore, SSD method was incorporated in DL based approach as a preprocessing step to increase the classification accuracy.

II. MATERIALS AND METHODS

A. Participants

Seven right-handed healthy subjects (four males) between 17 and 24 years with an average age of 21 years voluntarily participated in the study and gave their written informed consent. Ethical approval was obtained from the Local Ethics Committee of National Research University Higher School of Economics, Moscow.

B. Experiment setup

The experiments were performed by the author in a shielded dark room at the EEG laboratory of the Center for Cognition and Decision Making of National Research University Higher School of Economics, Moscow. The setup used in the study was same as a previous study [29]. During the experiments, subjects were sitting on a chair and looking at an LED monitor in front of them. EEG were recorded simultaneously with the electrodes attached to the scalp while the subjects were following the visual stimuli. Electrode cables were attached to the EEG amplifier and the amplifier was connected to the recording computer. Visual stimuli presented to the subjects were four circles on the LED monitor (Resolution: 1920 × 1080 pixels, refresh rate: 60 Hz). Each of the circles had a different flickering frequency: 5.45 Hz for upper circle, 8.57 Hz for lower circle, 12 Hz for the right circle, and 15 Hz for the left circle. These frequencies depend on the refresh rate of the monitor and were selected in a way that the first and second harmonics of the stimuli will not coincide. In Fig. 1, the locations of the circles on the LED monitor are presented.

The visual stimuli interface was prepared on Matlab software (<https://www.mathworks.com/products/matlab.htm>) with the help of Psychophysics Toolbox (<http://psychtoolbox.org/>).

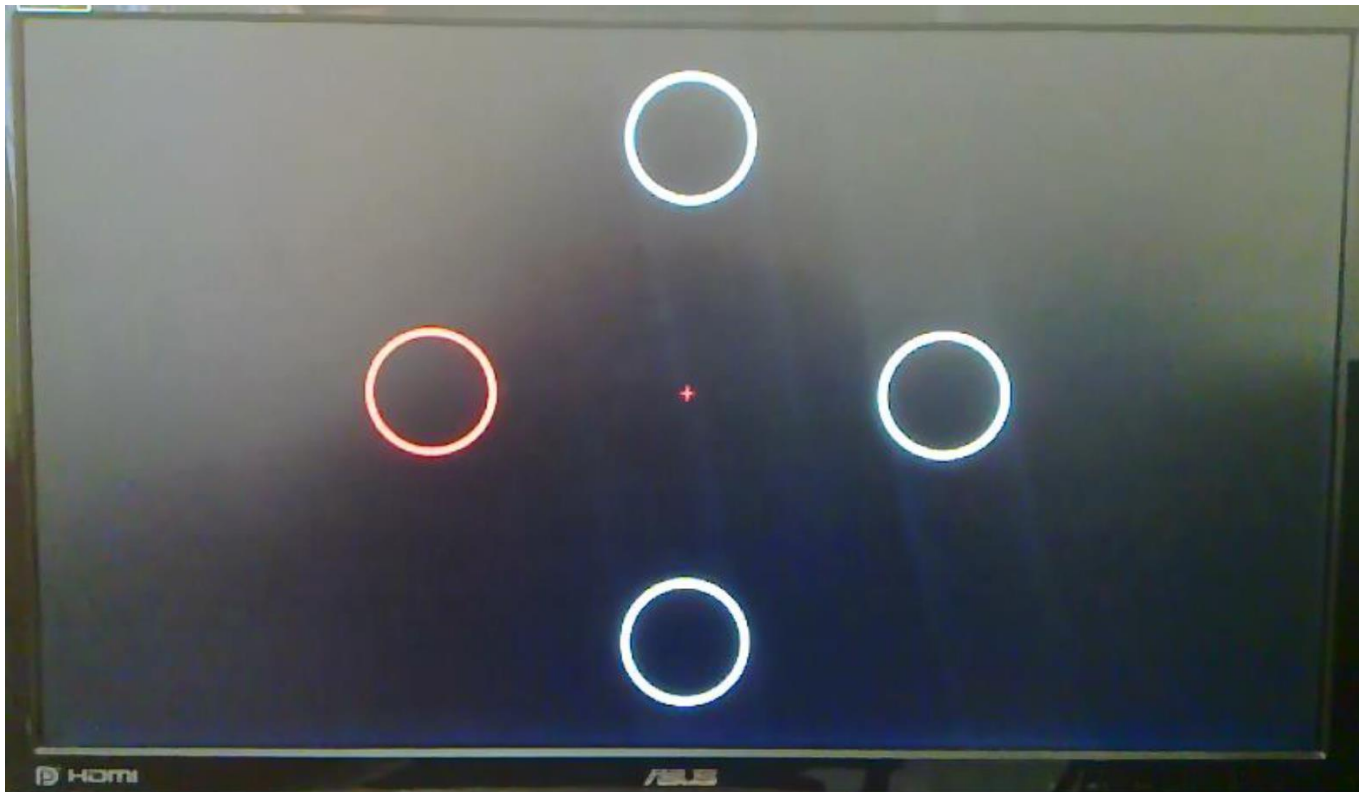


Fig.1. Locations of flickering circles: Up: 5.45 Hz, Down: 8.57 Hz, Right: 12 Hz, and Left: 15 Hz

Further details about the setup can be seen in [29].

Subjects focused their eyes on one of the four flickering circles that are specified with a visual cue (red frame) for three seconds. The order of the cues was permuted randomly. During one trial, subjects focused on each circle once.

There were 30 trials. Therefore, each subject had 120 (30 trials × 4 classes) SSVEP responses. During the experiment,

there was no feedback for the participants about the classification result (i.e. decision of the classifier). In Fig. 2, experiment blocks were presented with their timing. As it can be seen in Fig. 2, there are six different blocks in the experiment.

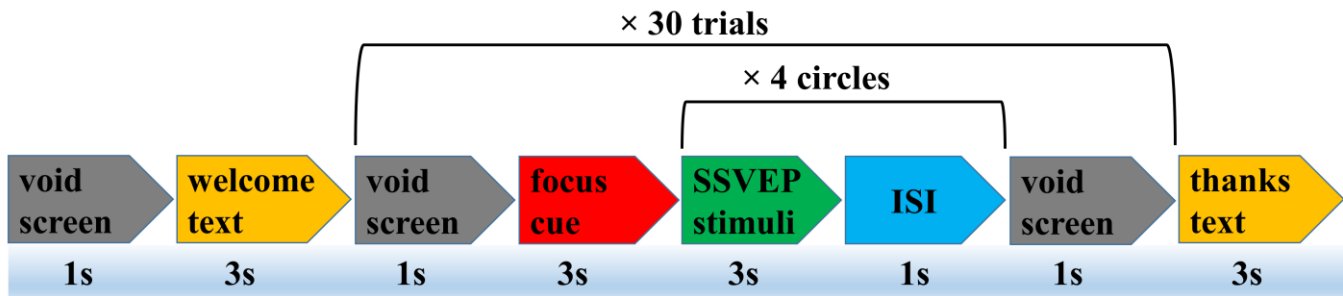


Fig.2. Blocks and timing of the experiment.

During the **void screen** block, the screen is blank (i.e. totally black). In the **welcome text** block, there is a welcome message for the participants. During the **focus cue** block, a message is displayed for the subject to focus on the presented circle. The **SSVEP stimuli** block is where four circles are displayed on the screen simultaneously with their corresponding flickering frequencies. Subjects fixate their eyes and give their attention to the specified flickering circle. After this block, there is an inter-stimulus interval (**ISI**) block which is another void screen between the consecutive **SSVEP stimuli** blocks. The **SSVEP stimuli** and the **ISI** blocks are

repeated four times (once for each circle). After all circles are presented once in a trial, this step is repeated 30 times. Hence, there are 120 SSVEP responses (each response lasts 3 s) in total. Finally, experiment thanks the participant in the **thanks text** block.

C. Data recording

Data were recorded from 60 EEG channels using active electrodes with ActiChamp system and a Python based software (PyCorder) of Brain Products, Germany. In Fig. 3, electrode locations in the experiments were presented.

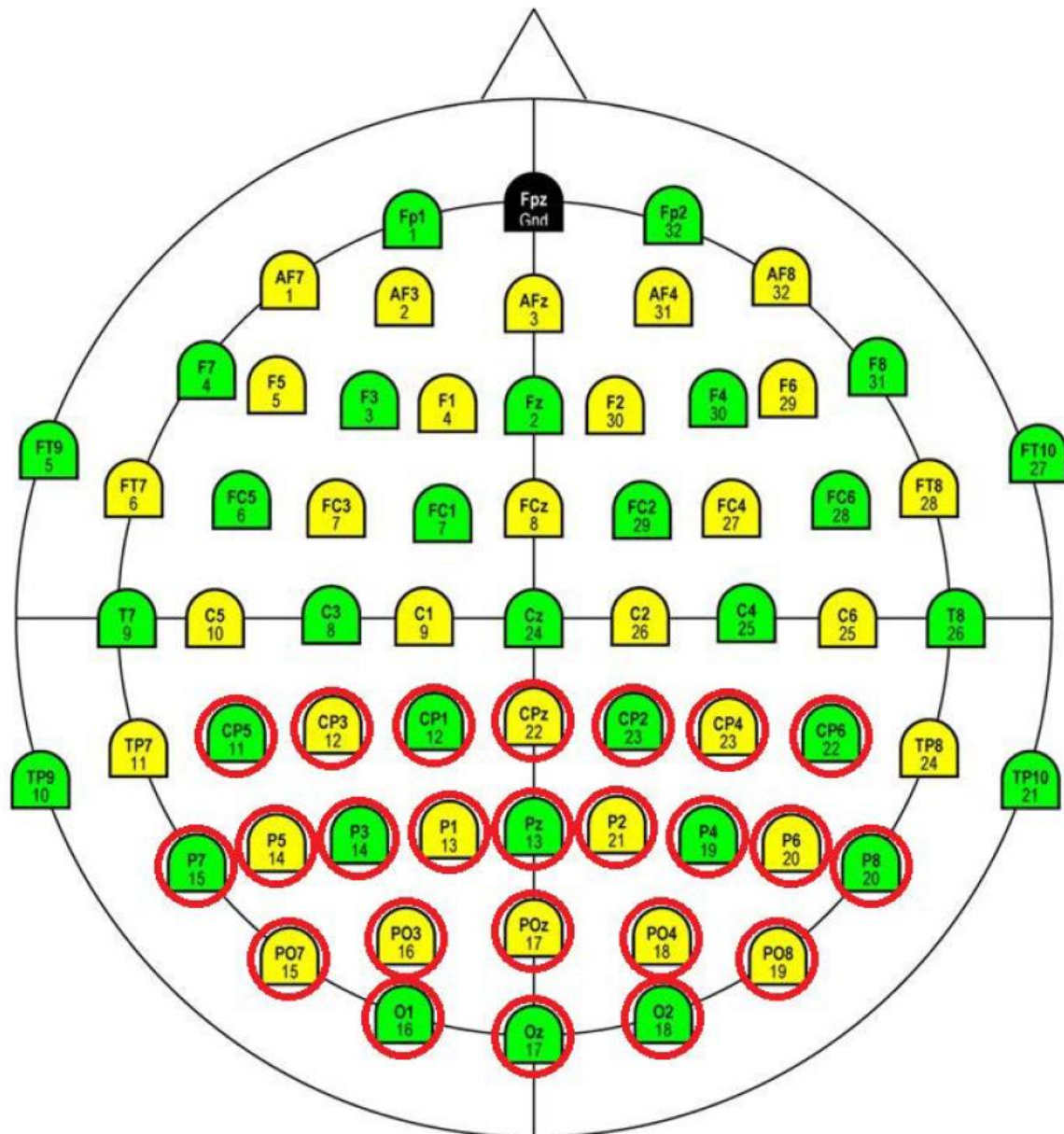


Fig. 3. Standard electrode locations (Addendum actiCHamp Version 002 05/2012) for 64 channel EEG recording system and selected 24 channels (in red circles) for smaller channel set.

Three electrodes (TP9, TP10, and FT10) were used as electrooculogram (EOG) electrodes and FT9 electrode was used for reference. Placement of three EOG electrodes were according to [30]. Reference electrode was attached to the left mastoid. Other 60 electrodes were placed on their standard locations. Impedance values between scalp and electrodes were less than 20 k Ω and sampling frequency was set to 1 kHz. Data were analyzed offline.

D. Classification

Data were resampled to 100 Hz and split into a training (2/3 of the samples) and a test set (1/3 of the samples) similar to [31]. A defected electrode channel (F5) was removed from the analyses. The classifications were done using all (59) vs. selected (24) channels. These channels were selected from the scalp areas where the SSVEP responses are known to be

stronger (See Fig. 3). Trial numbers per class were kept constant (20 train / 10 test) or increased (up to 240 train / 120 test) for shorter stimuli lengths. Before starting feature extraction process, trends in the segmented EEG were removed. A traditional machine learning approach and a deep learning approach were used for classification. The details of these approaches are given below.

E. Traditional machine learning approach

Traditional machine learning (ML) based approach needs a feature extraction step to obtain critical features and then these features are classified with classification algorithms. In Fig. 4, traditional ML approach used in the study was presented with the CCA as the feature extracting method.

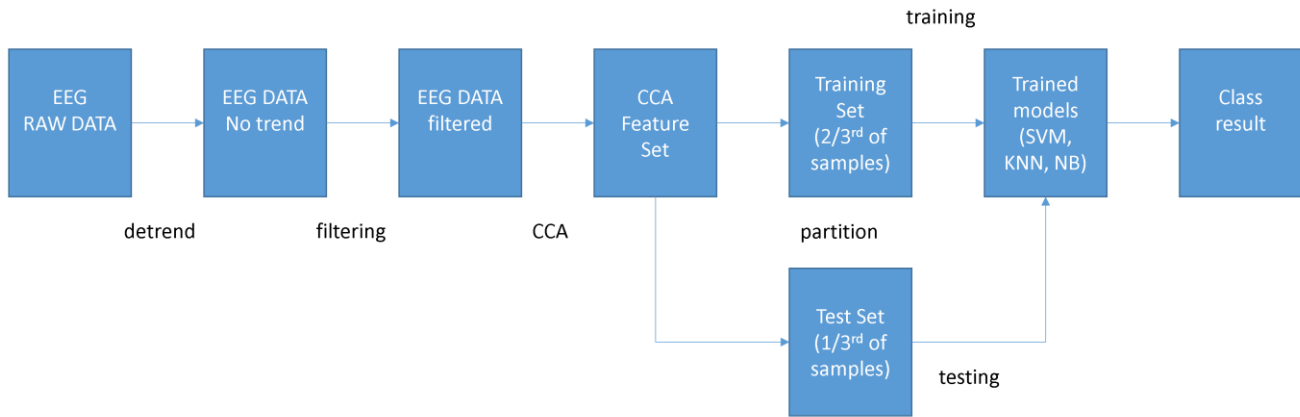


Fig.4. Blocks of traditional machine learning approach using SVM, KNN, and NB classifiers with features obtained from canonical correlation analysis.

First of all, trends in the EEG data were removed. Using a band-pass filter (0.53 - 40 Hz) very slow signal variations and high frequency artifacts including power line noise (50 Hz) were eliminated. Other artifact removal methods were not used as SSVEPs are not very sensitive to eye movements or other physiological artifacts [32].

Canonical correlation analysis (CCA) based feature extraction method was used for all EEG channels. These features were extracted from the segmented EEG records corresponding to the **SSVEP stimuli** blocks (3 seconds each).

CCA method tries to maximize the correlation between linear combinations (canonical variables) of two given data sets [33]. In order to use the method, one set of the data was taken from the recorded EEG segment, and the other set of the data was artificially generated using sine and cosine functions at the stimulation frequencies and their second harmonics. In the end, 16 canonical correlation features were generated. Details on how to implement the method were given in a former study [34].

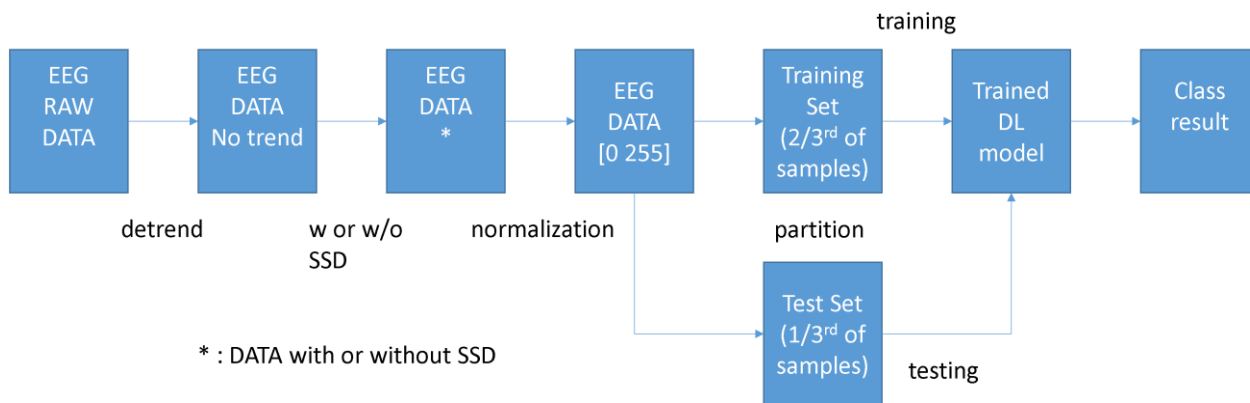
Naïve Bayes, K-Nearest Neighbor (KNN), and Decision Tree classifiers were used to determine the class of the feature vectors obtained from SSVEP responses. These

classifiers have been used in literature for SSVEP-based BCIs [35]–[37].

Naïve Bayes is a probabilistic classifier that assigns the new sample to the most likely class. In this study, Naïve Bayes classifier was used with kernel density estimation for all features. KNN is a non-parametric instance based classifier where the new sample is assigned by calculating the distance from the existing K neighboring samples. In the study, $K = 5$ and the distance metric was euclidean. Decision Tree classifiers build a tree structure by checking the values of features at each node and generating new branches until arriving a leaf corresponding to the class label. Here, a decision tree with binary split was used for classification. All these three classifiers were provided in Matlab R2020b.

F. Deep learning approach

In Fig. 5, proposed DL based approach was presented. Here the DL block is based on a convolutional neural network (CNN) architecture.



* : DATA with or without SSD

Fig.5. Blocks of deep learning approach with or without SSD

After detrending the data, there were two cases: In the first one, SSD method was applied to extract the oscillatory signals due to the visual stimuli. In Table 1, the cut-off frequencies of

the extracted oscillations, flanking intervals, and the band-stop intervals are given for the presented stimuli frequencies.

TABLE I
THE SELECTED CUT-OFF FREQUENCIES OF THE EXTRACTED OSCILLATIONS, FLANKING INTERVALS, AND THE BAND-STOP INTERVALS (IN HZ) FOR THE STIMULI FREQUENCIES IN THE SSD METHOD

Stimulus frequency (Hz)	Frequency of interest	Flanking interval	Band-stop interval
15	[14 - 16]	[12 - 18]	[13 - 17]
12	[11 - 13]	[9 - 15]	[10 - 14]
8.57	[7.5 - 9.5]	[5.5 - 11.5]	[6.5 - 10.5]
5.45	[4.5 - 6.5]	[2.5 - 8.5]	[3.5 - 7.5]

As there are four stimuli frequencies, SSD method generated four different data from the original data. Therefore, the method increased the data size by four. In the second case, SSD method was not applied.

After this block, the data were normalized between 0 and 255 in order to save them as 8 bit images. In the proposed approach, the EEG Data were considered as gray level images (channels × samples).

CNN based DL architecture can be summarized as follows:

There is an input layer with the same size as the images (i.e. normalized EEG data). Then there is a convolution layer with a filter size 4, and a filter number 20. This layer is convolving the input by moving the filters along the input vertically and horizontally and computing the dot product of the weights and the input, and then adding a bias term. Next, there is a batch normalization layer that normalizes the activations and gradients propagating through the neural network, making network training an easier optimization problem. This layer is followed by a rectified linear unit which basically sets each negative element of its input to zero. This layer is followed by a pooling layer with a size (2 x 2) and a stride value of 1. After this layer, there is a fully connected layer with an output size 4. This layer is followed by a softmax layer that transformed the values into probability values. Last layer is the

classification layer that calculates the cross entropy loss. In Fig. 6, all layers of the proposed CNN structure was given.

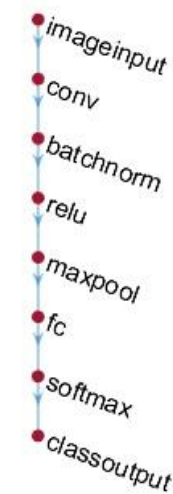


Fig.6. Layers of the proposed CNN

In the training of CNN, stochastic gradient descent with momentum (SGDM) optimizer was used. Initial learning rate was set to 0.0001. Maximum number of epochs was 50. Factor for L2 regularization was 0.0001.

III. RESULTS

Accuracy of the traditional ML and proposed DL approaches with (w) and without (w/o) SSD were compared in terms of number of channels (i.e. all (59) vs selected 24) and stimuli length for constant (20 Train / 10 Test trials per class) and increasing (240 Train / 120 Test trials per class) number of trials in Tables 2 and 3, respectively. These results were visualized in Fig. 7.

TABLE II
CLASSIFICATION ACCURACIES (%) FOR TRADITIONAL ML VS. DL APPROACHES (CONSTANT NUMBER OF TRIALS)

# of channels	Train / Test trials per class	Length (s)	Traditional ML algorithm			Deep learning algorithm	
			Decision Tree	Naïve Bayes	KNN	DL w/o SSD	DL with SSD
24	20 / 10	0.25	20.36	26.07	23.93	29.64	35.71
		0.5	30.36	36.43	31.43	61.43	62.50
		0.75	59.64	68.57	60.00	68.93	80.00
		1	72.14	81.79	80.36	72.14	84.64
		1.5	88.21	92.86	90.36	76.79	86.07
		2	92.86	94.64	95.36	77.86	89.29
		3	93.93	97.50	97.14	82.14	91.07
59	20 / 10	0.25	31.43	27.86	26.43	31.43	27.14
		0.5	25.71	22.50	23.93	54.64	53.21
		0.75	33.93	36.43	27.86	65.71	66.43
		1	55.00	56.07	47.50	64.29	76.79
		1.5	77.86	80.00	81.07	72.14	79.29
		2	88.57	91.07	95.00	72.14	82.50
		3	94.64	95.71	96.07	75.36	83.93

TABLE III
CLASSIFICATION ACCURACIES (%) FOR TRADITIONAL ML VS. DL APPROACHES (VARYING NUMBER OF TRIALS)

# of channels	Train / Test trials per class	Length (s)	Traditional ML algorithm			Deep learning algorithm	
			Decision Tree	Naïve Bayes	KNN	DL w/o SSD	DL with SSD
24	240 / 120	0.25	25.42	23.99	24.64	48.24	59.14
	120 / 60	0.5	59.11	68.04	53.51	53.10	74.40
	80 / 40	0.75	76.07	83.30	77.95	61.61	80.27
	60 / 30	1	82.26	88.57	87.62	65.83	81.79
	40 / 20	1.5	91.07	93.57	92.32	64.46	83.04
	20 / 10	2	92.86	94.64	95.36	77.86	89.29
	20 / 10	3	93.93	97.50	97.14	82.14	91.07
59	240 / 120	0.25	25.00	25.09	24.61	46.10	55.15
	120 / 60	0.5	25.36	25.06	25.12	49.05	66.25
	80 / 40	0.75	42.14	50.54	37.14	55.45	74.91
	60 / 30	1	65.60	75.00	61.79	60.71	71.90
	40 / 20	1.5	80.71	86.43	86.07	64.29	74.82
	20 / 10	2	88.57	91.07	95.00	72.14	82.50
	20 / 10	3	94.64	95.71	96.07	75.36	83.93

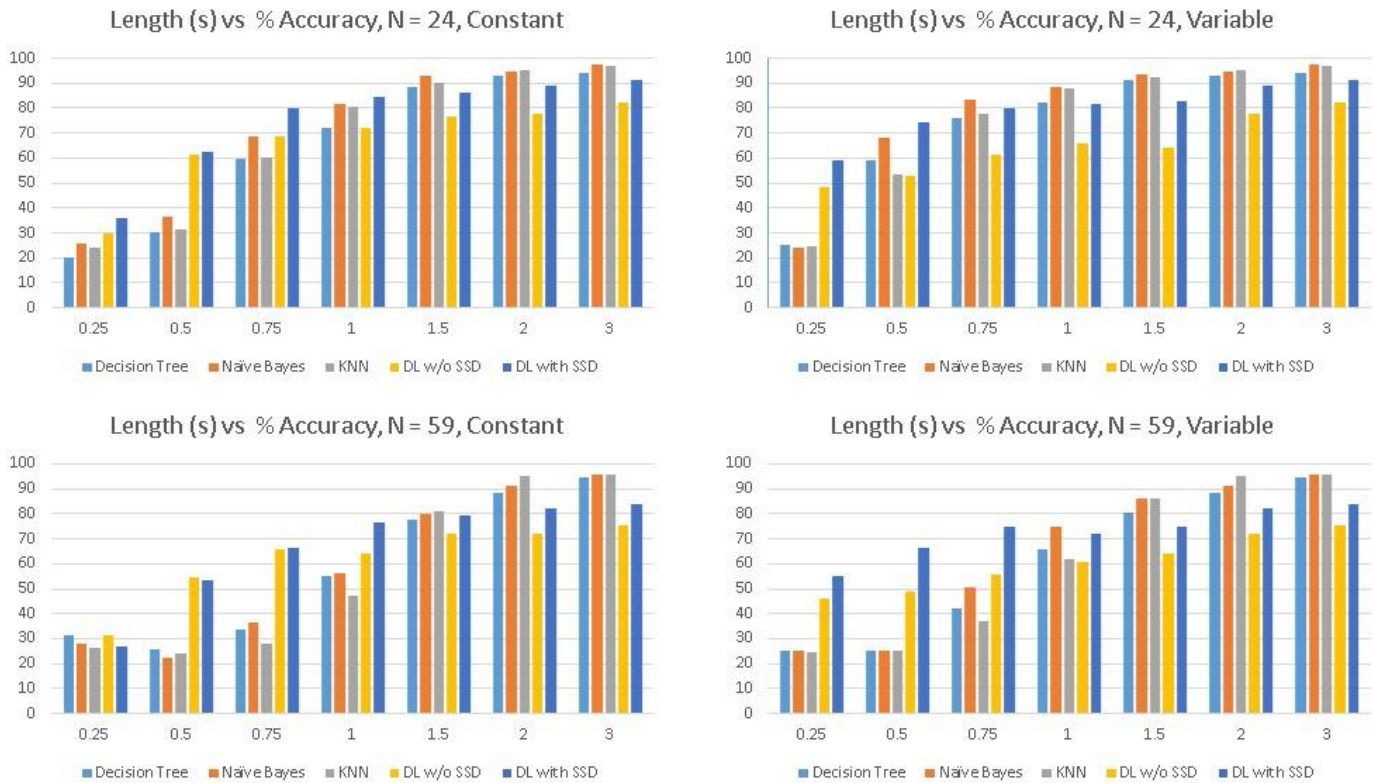


Fig.7. Stimuli length vs. (%) accuracies for constant (left) and variable (right) number of trials using 24 (top) and 59 (bottom) channels.

IV. DISCUSSION

In Tables 2 and 3, very high (>90%) accuracy values were obtained by traditional machine learning algorithms when the stimuli length were long enough >1.5 s. These results were expected as SSVEP based BCIs that use CCA features can generate very high performance in overt attention [38]. In fact, traditional ML algorithms generated higher accuracies than the proposed DL approach for stimuli length >1 s. The advantage of the DL approach over the traditional ML algorithms was

pronounced for shorter stimuli length. For the shortest stimuli length (i.e. 0.25 s), traditional methods performed around chance level (25%) whereas the proposed DL method reached to 59.14% for the varying number of trials using 24 channels. The confusion matrix related to this case was given in Fig. 8. The results emphasize the advantage of DL methods when including more trials in the classification.

Accuracy: 59.14%

	1	2	3	4
1	57.9% 7.185714e+01	8.1% 10	15.3% 1.785714e+01	17.4% 2.028571e+01
2	8.8% 1.085714e+01	71.2% 8.742857e+01	7.3% 8.571429e+00	11.3% 1.314286e+01
3	16.9% 21	7.2% 8.857143e+00	56.4% 6.585714e+01	20.9% 2.428571e+01
4	16.4% 2.028571e+01	13.5% 1.657143e+01	20.9% 2.442857e+01	50.4% 5.871429e+01
	1	2	3	4

Target Class

Fig. 8. Confusion matrix for DL with SSD, N = 24, length 0.25 s, varying number of trials

Accuracies for the DL approach that used the SSD as a preprocessing step were in general higher than those that belong to the DL without SSD. There were only two exceptions of this statement for the constant trial case, for a stimulus length ≤ 0.5 s, and 59 channels. This proves that the SSD is a proper preprocessing step in DL approaches for SSVEP based BCIs.

Another result of the study is that selected 24 channels gave higher accuracies than all channels for both ML and DL based approaches. This is a promising result, as using higher number of channels is not desired for BCIs due to practical reasons.

One limitation of the study is that there were few participants in the study. In the future, this approach should be tested on a bigger dataset to validate the generalization capability of the presented results. This will also help in evaluation of the results using statistical analyses. Moreover, by integrating an optimization method for channel selection (e.g. genetic algorithm), upper limit of the classification accuracies can be enhanced. Besides, DL performance presented here could be suboptimal due to the selection of the parameters. Determination of the optimal parameter values can boost the classification accuracies.

This study shows a systematic comparison of traditional ML approaches and DL approaches in an SSVEP based BCI experiment in terms of stimuli length, number of channels, and number of trials. This is the first study that incorporates SSD with a CNN based classification. The results of this study pave the way for combining DL with SSD in different fields where there are expected oscillatory activities in the recorded signal.

ACKNOWLEDGMENT

I would like to thank Elena Barulina and Vadim Nikulin for their help in the study which was partly funded by the Russian Academic Excellence Project '5-100'.

REFERENCES

- [1] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain-computer interfaces for communication and control," *Clin. Neurophysiol.*, vol. 113, 2002.
- [2] M. Sengelmann, A. K. Engel, and A. Maye, "Maximizing Information Transfer in SSVEP-Based Brain-Computer Interfaces," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 2, pp. 381–394, 2017.
- [3] X. Chen, Y. Wang, M. Nakanishi, X. Gao, T.-P. Jung, and S. Gao, "High-speed spelling with a noninvasive brain-computer interface," *Proc. Natl. Acad. Sci.*, vol. 112, no. 44, pp. E6058–E6067, Nov. 2015.
- [4] Z. Yan, X. Gao, G. Bin, B. Hong, and S. Gao, "A half-field stimulation pattern for SSVEP-based brain-computer interface," *Conf. Proc. ... Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Conf.*, vol. 2009, pp. 6461–6464, 2009.
- [5] W. Nan, C. M. Wong, B. Wang, F. Wan, P. U. Mak, P. I. Mak, and M. I. Vai, "A comparison of minimum energy combination and canonical correlation analysis for SSVEP detection," *2011 5th International IEEE/EMBS Conference on Neural Engineering*. pp. 469–472, 2011.
- [6] M. Simões, D. Borra, E. Santamaría-Vázquez, GBT-UPM, M. Bittencourt-Villalpando, D. Krzemiński, A. Miladinović, Neural_Engineering_Group, T. Schmid, H. Zhao, C. Amaral, B. Direito, J. Henriques, P. Carvalho, and M. Castelo-Branco, "BCIAUT-P300: A Multi-Session and Multi-Subject Benchmark Dataset on Autism for P300-Based Brain-Computer-Interfaces," *Frontiers in Neuroscience*, vol. 14, 2020.
- [7] A. Barachant and R. Cycon, *Pushing the limits of BCI accuracy: Winning solution of the Grasp & Lift EEG challenge*. 2016.
- [8] B. Zang, Y. Lin, Z. Liu, and X. Gao, "A deep learning method for single-trial EEG classification in RSVP task based on spatiotemporal features of ERPs," *J. Neural Eng.*, vol. 18, no. 4, Aug. 2021.
- [9] N.-S. Kwak, K.-R. Müller, and S.-W. Lee, "A convolutional neural network for steady state visual evoked potential classification under ambulatory environment," *PLoS One*, vol. 12, no. 2, p. e0172578, Feb. 2017.
- [10] A. Ravi, N. H. Beni, J. Manuel, and N. Jiang, "Comparing user-dependent and user-independent training of CNN for SSVEP BCI," *J. Neural Eng.*, vol. 17, no. 2, p. 26028, 2020.
- [11] A. Ikeda and Y. Washizawa, "Steady-State Visual Evoked Potential Classification Using Complex Valued Convolutional Neural Networks," *Sensors*, vol. 21, no. 16, 2021.
- [12] O. B. Guney, M. Oblokulov, and H. Ozkan, "A Deep Neural

- Network for SSVEP-Based Brain-Computer Interfaces,” *IEEE Trans. Biomed. Eng.*, vol. 69, no. 2, pp. 932–944, 2022.
- [13] X. Zhao, Y. Du, and R. Zhang, “A CNN-based multi-target fast classification method for AR-SSVEP,” *Comput. Biol. Med.*, vol. 141, p. 105042, 2022.
- [14] S. Tang, S. Yuan, and Y. Zhu, “Data Preprocessing Techniques in Convolutional Neural Network Based on Fault Diagnosis Towards Rotating Machinery,” *IEEE Access*, vol. 8, pp. 149487–149496, 2020.
- [15] V. V. Nikulin, G. Nolte, and G. Curio, “A novel method for reliable and fast extraction of neuronal EEG/MEG oscillations on the basis of spatio-spectral decomposition,” *Neuroimage*, vol. 55, no. 4, pp. 1528–1535, Apr. 2011.
- [16] T. Alotaiby, F. E. A. El-Samie, S. A. Alshebeili, and I. Ahmad, “A review of channel selection algorithms for EEG signal processing,” *EURASIP J. Adv. Signal Process.*, vol. 2015, no. 1, p. 66, Aug. 2015.
- [17] Y. Wang, S. Gao, and X. Gao, “Common Spatial Pattern Method for Channel Selection in Motor Imagery Based Brain-computer Interface,” *Conf. Proc. ... Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Conf.*, vol. 5, pp. 5392–5395, 2005.
- [18] M. Arvaneh, C. Guan, K. K. Ang, and C. Quek, “EEG channel selection using decision tree in brain-computer interface,” in *Proceedings of the Second APSIPA Annual Summit and Conference, Biopolis, Singapore, 14-17 December*, 2010, pp. 225–230.
- [19] M. Arvaneh, C. Guan, K. K. Ang, and C. Quek, “Optimizing the Channel Selection and Classification Accuracy in EEG-Based BCI,” *IEEE Trans. Biomed. Eng.*, vol. 58, no. 6, pp. 1865–1873, 2011.
- [20] J. K. Feng, J. Jin, I. Daly, J. Zhou, Y. Niu, X. Wang, and A. Cichocki, “An Optimized Channel Selection Method Based on Multifrequency CSP-Rank for Motor Imagery-Based BCI System,” *Comput. Intell. Neurosci.*, vol. 2019, p. 8068357, May 2019.
- [21] J. Zhang, C. Yan, L. Cao, and X. Gong, “Optimal channel set selection for SSVEP-based BCI using spatial temporal correlation,” in *2017 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)*, 2017, pp. 2038–2042.
- [22] L. Meng, J. Jin, and X. Wang, “A comparison of three electrode channels selection methods applied to SSVEP BCI,” in *2011 4th International Conference on Biomedical Engineering and Informatics (BMEI)*, 2011, vol. 1, pp. 584–587.
- [23] E. Webster, H. Habibzadeh, J. J. S. Norton, T. M. Vaughan, and T. Soyata, “An Unsupervised Channel-Selection Method for SSVEP-based BCI Systems,” in *2018 9th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*, 2018, pp. 626–632.
- [24] S. N. Carvalho, T. B. S. Costa, L. F. S. Uribe, D. C. Soriano, G. F. G. Yared, L. C. Coradine, and R. Attux, “Comparative analysis of strategies for feature extraction and classification in SSVEP BCIs,” *Biomed. Signal Process. Control*, vol. 21, pp. 34–42, 2015.
- [25] Y. Zhang, D. Guo, P. Xu, Y. Zhang, and D. Yao, “Robust frequency recognition for SSVEP-based BCI with temporally local multivariate synchronization index,” *Cogn. Neurodyn.*, vol. 10, no. 6, pp. 505–511, Dec. 2016.
- [26] M. Schroder, M. Bogdan, T. Hinterberger, and N. Birbaumer, “Automated EEG feature selection for brain computer interfaces,” in *First International IEEE EMBS Conference on Neural Engineering, 2003. Conference Proceedings.*, 2003, pp. 626–629.
- [27] D. A. Peterson, J. N. Knight, M. J. Kirby, C. W. Anderson, and M. H. Thaut, “Feature Selection and Blind Source Separation in an EEG-Based Brain-Computer Interface,” *EURASIP J. Adv. Signal Process.*, vol. 2005, no. 19, p. 218613, 2005.
- [28] L. Wang, D. Han, B. Qian, Z. Zhang, Z. Zhang, and Z. Liu, “The Validity of Steady-State Visual Evoked Potentials as Attention Tags and Input Signals: A Critical Perspective of Frequency Allocation and Number of Stimuli,” *Brain Sci.*, vol. 10, no. 9, Sep. 2020.
- [29] Z. İşcan and V. V. Nikulin, “Steady state visual evoked potential (SSVEP) based brain-computer interface (BCI) performance under different perturbations,” *PLoS One*, vol. 13, no. 1, 2018.
- [30] A. Schlögl, C. Keinrath, D. Zimmermann, R. Scherer, R. Leeb, and G. Pfurtscheller, “A fully automated correction method of {EOG} artifacts in {EEG} recordings,” *Clin. Neurophysiol.*, vol. 118, no. 1, pp. 98–104, 2007.
- [31] D. Ibáñez-Soria, A. Soria-Frisch, J. Garcia-Ojalvo, and G. Ruffini, “Characterization of the non-stationary nature of steady-state visual evoked potentials using echo state networks,” *PLoS One*, vol. 14, no. 7, p. e0218771, Jul. 2019.
- [32] W. M. Perlstein, M. A. Cole, M. Larson, K. Kelly, P. Seignourel, and A. Keil, “Steady-state visual evoked potentials reveal frontally-mediated working memory activity in humans,” *Neurosci. Lett.*, vol. 342, no. 3, pp. 191–195, May 2003.
- [33] G. Y. Bin, X. R. Gao, Z. Yan, B. Hong, and S. K. Gao, “An online multi-channel SSVEP-based brain-computer interface using a canonical correlation analysis method,” *J Neural Eng.*, vol. 6, 2009.
- [34] Z. İşcan and Z. Dokur, “A novel steady-state visually evoked potential-based brain-computer interface design: Character Plotter,” *Biomedical Signal Processing and Control*, 2013.
- [35] T. Bender, T. W. Kjaer, C. E. Thomsen, H. B. D. Sorensen, and S. Puthusserypady, “Semi-supervised adaptation in ssvep-based brain-computer interface using tri-training,” *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 4279–4282, 2013.
- [36] N.-S. Kwak, K.-R. Muller, and S.-W. Lee, “A lower limb exoskeleton control system based on steady state visual evoked potentials,” *J. Neural Eng.*, vol. 12, no. 5, p. 56009, Oct. 2015.
- [37] S. M. T. Muller, P. F. Diez, T. F. Bastos-Filho, M. Sarcinelli-Filho, V. Mut, and E. Laciari, “SSVEP-BCI implementation for 37-40 Hz frequency range,” *Conf. Proc. ... Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Conf.*, vol. 2011, pp. 6352–6355, 2011.
- [38] M. Nakanishi, Y. Wang, Y.-T. Wang, and T.-P. Jung, “A Comparison Study of Canonical Correlation Analysis Based Methods for Detecting Steady-State Visual Evoked Potentials,” *PLoS One*, vol. 10, no. 10, p. e0140703, Oct. 2015.

BIOGRAPHY



ZAFER İŞCAN received the B.S. degree (2002) in electronics and communication Engineering, M.S. degree (2005) in biomedical engineering, and Ph.D. degree (2012) in electronics engineering from Istanbul Technical University.

From 2005 to 2012, he was a research and teaching assistant in Department of Electronics and Communication Engineering at Istanbul Technical University. He worked as a postdoctoral researcher in Department of Psychiatry, Stony Brook University, US (2013 - 2014), Center for Cognition and Decision Making, Higher School of Economics, Russia (2014 - 2016), and Neurospin, INSERM, France (2016 - 2018).

Since 2018, he has been an Assistant Professor in the department of Electrical and Electronics Engineering, Bahçeşehir University. His research interests include computational and cognitive neuroscience, brain computer interfaces, pattern recognition, machine learning, artificial neural networks, and image processing.

Publication Ethics

The journal publishes original papers in the extensive field of Electrical-electronics and Computer engineering. To that end, it is essential that all who participate in producing the journal conduct themselves as authors, reviewers, editors, and publishers in accord with the highest level of professional ethics and standards. Plagiarism or self-plagiarism constitutes unethical scientific behavior and is never acceptable.

By submitting a manuscript to this journal, each author explicitly confirms that the manuscript meets the highest ethical standards for authors and coauthors

The undersigned hereby assign(s) to *Balkan Journal of Electrical & Computer Engineering* (BAJECE) copyright ownership in the above Paper, effective if and when the Paper is accepted for publication by BAJECE and to the extent transferable under applicable national law. This assignment gives BAJECE the right to register copyright to the Paper in its name as claimant and to publish the Paper in any print or electronic medium.

Authors, or their employers in the case of works made for hire, retain the following rights:

1. All proprietary rights other than copyright, including patent rights.
2. The right to make and distribute copies of the Paper for internal purposes.
3. The right to use the material for lecture or classroom purposes.
4. The right to prepare derivative publications based on the Paper, including books or book chapters, journal papers, and magazine articles, provided that publication of a derivative work occurs subsequent to the official date of publication by BAJECE.
5. The right to post an author-prepared version or an official version (preferred version) of the published paper on an internal or external server controlled exclusively by the author/employer, provided that (a) such posting is noncommercial in nature and the paper is made available to users without charge; (b) a copyright notice and full citation appear with the paper, and (c) a link to BAJECE's official online version of the abstract is provided using the DOI (Document Object Identifier) link.



ISSN: 2147- 284X
Year: July 2022
Volume: 10
Issue: 3

CONTENTS

B. Küçükkaraca, B. Barutcu; Life Cycle Assessment of Wind Turbine in Turkey,.....	230 - 236
C. Andiç, A. Öztürk, S. Tosun; Dynamic Economic Dispatch with Valve-Point Effect Using Crow Search Algorithm,.....	237 - 244
A. T. Çakmak; 4 Wheel Steering System Control Unit Design,.....	245 - 251
B. Ağgül, G.Erdemir; Development of a Counterfeit Veh. License Plate Det. System by Using Deep Learning,.....	252 - 257
M. Öztürk, A. Küçükmanisa, O. Urhan; Drowsiness . System Based on Machine Learning Using Eye State,.....	258 - 263
M. Uzun, O. Abul; Proactive Metering of Mobile Internet User Experience,.....	264 - 272
N. Ekren, A. S. Sarkin; Semi-conductor Applications to Printed Circuits on Flexible Surfaces,.....	273 - 277
H. A. Yıldız, O. Aydın; Design Consideration for Active–Only Memcapacitor Emulator Circuits,.....	278 - 285
E. Kizilkaplan, F. Yalçinkaya; The New HEMS Modelling of Human Heart,.....	286 - 294
T. Kocaer, Y. Öner; Axial Flux Motor Design for Ventilation Fans Used in The Automotive Industry,.....	295 - 299
M. Sülü, R. Daş; Graph Visualization of Cyber Threat Intelligence Data for Analysis of Cyber Attacks,.....	300 - 306
Y.Canbay, Ş.Sağiroğlu, Y.Vural; A New Anonymization Mod. for Priv.Preserving Data Publishing: CANON,.....	307 - 316
F. Asadi; Reduction of Output Impedance of Buck Converter with Genetic Algorithm,.....	317 - 322
B.Gecer, İ.Kiyak; Wavelength tune of InGaN based blue LEDs by changing indium perc.and op. volt.variables,.....	323 - 327
E. Gündüzalp, G. Yıldırım, Y. Tatar; Effic. Task Scheduling in Cloud Systems with Adap. Discrete Chimp Algorithm,.....	328 - 336
T. Inan, U. Kacar; Ear semantic segmentation in natural images with Tversky loss function supported DeepLabv3+ convolutional neural network,.....	337 - 346
Z. Işcan; Comparison of Deep Learning and Traditional Machine Learning Classification Performance in a SSVEP Based Brain Computer Interface,.....	347 - 355

BALKAN JOURNAL OF ELECTRICAL & COMPUTER ENGINEERING

(An International Peer Reviewed, Indexed and Open Access Journal)

Contact

Batman University
Department of Electrical-Electronics Engineering
Bati Raman Campus Batman-Turkey

Web: <http://dergipark.gov.tr/bajece>
<http://www.bajece.com>
e-mail: bajece@hotmail.com

