



Balkan Journal of Electrical & Computer Engineering

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

www.bajece.com

Vol : 10

No : 1

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, iijFactor. 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 : 1
Year: January 2022

CONTENTS

- Ş. Sivritaş, A. Tekin;** Autonomous 3D Target Tracking via UAV,.....1 - 9
- E. Kuşkan, M.A. Sahraei, M.Y. Çodur;** Classification of Aviation Accidents Using Data Mining Algorithms,.....10 - 15
- T. Ö. Onur, G. U. Kaya;** Application of Binary Genetic Algorithm for Holographic Vascular Mimicking Phantom Reconstruction,.....16 - 22
- H. Mutlu, C. Akuner, G. Akgün;** An Intelligent Machine Condition Monitoring Model for Servo Systems,.....23 - 29
- T. Tut;** Broadband Low Reflection Surfaces with Silicon Nano-pillar Square Arrays for Energy Harvesting,.....30 - 34
- H. Firat, M. E. Asker, D. Hanbay;** Hybrid 3D Convolution and 2D Depthwise Separable Convolution Neural Network for Hyperspectral Image Classification,.....35 - 46
- N. Noğay, H. S. Noğay;** Estimation of Survival According to Body Mass Index (BMI), Hypertension, Diabetes and Heart Disease with Optimizable Decision Trees,..... 47 - 52
- H. Sezer, A. Güven, O. B. Tör, M. E. Cebeci, S. Teimourzadeh;** A Cost-Based Approach to Elicit Ancillary Service Tariffs from Unified Capacity and Energy Tariffs,.....53 - 63
- F. A. Kazan, R. Akkaya;** The Effect of Road Slope on Parameters of Disabled Vehicle Driven by PMDC Motor,.....64 - 74
- A. Gunagwera, A. T. Zengin;** Longitudinal Inter-Vehicle Distance Control of Autonomous Vehicle Platoons Subjected to Internal and External Disturbances,.....75 - 84
- T. C. Akinci, A. A. Martinez-Morales;** Cognitive Based Electric Power Management System,.....85 - 90
- I. B. Çiçek, I. Sel, F. H. Yağın, C. Çolak;** Development of a Python-Based Classification Web Interface for Independent Datasets,.....91 - 96
- B. Kocaman, Y. Yiğit;** Real-Time Monitoring, Analysis and Control of Power Parameters in Residential Houses Using LabVIEW,.....97 - 105
- M. E. Yıldırım;** RSSI Based Indoor Localization with Reduced Feature Dimension,.....106 - 109

**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

Autonomous 3D Target Tracking via UAV

Şeyma Sivritaş and Ahmet Tekin


Abstract— In recent years, unmanned aerial vehicles (UAVs) have been integrated into many applications. These vehicles are used by the governments and the private sector for many different purposes. Some of the popular applications are field scanning, target tracking, defense equipment, search and rescue, map creation, underground resource search, photography or entertainment. Recently many academic institutions have shown significant amount of interest in this field and many of them have established aviation departments and societies. In the recent years, many thesis studies and academic publications related to UAVs have been presented. The aim of this particular study is to detect a small object identified as a target by UAV and to provide autonomous tracking of the detected target in three axes. As the outcome of this work, the UAV is coded to lock on the target and to maintain a fixed distance on the x, y and z axes after the initial lock. Within the scope of the study, a functional system was designed by evaluating the most important parameters for UAVs such as weight, cost and flight time. An algorithm was designed that uses the image processing methods to detect the target and movement direction of the target. The system utilizes MAVLink communication protocol to control the UAV. In addition, two different tests and test results corresponding to two distinct methods were reported which were done to keep the distance constant on the z-axis by using an external distance sensor or comparison algorithm. The project was successfully implemented using single camera only.

Index Terms—Unmanned aerial vehicle, UAV, autonomous system, 3D tracking, image processing, fast algorithm.


I. INTRODUCTION

IN RECENT DECADES, unmanned aerial vehicles (UAVs) have become quite widespread and popular. These vehicles may be deployed for many different purposes. The main use cases for these technologies are field scanning, target detection, target tracking, attack, search and rescue, map creation, photography or entertainment. UAVs are also popular topic for academic research [1-3].

ŞEYMA SİVRİTAŞ, is with Department of Electrical and Electronics Engineering, Ozyegin University, Istanbul, Turkey, (e-mail: seyma.sipar@gmail.com).

 <https://orcid.org/0000-0001-5957-0890>

AHMET TEKİN, is with Department of Electrical Engineering, Istanbul Technical University, Istanbul, Turkey, (e-mail: ahmettekin@itu.edu.tr).

 <https://orcid.org/0000-0001-8549-2582>

Manuscript received Jun 22, 2020; accepted Dec 23, 2021.
DOI: [10.17694/bajece.756443](https://doi.org/10.17694/bajece.756443)

Tracking an object by a drone is widely implemented project utilizing multiple simultaneously operating sensors. In one of these projects, mainly two embedded cameras were used for object detection and tracking. An algorithm was used to provide tracking in the 3th axis without a dedicated distance sensor [1].

There is another prominent work in the literature about autonomous flight. A reinforcement learning method was also applied in this work. In the particular system, a camera subsystem, which consists of 24 cameras, was used for this study [4].

In 2016, a group has reported a real time object tracking project and the project outcome was successfully demonstrated. In the particular project, a special motion capture camera system was used to detect the position of the moving object and Matlab Simulink was used to process the information. In addition, an external computer was used as a base station [5].

In another similar project, a visual detection and tracking implementation was realized with UAVs. In this project, Haar cascade classifier was used to detect the object and two cameras were used to implement the target locking function [6].

Within the scope of this work, a programmable autonomous UAV system was designed and a test mission was proven successfully. The mission was detection of a target object and 3D (3-dimensional) tracking (on x, y, z axes) of the target autonomously. In this paper, main components of the autonomous UAV system are listed and described. Also, design steps of the image processing subsystem were summarized. The method of detecting the position of the moving target and calculation of the updated coordinates according to displacement of the target were presented. Two different techniques for altitude tracking were analyzed and the results were compared.

II. METHODOLOGY AND DESIGN

Overall system consists of two main platforms; hardware and software platforms. The mission can be analyzed in three sub-sections to achieve autonomous 3D tracking task successfully.

A. Hardware Platform of the System

A UAV was used with SDK (Software Development Kit) flight controller to achieve the mission. Some auxiliary components (sensors) were also used to complete the autonomous system design. These components are:

- **GPS** was used to determine the located coordinates of the UAV.
- **Compass** was used to determine the directions.
- **Accelerometer** was used to determine the acceleration in three axes (x, y, z).
- **Gyroscope** was used to determine the motion in three axes (x, y, z).
- **Barometer** was used to determine the altitude by using pressure value.
- **On-board computer** was used to provide the communication with the flight controller of the UAV for behavior commands. It was also used as hardware platform for image processing software.
- **Camera** was used to get the image for image processing.
- **Distance sensor** was used to provide altitude tracking.

B. Software Platform of the System

1) Image Processing

Image processing application was used for detection and tracking of the target. All image processing steps were implemented online with an on-board computer. Within the scope of image processing, Python programming language and OpenCV library were preferred.

Python is an interpreted programming language and doesn't need compiling process [7]. OpenCV is a library for programming languages. It is open source and very helpful. This library provides significant support in coding and hence saves time [8]. Both of Python and OpenCV are widely used in an image processing area [9].

1.1 Color space

The image is captured in RGB color space from camera but it is converted to HSV (Hue, Saturation, Value) color space to have more details about color such as saturation and light intensity. The most accurate HSV values can be obtained by experimentally masking the target in the mission environment [10, 11].

1.2 Gaussian filtering (Blurring)

Gaussian filtering is a kind of low-pass filter and this function is used to blur an image. This filter prevents sharp transition (high frequency) in the image and provides smoothing. Noise components can be attenuated with the help of this filter [8, 12].

1.3 Masking

The desired object is masked from whole image by making the color of the target object white and color of the rest of the image black. This process is necessary to implement Morphological Transformations to be described in next section to obtain better quality from captured image by filtering out the noise [10, 11].

1.4 Erosion

The Erode function is used for eroding boundaries of the objects and this function provides decreasing the area of the objects. The Erode function is very useful for removing noise. To protect the size of the original objects, the "Dilation" function is used after erosion process [13-16].

1.5 Dilation

The dilate function is used to expand the boundaries of the objects and this function provides enlarging the area of the objects. The Dilate function is generally used with "Erode" function, after erosion [13-16].

2) Control Mechanism

The stabilization of the UAV was provided with the help of PID control mechanism integrated into flight controller. PID (Proportional, Integral, and Derivative) controller which controls the process variables by using a control loop feedback mechanism. This system is used to minimize the error of the system dynamically. The difference between a desired set point and a measured process variable is calculated by PID controller. There are three parameters in PID controller calculation. The first component is proportional control; it controls the actual value of the error. The second component is integral control; it takes into account the error history. Finally, third component is derivative control, it acts based on the rate of error [17, 18]. The block diagram of the PID controller is shown in Fig. 1 and the calculation method is shown in equation (1).

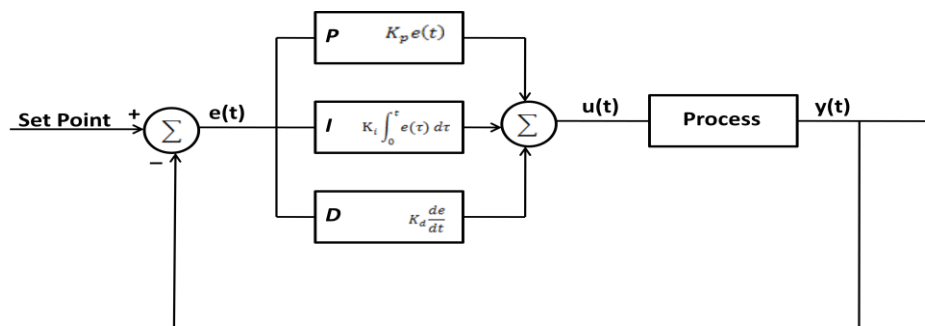


Fig.1. Block diagram of PID controller

$$u(t) = k_p e(t) + k_i \int_0^t e(\tau) d\tau + k_d \frac{de}{dt} \quad (1)$$

where,

Kp: Proportional gain

Ki: Integral gain

Kd: Derivative gain

u: Control signal

e: Error

3) Communication Protocol

MavLink (The Micro Air Vehicle Link) is one of the most popular communication protocols for Unmanned Aerial Vehicles. MAVLink protocol was used for communication with UAV and on-board computer. This protocol also can be used between UAVs and GCS (Ground Control Station) over a wireless channel [19].

C. Design and Mission Steps

Before starting the mission, the target was determined as a red point and a red ball with a diameter of 5 cm was used during the mission trials. The determined color was defined to the system with the HSV color space parameters. Since the color detection sensitivity of the system changes according to the ambient brightness, the appropriate color range was decided by experimenting. Detection of objects smaller than 3 pixels was inhibited to prevent unwanted pixel noise from being detected as a target. As a performance criterion, the system could detect and track a red point larger than 3 pixels which is within the capacity of the ultrasonic sensor. The mission can be divided into four parts to be able to analyze it in detail.

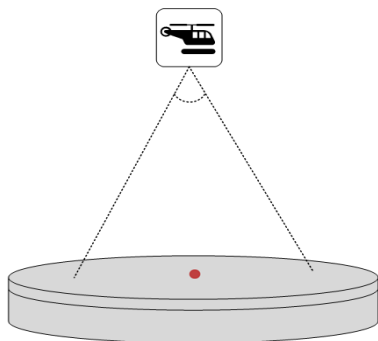


Fig.2. Illustration of angle of view

1) Target Detection

When the red point target is detected by the UAV, the UAV remains over the target. The resolution of the camera was set as 640 x 480 pixels in this case. This resolution values show that the horizontal area of the camera will be divided into 640 equal parts (pixels) and the vertical area will be divided into 480 equal parts (pixels). These horizontal and vertical areas

are related to angle of view of the camera. Illustration of the angle of view can be seen from Fig. 2.

The angle of view values of the used camera are as follows [20]:

Horizontal viewing angle: 53.50°

Vertical viewing angle: 41.41°

The field of view of the camera is calculated using angles of view of the camera with the help of cosine theorem. Illustration of the field of view is given in Fig. 3. Equation (2) and (3) can be used for these calculations.

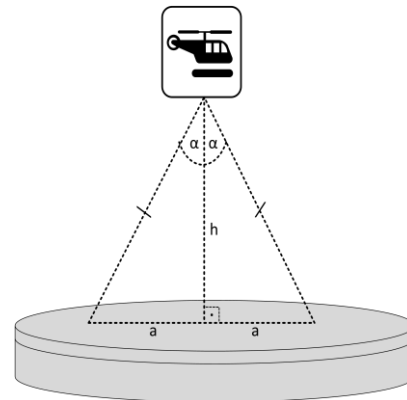


Fig.3. Illustration of field of view

$$\frac{h}{\cos(\alpha)} = \frac{a}{\sin(\alpha)} \quad (2)$$

$$a = h \frac{\sin(\alpha)}{\cos(\alpha)} \quad (3)$$

In above equations, h (altitude) is measured by the barometer and this data is taken from barometer output. The calculated field of view is divided by 640 for the horizontal axis and the length corresponding to a single pixel on the horizontal axis (x-axis) is obtained. Similarly, the calculated field of view of the vertical axis is divided by 480 and the length corresponding to a pixel on the vertical axis (y axis) is calculated. These calculations can be formulated as shown in equation (4) and (5).

$$P_h = \frac{FoV_h}{640} \quad (4)$$

$$P_v = \frac{FoV_v}{480} \quad (5)$$

where,

P_h: Unit pixel length on horizontal axis

P_v: Unit pixel length on vertical axis

FoV_h: Field of view on horizontal axis

FoV_v: Field of view on vertical axis

2) Calculation of New Coordinate of Moving Target

If the detected target moves out of the center of the camera, the autonomous system follows the target and gets the target position back to the center of the field of view. In order to do this, first, the new location of the detected target is determined. By comparing the pixel location at the center of the target with the pixel location at the center of the field of view, the target location can be obtained with respect to the center frame. Then the distance of the target from the center is calculated in meters by multiplying the pixel distance with pixel length. The illustration of this step can be seen from Fig. 4.

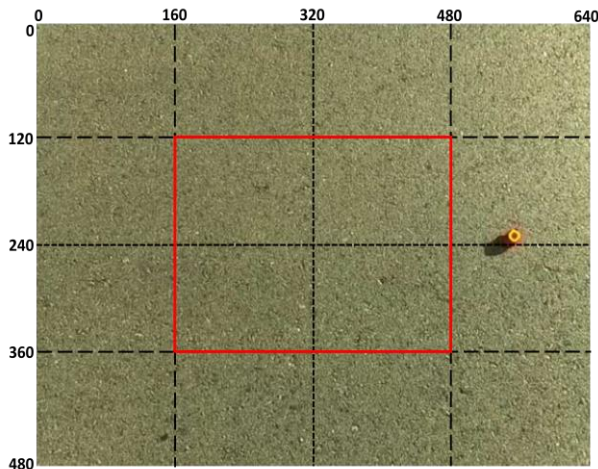


Fig. 4. Illustration of target pixel position

The center coordinate of the field of view is (320, 240) pixels on horizontal (x) and vertical (y) axes, respectively. The four sides (left, right, up, down) of the center were divided into two parts. The intersection points in all quarter parts specify the corner points of the tolerance frame. This frame is shown with red color in Figure 4. In this figure, target was detected at the pixel position of (546, 228) as seen from Fig. 5. Since the vertical 228th pixel is inside the tolerance frame, movement on the vertical axis is not required. However, the horizontal 546th pixel is out of tolerance frame, and the UAV should move to the right direction on the horizontal axis to make the target placed in the tolerance frame. General formulation to determine the required movement distance to get the target into center of the frame can be shown in the relation between Equations (6) and (9).

$$Pd_h = 320 - x \quad (6)$$

$$Pd_v = 240 - y \quad (7)$$

$$M_h = Pd_h \times P_h \quad (8)$$

$$M_v = Pd_v \times P_v \quad (9)$$

where,

Pd_h: Pixel distance from center on horizontal axis

Pd_v: Pixel distance from center on vertical axis

M_h: Required movement distance on horizontal axis

M_v: Required movement distance on vertical axis

```

pi@raspberrypi: ~
Dosya Düzenle Sekmeler Yardım
Searching for red point target
Red point is detected!
x : 546.5
y : 228.5
Go right!

```

Fig. 5. Command output window for detection of target pixel position

Calculated distance data is converted to coordinate by using coordinate calculation method. Latitude angle and Longitude angle are used as coordinate parameters. These angles can be seen from Fig. 6. Equations (10) and (11) can be used for calculation of the angle difference [21]. Calculated angle difference values are then added to original angle values to obtain the new coordinates as seen from equation (12) and (13).

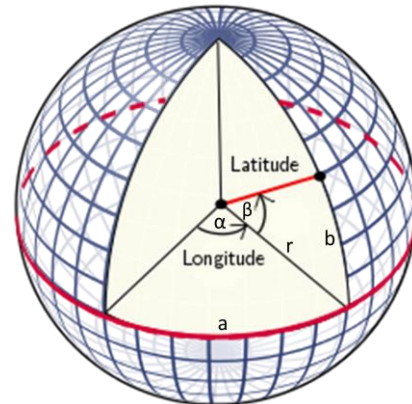


Fig. 6. Latitude and Longitude angles [22]

$$\beta = \frac{b}{r} \quad (10)$$

$$\alpha = \frac{a}{r \cos \beta} \quad (11)$$

$$\text{New Longitude angle} = \alpha_{\text{org}} + \alpha \quad (12)$$

$$\text{New Latitude angle} = \beta_{\text{org}} + \beta \quad (13)$$

where,

α: Longitude angle difference (radian)

β: Latitude angle difference (radian)

a: Displacement in latitude (m)

b: Displacement in longitude (m)

3) *Moving to New Coordinate*

The calculated coordinate values are sent to the UAV via the MAVLink protocol and the UAV is made to move over the target. This part can be seen from "Results" section clearly.

4) *Altitude Tracking*

After the x and y axes were checked and the position of the target was brought to the center of the frame by the autonomous UAV, the third axis, z axis (height), is checked. Then, the current distance value is compared with the target distance and the UAV behaves according to the comparison result and reaches the required height. Two different techniques have been implemented within the scope of the study.

4.1. *Distance calculation with algorithm*

In the camera based distance calculation technique, single camera and a simple pixel algorithm are used. No other components are necessary to obtain the approximate distance.

In this technique, a change in the radius of the object detected as a target is observed. The current altitude and radius are recorded when the target was first detected. If the target moves up or down while the UAV is hovering at a constant altitude after detection, the measured radius will also change. This change will be positive if the object moves upwards and approaches the UAV, and if the object moves downwards and away from the UAV the change will be negative. After the trigonometric calculations, the new radius and the previous radius are compared to determine the amount and direction of change in distance. Calculation can be made by using triangle similarity theorems. This process can be understood with the help of Fig. 7, equation (14) and (15).

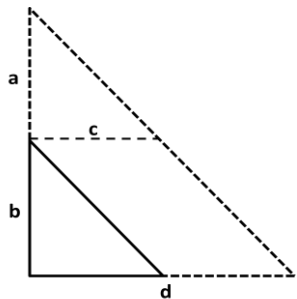


Fig. 7. Illustration of triangle similarity

$$\frac{a}{a + b} = \frac{c}{d} \tag{14}$$

Here, "d" represents the radius of the object (r_{ex}) and "a+b" represents the altitude whereas "c" represents the length of the new radius (r_{new}) and "a" value represents the new altitude.

$$\text{ratio} = \frac{r_{new}}{r_{ex}} \tag{15}$$

If ratio > 1;

it is understood that the object is rising and approaching the UAV. The distance is calculated by multiplying the current

height by calculated ratio. The obtained altitude value is sent to the UAV as an ascent command.

If ratio < 1,

it is understood that the object is descending and moving away from UAV. The distance is calculated by multiplying the current height by calculated ratio. The obtained altitude value is sent to the UAV as a descend command.

There is a possible problem while calculating the distance with this technique. In some cases, even if the position of the object does not change, its radius can be measured smaller than it appears. As a result, since the autonomous UAV perceive that the object is farther from the UAV, the UAV may want to approach the object. The illustration of this relation can be seen from Fig. 8.

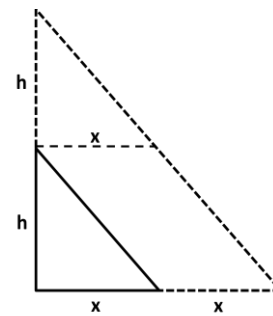


Fig. 8. Illustration of radius measurement

When the altitude of the UAV is "2h" and the radius of the target is "2x", if the radius of the object is measured as "x" instead of "2x", the UAV wants to reduce the altitude by the same radius ratio. Thus, the UAV will move downwards until the altitude is "h". This type of system behavior is undesirable. The source of this problem may be some factors such as image quality, camera resolution, and variation of light intensity or change of light angle. Some examples of this kind of problem can be seen in Fig. 9 through Fig. 12.

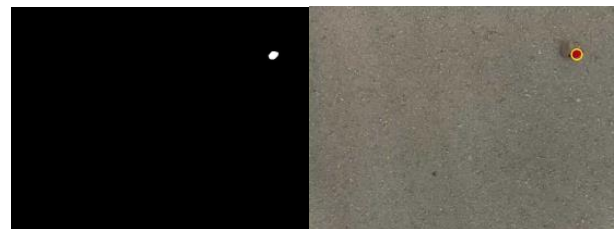


Fig. 9. Correct radius measurement

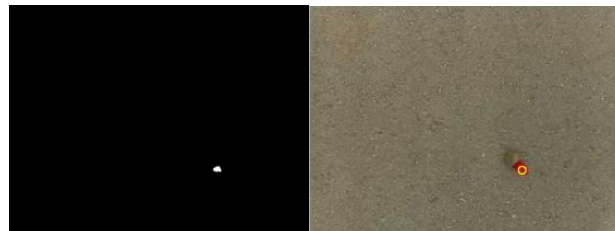


Fig. 10. Incorrect radius measurement



Fig. 11. Correct radius measurement



Fig. 12. Incorrect radius measurement

In this experiment, it was observed that the probability of obtaining an incorrect value as a result of the distance measurement made by the algorithm method is higher than the measurement method made while utilizing a precision distance sensor. Therefore, although the distance calculation method with the algorithm is lower in cost, lighter in weight and simpler in terms of coding, the mission is carried out by employing an additional distance sensor.

4.2. Distance calculation with distance sensor

Different distance sensors can be used in this method. Two commonly used distance sensors are LIDAR and Ultrasonic sensors.

4.2.1. Lidar sensor

Laser technology is used in LIDAR (Light Detection and Ranging) measurement sensor [23]. The single-beam narrow-band laser is generated and then it is reflected from target object [24]. The illustration of the working principle of Lidar sensor is shown in Figure 13. The distance can be calculated by the half of duration when the laser pulse being emitted and the time when the reflected laser pulse being sensed. The distance can be calculated by multiplying the half of the total travel duration by the speed of light (299,792,458 m/s). The result will be in meters and the formula can be seen from equation (16) [25].

$$D = t \cdot c \tag{16}$$

where,

D: Distance from object (m)

t: Half of time interval between generated and reflected laser pulse

c: speed of light (299,792,458 m/s)

4.2.2. Ultrasonic sensor

Sound wave technology is used for this type of distance sensor. The sensor generates 40 kHz sound pulse when it is

triggered then waits and listens to reflected sound wave. Therefore, the total travel time (2t) of the sound wave can be measured as transmitted and reflected duration [26]. The illustration of the working principle of ultrasonic sensor is shown in Fig. 14. The distance can be calculated by multiplying the half of the total travel duration by the speed of the sound wave (343 m/s) [27]. The result will be in centimeters and the formula can be seen from equation 17.

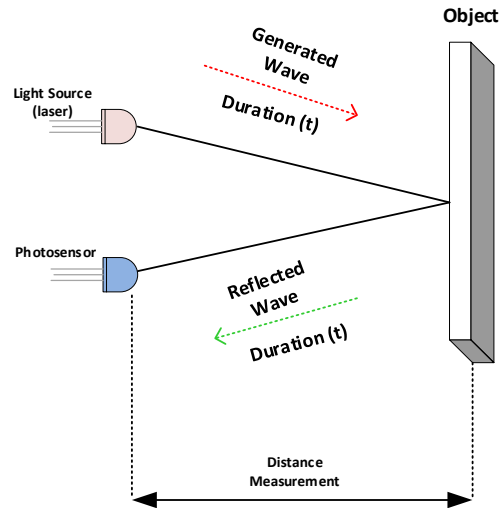


Fig. 13. Working principle of Lidar Sensor

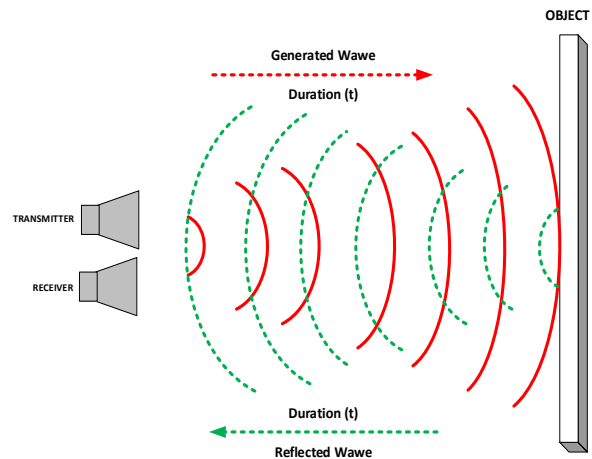


Fig. 14. Working principle of Ultrasonic Sensor

$$D = t \cdot 34300 \tag{17}$$

where,

D: Distance from object (cm)

t: Half of time interval between generated and reflected sound wave

From these two options, using LIDAR is high in terms of cost and hence, it was preferred to use an ultrasonic sensor in the scope of the study. The range of the chosen ultrasonic sensor is 400 cm.

The ultrasonic sensor is located very close point to the lens of the camera used for image processing. The sensor thus calculates the distance corresponding to the center of the field

of view. The distance from the target at the time the target was first detected during the task is recorded as the "target distance". A hysteresis loop-style control mechanism was integrated to prevent the oscillation because of minor measurement error and small swings of the drone. In this context, a tolerance of 10% was allowed for distance change. That is, if the object is between 0.9 x target distance and 1.1 x target distance, the UAV will continue to maintain its current position. If the height of the target changes more than the specified tolerance, the result obtained by sensor measurement is compared with the first value recorded, target distance. If the difference is smaller than the target distance, it is understood that the object rises and approaches the UAV. The new altitude is calculated by adding the obtained difference with the current altitude. The calculated altitude value is sent to the UAV as an ascent command. If the difference is greater than the target distance, it is understood that the object is lowered away from the UAV. The new altitude is calculated by subtracting the obtained difference from the current altitude. The calculated altitude value is sent to the UAV as a descend command.

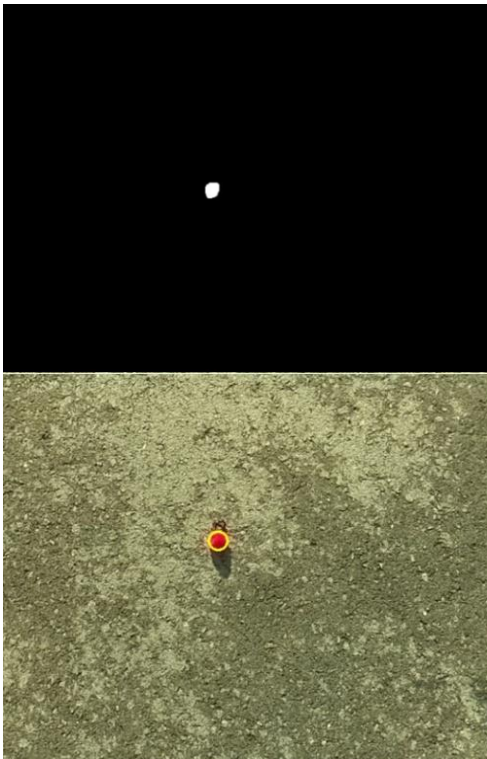


Fig. 15. Detection of the target

This process can be represented with the following equations (18) and (19). Altitude tracking photos can be seen from Fig. 15 to Fig. 17.

$$D = C - T \quad (18)$$

$$N = A - D \quad (19)$$

where,

A: Current Altitude (obtained from barometer)

C: Current Distance. The current distance between UAV and target (obtained from ultrasonic sensor).

D: Difference from target distance

N: New Altitude

T: Target Distance. The distance between UAV and target at first target detection (constant, obtained from ultrasonic sensor).

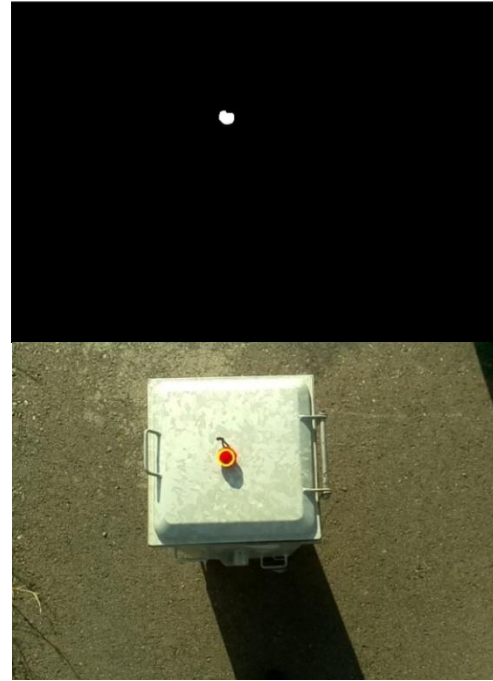


Fig. 16. Altitude tracking of the target



Fig. 17. Measurement of displacement in z-axis (height)

III. RESULTS

A simple mission is illustrated in Fig. 18 and tracking commands are given as shown in Fig. 19. The command window output was simplified to reduce the amount of writing and facilitate intelligibility. The basic functional behavior of the autonomous UAV system can be observed from the test track shown in Fig. 18. This figure is taken from GCS and the UAV was illustrated as green colored quad-copter on this satellite street view. The UAV was taken off from Home

position (represented by H letter with green color). While searching the field, a target was detected at the point shown as "Tracking is Starting" with yellow color. Then the related commands were sent to the flight controller by on-board computer according to movement of the target. Movement positions were drawn by ground control station with purple line and directions were shown with red arrows. Commands were added on the figure with red color as well. At point "A", the target was placed 1 meter above the ground and this difference was detected by UAV as seen from the command window.

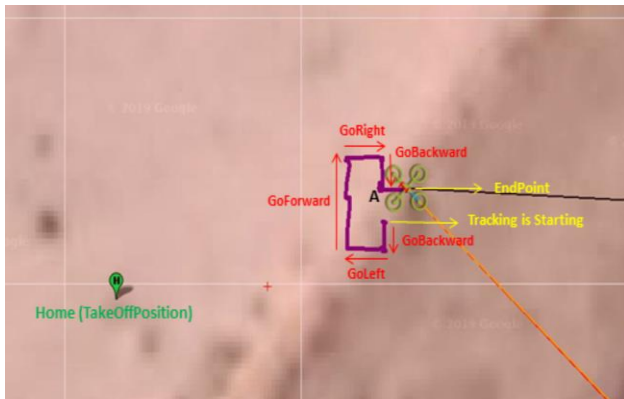


Fig. 18. Illustration of 3D tracking mission

```

pi@raspberrypi: ~/ball
Dosya Düzenle Sekmeler Yardim
Searching for red point target
Red point is detected!
Go backward!
Locked
Measuring distance...
Target Distance: 157,48 cm
Go left!
Locked
Measuring distance...
Distance is in range
Go forward!
Go forward!
Go right!
Locked
Measuring distance...
Distance is in range
Go backward!
Locked
Measuring distance...
Distance is below range
Distance: 52,73 cm
Go up!
Measuring distance...
Distance is in range
Go right!
Locked
Measuring distance...
Distance is above range
Distance: 249,14 cm
Go down!
Measuring distance...
Distance is in range
  
```

Fig. 19. Command output window of 3D tracking mission



Fig. 20. Photo of 3D tracking mission

Finally, target was placed in the previous position again and the UAV was moved down until it reached target distance as seen in the command window. At this point, the mission was complete and the point was marked as "End Point" and indicated by the yellow arrow. The photo of the autonomous 3D target tracking mission in action is shown in Fig. 20.

IV. DISCUSSION AND CONCLUSION

A target object can be detected and can be tracked on all three axes (x, y, z axes) with the aid of a UAV, an image processing system and a distance sensor. The distances to be moved for right-left (x axis/roll) and forward-backward directions (y axis/pitch) were calculated by using coordinate calculation methods, subsequently they are sent to flight controller and the UAV reached the target points successfully. It has been observed that using a single camera and basic distance algorithm option for tracking altitude (z axis) is not a good enough solution to get stable results due to environmental factors and systemic offsets in the camera sensor. The mission of three-dimensional (x, y, z axes) tracking of an object was successfully completed using a single camera, an additional distance sensor and a simpler algorithm. All project components were used by taking into account factors such as cost, weight and ease of application. Such a system can be customized for various tracking applications. The designed UAV can be used for applications such as; entertainment, scanning and tracking, with an even more attractive use case in the defense markets. If an object with complex boundaries is desired to be detected, it will be a little more complex and difficult to define these boundaries clearly, but, progressively smarter algorithms can enable these applications in the near future. The system can be improved significantly by using additional long-range Lidar sensors and high-resolution cameras for applications where the cost is not as important concern. Thus, tracking tasks can be performed more accurately and precisely with such an upgraded system.

REFERENCES

- [1] R. Barták, A. Vyškovský, "Any Object Tracking and Following by a Flying Drone." Fourteenth Mexican International Conference on Artificial Intelligence. Cuernavaca, US, 2015.
- [2] A. Hinas, J.M. Roberts, F. Gonzales, "Vision-Based Target Finding and Inspection of a Ground Target Using a Multirotor UAV System", *Sensors*, vol. 17, no.12, 2017, pp. 1-17.
- [3] A. Audi, M. Pierrot-Deseilligny, C. Meynard, C. Thom, "Implementation of an IMU Aided Image Stacking Algorithm in a Digital Camera for Unmanned Aerial Vehicles", *Sensors*, vol. 17, no.7, 2017, pp. 1-21.
- [4] J. Junell, E.J. Kampen, C. Visser, Q. Chu, "Reinforcement Learning Applied to a Quadrotor Guidance Law in Autonomous Flight." AIAA Guidance, Navigation, and Control Conference, Florida, US, 2015.
- [5] A. Mashood, A. Dirir, M. Hussein, H. Noura, F. Awwad, "Quadrotor Object Tracking using Real-Time Motion Sensing." 5th International Conference on Electronic Devices, Systems, and Applications, Ras Al Khaimah, United Arab Emirates, December 2016.
- [6] D. A. Mercado-Ravell, P. Castillo and R. Lozano, "Visual Detection and Tracking with UAVs, Following a Mobile Object", *Advanced Robotics*, vol.33, pp. 1-15, 2019.
- [7] M. Lutz, *Learning Python*, O'Reilly Media, 2007, California, pp.700.
- [8] S. Brahmhatt, *Practical OpenCV*, Apress, USA, pp.223.
- [9] R. Chityala, S. Pudipeddi, *Image Processing and Acquisition using Python*, CRC Press, Florida, pp.350.
- [10] A. Mordvintsev, K. Abid, Changing Colorspaces. Accessed: Jul. 2019. [Online]. Available:

- https://opencv-python-tutroals.readthedocs.io/en/latest/py_tutorials/py_imgproc/py_colorspaces/py_colorspaces.html
- [11] OpenCV, Thresholding Operations using inRange. Accessed: Jul. 2019. [Online]. Available: https://docs.opencv.org/3.4/da/d97/tutorial_threshold_inRange.html
- [12] OpenCV, Smoothing Images. Accessed: Jul. 2019. [Online]. Available: https://docs.opencv.org/3.4.2/d4/d13/tutorial_py_filtering.html
- [13] A. Koschan, M. Abidi, *Digital Color Image Processing*, John Wiley & Sons, Inc., New Jersey, US, pp.375.
- [14] B. Jähne, *Digital Image Processing*, 6th, revised and extended edition, 2005, Springer, Berlin, Germany, pp.607.
- [15] W.K. Pratt, *Digital Image Processing: PIKS Scientific inside*, 4th edition, 2001, John Wiley&Sons, Inc., New Jersey, US, pp.807.
- [16] OpenCV, Morphological Transformations. Accessed: Aug. 2019. [Online]. Available: https://docs.opencv.org/3.4.2/d9/d61/tutorial_py_morphological_ops.html
- [17] R. Gamasu, V.R.B. Jasti, "Robust Cohen-Coon PID Controller for Flexibility of Double Link Manipulator", *International Journal of Control and Automation*, vol. 7, no.1, 2014, pp. 357-368.
- [18] K.J. Åström, T. Hägglund, *PID Controllers: Theory, Design, and Tuning*, 2nd edition, Instrument Society of America, 1995, USA, pp.338.
- [19] S. Atoev, K. Kwon, S. Lee, K. Moon, "Data Analysis of the MAVLink Communication Protocol." *International Conference on Information Science and Communications Technologies*. Tashkent, Uzbekistan, 2017.
- [20] Raspberry pi, Camera Module. Accessed: May. 2019. [Online]. Available: <https://www.raspberrypi.org/documentation/hardware/camera/>
- [21] J. Case, *Astro Navigation Demystified*, Revised edition, Bookcase Learning Resources, 2016, Weymouth, UK, pp. 324.
- [22] A. Singh, Latitude and Longitude are angles, not just some numbers on the earth's surface. Accessed: Aug. 2019. [Online]. Available: <https://medium.com/@atinders/latitude-and-longitude-are-angles-not-just-some-numbers-on-the-earths-surface-d16019aaf24a>
- [23] J. Laconte, S. Deschênes, M. Labussière, F. Pomerleau, "Lidar Measurement Bias Estimation via Return Waveform Modelling in a Context of 3D Mapping." *International Conference on Robotics and Automation*. Montreal, Canada, 2019.
- [24] X. Li, Y. Guo, "Application of LiDAR Technology in Power Line Inspection." *International Conference on Advanced Materials, Intelligent Manufacturing and Automation*. China, 2018.
- [25] G. Duh. Light Detection and Ranging (LiDAR), Lecture notes. Portland State University.
- [26] N.A. Latha, B. R. Murthy, K. B. Kumar, "Distance Sensing with Ultrasonic Sensor and Arduino", *International Journal of Advance Research, Ideas and Innovations in Technology*, vol. 2, no.5, 2016, pp. 1-5.
- [27] A. Pantano, D. Cerniglia, "Simulation of laser-generated ultrasonic wave propagation in solid media and air with application to NDE", *Applied Physics A*, vol. 98, no.2, 2010, pp. 327-336.



AHMET TEKİN has received his EE PhD degree from University of California Santa Cruz, CA, EE MS degree from North Carolina A&T State University, Greensboro, NC and EE BS degree from Bogazici University, Istanbul, Turkey in 2008, 2004, and 2002, respectively. In addition to academic research in microelectronics, he worked for multiple innovative semiconductor design companies such as; Multigig, Inc., Newport Media, Aydeekay LLC, Broadcom corp., Semtech Corp., Nuvoton Technology Corp., Qualcomm and Waveworks Inc., leading semiconductor designs for communications, consumer and medical markets. His main focus area is analog/RF/mixed-signal integrated circuit design for communication and biomedical applications. He is currently engineering faculty member of Istanbul Technical University and serves in the board of directors of Waveworks Inc, Mission Viejo, CA.

BIOGRAPHIES

ŞEYMA SİVRİTAŞ received her B.Sc. degree of Electrical and Electronics Engineering from Dokuz Eylül University, İzmir, Turkey in 2015. She worked at Vestel Electronics between 2015 and 2018. She received her M.Sc. degree in the department of Electrical and Electronics Engineering from Özyeğin University, İstanbul, Turkey in 2020. She has been working at Turkish Aerospace



as an Avionics System Engineer since 2018. She has experience in aircraft avionics systems design, power distribution system design and analysis of communication systems.

Classification of Aviation Accidents Using Data Mining Algorithms

Emre Kuşkan, Mohammad Ali Sahrei and Muhammed Yasin Çodur


Abstract— Air transportation is a very preferred type of transportation for long-distance trips worldwide. This type of transportation has made great progress with the development of technology. In addition to its technological developments, passenger capacity is gradually increasing due to its fast and secure access. In contrast, the mortality rate is quite high in the case of an airplane accident, and hundreds of people die in a single accident. This research aims to classification several airplane accidents to find crucial factors and their overall impacts on the mentioned accident. In this study, appropriate data associated with said accidents worldwide since 2000 have been collected and then analyzed using sequential minimal optimization, decision tree (J48), and naive bayes. It is revealed that the decision tree algorithm provided the most accurate results for the study. Finally, appropriate comments were elaborated about each stage to reduce accidents. If these evaluations are taken into account, air transport will be much more reliable and thus loss of life will be minimized.

Index Terms—Aviation safety, sequential minimal optimization, decision tree, naive bayes.


I. INTRODUCTION

AN IMPORTANT part of fast and secure access in the world is provided by air transportation. It is highly preferred especially for fast access in long distance travels. Airway transportation was very dangerous for the safety of passengers when it first started while with the advancement of technology, it has become more reliable.


EMRE KUŞKAN, is with Department of Civil Engineering University of Erzurum Technical University, Erzurum, Turkey, (e-mail: emre.kuskapan@erzurum.edu.tr).

 <https://orcid.org/0000-0003-0711-5567>

MOHAMMAD ALI SAHRAEI, is with Civil Engineering Department, Faculty of Engineering, Girne American University, Girne, N. Cyprus Via Mersin, Turkey, (e-mail: mohammadalisahraei@gau.edu.tr)

 <https://orcid.org/0000-0002-9130-3685>

MUHAMMED YASİN ÇODUR, is with 1. Department of Civil Engineering University of Erzurum Technical University, Erzurum, Turkey, 2. Department of Civil Engineering of American University of the Middle East, Kuwait (e-mail: mycodur@erzurum.edu.tr).

 <https://orcid.org/0000-0002-7647-2424>

Manuscript received September 10, 2020; accepted Sep 18, 2021.

DOI: [10.17694/bajece.793368](https://doi.org/10.17694/bajece.793368)

However, due to increase the number of flights, plane accidents are dramatically raised. Since most of these accidents occur at higher speeds than the ground, deaths and injuries are also high. In this case, people are unlikely to survive if the planes fall from a certain height. In order to reduce these loss of lives, it is necessary to classify them and conduct detailed research related to the mentioned classifications. To this end, there are several studies worldwide to investigate plane crashes.

Studies offer different perspectives by examining plane crashes with various factors. Among these factors, human factors (pilot, cabin crew, passenger and air traffic controller) rank first among the factors affecting accidents [1-3]. The human factor causes about 75% of plane crashes and incidents. This factor consists of reasons such as alcohol, fatigue, carelessness, communication problem, non-compliance with the procedure. In this context, it is possible to reduce plane crashes, especially if the pilot and cabin crews pay attention to the stated reasons [4,5]. On the other hand, it is possible to prevent human factors that cause accidents by increasing the service quality in aircraft and by choosing qualified personnel [6]. Apart from this situation, a problem may arise from the passengers due to hijacking, disabling the cabin or pilots. To prevent this, it is necessary to raise awareness and increase security measures [7]. An increase in the occurrence of anxiety-related errors is observed due to exceeding the air traffic volume which causes to increases the likelihood of pilots making a mistake [8]. In addition to the human factor, aviation accidents can occur due to weather conditions, bird strikes, cabin pressure problems, technical and fuel problems. Technical defects on the aircraft may be caused by maintenance and production [9,10]. It can be analysed using various methods for factors affecting plane crashes which is aimed to prevent flight accidents. In one study, Cheng et al. [11] examined the reports of flight events received at the air traffic management centre. Using Heinrich's pyramid theory in line with these reports; established a quantitative relationship between major, minor and non-injury accidents. Thus, the effects of system failures on accident severity have been identified. Kaleta and Skorupski [12] have simulated the aircraft landing systems as an important element in air transportation, utilizing fuzzy logic approach. As a result of simulation experiments, the effect of the aircraft landing system on accidents was interpreted. In another study, aviation accidents were analysed by the Netherlands Air Traffic Control using data on traffic management. The types of incidents in

these aviation accidents were evaluated and the accident rates were calculated and the rates were interpreted [13]. Models containing Petri networks are also used to assess the possibility of aviation accidents. These networks create computer-aided fuzzy logic risk matrices to predict aircraft accident situations [14]. Fuzzy logic used to examine factors such as pilot's flight skill levels, airport traffic volume, weather conditions, airport procedures, and airport geometry. With this method, it is possible to calculate the probability of an event turning into an accident. In this way, weak points of security systems can be detected. Besides, probability estimates can be developed for different event situations [15, 16]. Another method which can be utilized for investigation of air traffic risk assessments and aviation accident models is Monte Carlo simulation. The risk of a collision between the aircraft taxiing with an aircraft taking-off may examine, which is based on dedicated Monte Carlo simulations in combination with a validation approach of the simulation results. The results particularly may be focused on the effectiveness of a runway incursion alert system that warns an air traffic controller, in reducing the safety risk for good and reduced visibility conditions [17]. Planning and controlling the flights are provided by air traffic management centres. The main goal of this management system is to control the risk of accidents. Because of the lack of communication between the air traffic controller and the pilot, it may cause crucial issues in the take-off and landing of the plane [18]. In a study on this subject, mobile technology was used in an application to increase the communication between the pilot and flight crew or air traffic controller. This system is generally designed as smartphones or tablets that pilots carry, use, and can be fixed in the cockpit inside the plane [19].

When the previous studies are examined, it is obviously shown that classifying the accidents into certain classes will facilitate the investigation of the causes of aviation accidents.

In this study, the aircraft accident data in the last 20 years in the world has been examined. These accidents are detailed as takeoff, initial climb, en-route, approach, and landing. Then, analysis was made using data mining algorithms. Accordingly, the analysis was performed using data mining algorithms utilizing sequential minimal optimization (SMO), decision tree (J48), and Naive Bayes. Then a comparison analysis among algorithms were conducted and the best algorithm determined and some suggestions to reduce aviation accidents were provided.

II. MATERIAL AND METHOD

A. Accident data

An average of 40 million flights have been performing annually worldwide. All flights are controlled by the air traffic control centres where it is possible to get instant information from these centres in case of any malfunction in the flights [20]. In general, aviation accidents may occur due to the pilot, cabin crew, passenger, air traffic control centre, weather conditions, and general condition of the aircraft. Classes are created using the stages in which plane crashes occur. The phases in which an airplane crashes are given in Figure 1. Information on the total number of aviation accidents occurring in the last 20 years, the number of death, and the phase of the accidents are given in Table 1.

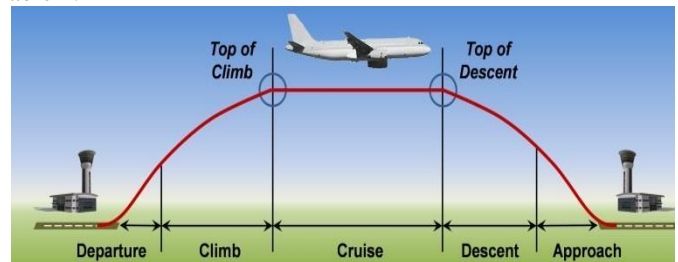


Fig. 1. Phases of aircraft accidents

TABLE I
AIRCRAFT ACCIDENT INFORMATION OCCURRING IN THE LAST 20 YEARS IN THE WORLD [21]

Year	Number of accidents	Number of deaths	Takeoff	Initial Climb	En Route	Approach	Landing
2000	43	1148	4	5	17	12	3
2001	36	879	3	3	13	14	1
2002	42	1000	2	0	16	20	2
2003	34	705	5	3	9	13	0
2004	35	462	2	2	13	12	4
2005	40	1075	2	6	15	9	4
2006	33	905	1	4	18	4	5
2007	32	774	2	4	15	4	7
2008	35	595	4	2	14	10	4
2009	32	763	4	6	9	8	4
2010	32	943	0	6	11	10	5
2011	36	525	1	3	16	12	3
2012	24	477	1	6	3	11	2
2013	28	232	3	1	7	13	4
2014	20	692	0	3	12	3	2
2015	14	186	0	4	9	1	0
2016	17	258	3	1	11	2	0
2017	14	59	0	2	5	4	2
2018	18	561	1	1	11	3	2
2019	23	288	4	1	10	4	4

Table 1 shows an average of 29 aviation accidents occurred annually in the world as well as an average of 626 people died annually in these accidents. Analysis of flight phases demonstrated that accidents are mostly happened during cruise and landing situations. Therefore, human errors are the biggest factor of accidents at these phases. The fatigue and loss of attention among pilots especially on long journeys support this idea. In addition, Table 1 displays that the number number of accidents and dies decreased in recent years compared to the beginning of the 2000s. While around 20 million flights were carried out annually in the early 2000s, today this number has doubled. Despite the increasing amount of flights and the number of passengers, the decrease in the number of accidents shows that the plane journey is becoming safer. However, the high mortality rate of major accidents occurring may cause anxiety in humans. In order to analyze and classify these accidents, data mining algorithms will be described in details in the following section.

B. Data mining

Data mining is the process of discovering patterns in large datasets that include methods of machine learning, the intersection of statistics, and database systems. Data mining is an interdisciplinary subfield of computer science and statistics, which has a general-purpose to extract information (by intelligent methods) from a dataset and turn it into an understandable structure for further use. In addition to the analysis step, it also includes database and data management aspects, data pretreatment, model and extraction issues, and criteria of interestingness, complexity issues, post-processing of discovered structures, visualization, and online updating. Many computer programs are used for the analysis of these algorithms. In this study, WEKA (Waikato Environment for Knowledge Analysis) software, i.e. a machine learning program, was used which includes data mining algorithms and methods [22].

1) Sequential minimal optimization (SMO)

SMO algorithm is derived by taking the idea of the decomposition method to its extreme and optimizing a minimal subset of just two points at each iteration. The power of this technique resides in the fact that the optimization problem for two data points admits an analytical solution, eliminating the need to use an iterative quadratic programming optimizer as part of the algorithm. In this method, the choice of the two points is determined by a heuristic, while the optimization of the two multipliers is performed analytically. Despite needing more iterations to converge, each iteration uses so few operations that the algorithm exhibits an overall speed-up of some orders of magnitude. Besides convergence time, other important features of the algorithm are that it does not store the kernel matrix in memory, since no matrix operations are involved, that it does not use other packages, and that it is fairly easy to implement [23]. Note that since standard SMO does not use a cached kernel matrix, its introduction could be used to obtain a further speed-up, at the expense of increased space complexity. Consider a binary classification problem with a dataset $(x_1, y_1), \dots, (x_n, y_n)$, where x_i is an input vector and $y_i \in \{-1, +1\}$ is a binary label corresponding to it. A soft-margin

support vector machine is trained by solving a quadratic programming problem, which is expressed in the dual form as follows:

$$\max \sum_{i=1}^n a_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n y_i y_j K(x_i x_j) a_i a_j \quad (1)$$

Subjected to: $0 \leq a_i \leq C$ for $i = 1, 2, 3, \dots, n$

$$\sum_{i=1}^n a_i y_i = 0 \quad (2)$$

where C is an SVM hyper parameter and $K(x_i x_j)$ is the kernel function, both supplied by the user; and the variables a_i are Lagrange multipliers [24].

2) Classification via naive Bayes

Naive Bayes is an algorithm that performs transactions based on probability calculations. It processes the found train data according to its formula and extracts a percentage for each case and performs the classification of the test set according to these probabilities [25]. $P(A/B)$ is the probability that event A occurs when event B occurs (see conditional probability), $P(B/A)$ is the likelihood that event B will occur when event A occurs. $P(A)$ and $P(B)$; A and B are the preliminary probabilities of events [26]. The algorithm explained Equation (3).

$$P(A / B) = [P(B/A) \times P(A)]/P(B) \quad (3)$$

3) Classification via J48 (decision trees) algorithm

This algorithm, i.e. J48 (decision trees), is a C4.5 decision tree developed to classify nonlinear and small size data. The decision tree approach is important in solving classification problems where a tree is created to model the classification process. After creation the tree, the classification process takes place by applying it to each data group in the database [27,28]. In this regard, the missing values are ignored during creating the tree. Thus, estimation is performed using the remaining data. The basic idea in the J48 method is to classify using the rules produced by decision trees [29]. After calculation of the entropy value, the information value is calculated for each predictive variable and then the information gain is calculated. The purpose of all these calculations is to find the predictive class that provides the highest level of knowledge. Accordingly, the entropy value is calculated using Equation (4) and indicates the probability of an unexpected situation occur. If the samples are homogeneous, the entropy value is zero and if the values are equal, entropy becomes one. The entropy equation can be calculated by Equation (5) as well as the information equation value provided by Equation (6). It is based on subtracting all data from entropy after dividing a data set on a feature. The c value according to the formulas gives the number of values that the target variable can take. The S value gives the target variable and the T value gives the predictive variable [30].

$$E(S) = \sum_{i=1}^c -p_i \log_2 p_i \quad (4)$$

$$E(T, X) = \sum_{c \in X} P(c) E(c) \quad (5)$$

$$\text{Gain}(T, X) = \text{Entropy}(T) - \text{Entropy}(T, X) \quad (6)$$

4) WEKA

WEKA contains many machine learning algorithms, which were developed using Java, for carrying out many data mining processes and was developed by the Machine Learning Group at the University of Waikato in New Zealand [31]. This is not one single program, contains many algorithms for analysis and predictive modeling where these algorithms can be directly applied to the dataset. WEKA comprises of many tools for the data mining activities like the classification, data pre-processing, clustering, regression, association rules, or visualization. This tool also helps in developing many additional machine learning techniques. Furthermore, it also contains many classes that could be easily accessed by other WEKA classes. The essential WEKA classes are the attribute and the instance. The attribute can be represented by any object of the class attributes that contain the attribute name, type, and the values of the nominal attributes [32].

5) Data mining performance and error scales analysis

While accomplishing performance analysis in data mining, basic success criteria concepts are used. These concepts are precision, sensitivity, F-measure, and ROC criteria. During calculation of the values of said concepts, the comparison of the estimated and available data is taken into account [33]. In the comparison process, True Positive-right (TP) means True Negative-right means false (TN), False Positive-false means (FP), and False Negative-false means wrong (FN) values are used.

		Predictive Values	
		Class =1	Class =0
Actual Values	Class=1	TP	FN
	Class =0	FP	TN

Fig. 2. Confusion matrix

Using the confusion matrix given in Figure 2, the accuracy values of the classification algorithms can be calculated. The precision statement is the ratio of the number of correct and positive estimated samples as class 1 to the number of estimated samples as class 1, as indicated in Equation (7) [34]. Sensitivity is defined as the ratio of the number of positive samples correctly classified in Equation (8) to the total number of positive samples. The F-criterion is stated as the harmonic mean of these two expressions in Equation (9) to evaluate both the sensitivity and precision expressions together [35]. The ROC value is obtained with the created curve to interpret the model performance in general. All of these performance values take values between 0 and 1.

$$\text{Precision} = \text{TP}/(\text{TP}+\text{FP}) \quad (7)$$

$$\text{Recall} = \text{TP}/(\text{TP}+\text{FN}) \quad (8)$$

$$\text{F-measure} = (2 \times \text{Recall} \times \text{Precision}) / (\text{Recall} + \text{Precision}) \quad (9)$$

$$\text{Accuracy} = (\text{TP}+\text{TN}) / (\text{TP}+\text{FP}+\text{FN}+\text{TN}) \quad (10)$$

The error scales of the model are determined by the accuracy rate, mean square error (MAE), root mean square error (RMSE), and Kappa statistics. Although the accuracy rate is shown in Equation (10) which is the most important criterion for the success of the algorithm, it indicates how appropriate the predicted value is to the real value. MAE is expressed as the average of the difference between estimated values and actual values of all data. The RMSE value is calculated by taking the square root of the mean of the difference between the values estimated by the model and the actual values obtained [44]. Kappa value, on the other hand, is a term expressed to measure the mismatch between observational. The closer this value is to 1, the better the agreement between observations. The kappa statistic is frequently used to test interrater reliability. The importance of rater reliability lies in the fact that it represents the extent to which the data collected in the study are correct representations of the variables measured [36]. The kappa statistic formulated as Equation (11).

$$K = \frac{p_o - p_e}{1 - p_e} \quad (11)$$

Where p_o and p_e are expectation and observation, respectively. The meaning of this calculation has come into question, and ranges for the measure vary. However, an example would include the following: $K < 0.20$ = poor agreement; $K = 0.21$ to 0.40 is fair; $K = 0.41$ to 0.60 is moderate; $K = 0.61$ to 0.80 is substantial; and $K > 0.81$ is good [37].

MAE as a model evaluation metric is used with regression models. MAE error of a model concerning a test set is the mean of the absolute values of the individual prediction errors over all instances in the test set. Each prediction error is the difference between the true value and the predicted value for the instance. MAE, as provided by Equation (12), measures the closeness of the predictions to the eventual outcomes.

$$MAE = \frac{1}{N} \sum_{i=1}^N |x_{f,i} - x_{o,i}| \quad (12)$$

Where $x_{f,i}$ and $x_{o,i}$ are the i th expectation and observation, respectively.

RMSE is the square root of mean squared error. RMSE measures the differences between values predicted by a hypothetical model and the observed values. In other words, it measures the quality of the fit between the actual data and the predicted model. RMSE, as provided by Equation (13), is the most frequently used measures of the goodness of fit of generalized regression models [38].

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_{f,i} - x_{o,i})^2} \quad (13)$$

Where $x_{f,i}$ and $x_{o,i}$ are the i th expectation and observation, respectively.

III. RESULT AND DISCUSSION

In order to examine aviation accidents in detail, the relationship between the variables can be determined with a

single scatter chart where each point, as shown in Figure 2, represents an existing data. Accordingly, the relationship between each of the flight phases can be seen. Based on these analysis, the structure of each class in accidents allow researchers to create several comments on aircraft, pilots, or weather conditions. Thus, it is possible to go into more detail on the problems in aviation accidents. The plot matrix formed between determined classes for aviation accidents, accident year, the number of accidents, and death is shown in Figure 3.

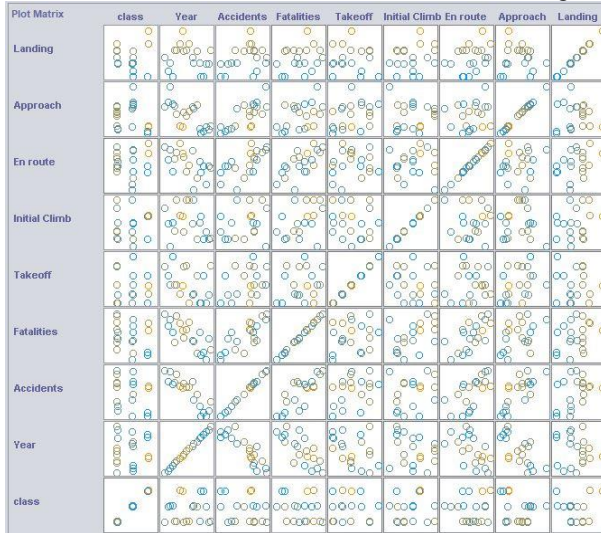


Fig. 3. Plot matrix showing the relationship between variables

After creating the relationship matrix between the variables, the performance values of each algorithm were classified into these variables, and performance values analysed. These values provide important information about the general structure of the algorithm. Precision, recall, F-measure, and Roc area accuracy values of the algorithms are shown in Figure 4.

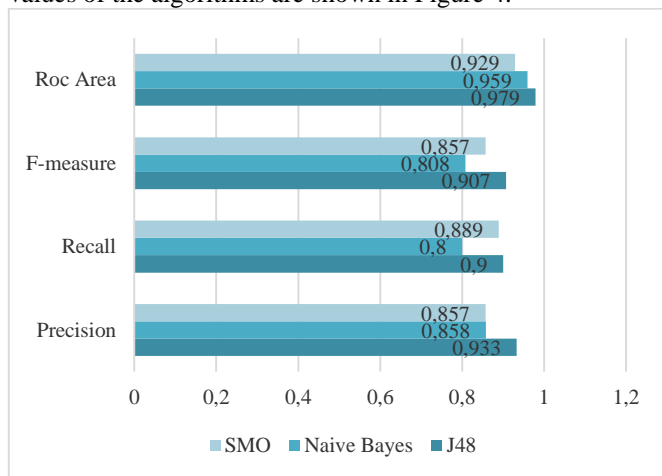


Fig. 4. Accuracy values of the algorithms

Comparison analysis of all 3 algorithms related to the accuracy parameters utilizing WEKA software demonstrated that the algorithm with the highest accuracy in terms of all parameters is the J48 algorithm. Comparing SMO and Naive Bayes algorithms showed that ROC area and precision parameters were higher in Naive Bayes algorithm while f-measure and recall parameters were higher in SMO algorithm. In this case, it is difficult to decide which of the two algorithms

gives better results. For this reason, it is necessary to look at the error values of the algorithms and the kappa statistical value, as shown in Figure 5.

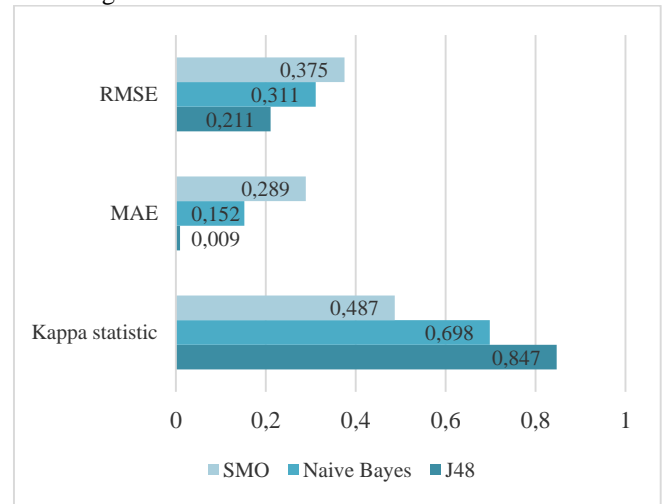


Fig. 5. Error scales of the algorithms

Based on the error scales of the algorithms, it clearly showed that the J48 algorithm has the highest value as in the accuracy values. Due to its high accuracy, the kappa statistical value was also high. On the other hand, the MAE and RMSE error values of the Naive Bayes algorithm were lower than the SMO algorithm and the kappa statistic value was higher than the SMO algorithm. This demonstrates that the Naive Bayes algorithm gives better results. Consequently, it is revealed that the J48 algorithm gives better results compared to other algorithms in the classification process of aviation accidents.

IV. CONCLUSION

In this study, classification of aviation accidents and mortality rate throughout the last 20 years utilizing decision tree (J48), Naive Bayes, and SMO data mining algorithms were carried out. For classification, take-off, first climb, route, approach, and landing stages showing the status of the aircraft were used. The results of this analysis demonstrated that the J48 algorithm provide the best outcomes in terms of both performance analysis and error scales. On the other hand, the Naive Bayes algorithm yielded better results than the SMO algorithm. Lastly, it is recommended that the classification made by the J48 algorithm is taken as a basis for analyzing accidents. This study revealed which of the algorithms gave better results for studies with similar data sets. It is possible to obtain more accurate results in analyzes to be made with this algorithm. This study reveals which algorithm should be preferred to researchers and readers in future analysis studies for similar data sets. In this respect, it provides convenience for future studies. In addition, according to this study data, the higher accident factors for aviation accidents were cruise (en-route) and descend (approach, landing) processes. This situation is thought to be caused by the loss of attention and communication problems between the tower and the pilot due to the fatigue of the pilots, considering the human factors that have the largest share in aircraft accidents. Significant reductions in aircraft

accidents are anticipated, as countries' decision-makers work to address this situation.

REFERENCES

- [1] M. Terzioğlu, Human errors as a cause of aviation accidents can be reduced through crew resource management, Master thesis, Department of Human Resources Dokuz Eylül University, İzmir, Turkey, 2007.
- [2] K. Dönmez, S. Uslu, "A Study on communication induced accidents and incidents in aviation." *The Journal of International Social Research*, vol. 45, 2016, pp. 1074-1079.
- [3] S. Metin, "Human Factors in Aviation Accidents in Last Years. "1. National Aviation Medicine Congress," pp. 22–24, Eskişehir, Turkey, 2014.
- [4] H. Kharoufah, J. Murray, G. Baxter, G. Wild, "A review of human factors causations in commercial air transport accidents and incidents: From to 2000–2016." *Progress in Aerospace Sciences*, vol. 99, 2018, pp. 1-13.
- [5] V. Andersen, T. Bove, "A feasibility study of the use of incidents and accident reports to evaluate effects of team resource management in air traffic control. " *Safety Science*, vol. 35(1-3), 2000, pp. 87-94.
- [6] V. Socha, L. Socha, V. Němec, "Air accidents, their investigation and prevention. " *eXclusive e-Journal*, vol. 4, 2014, pp. 1-9.
- [7] M. Bazargan, V. S. Guzhva, "Impact of gender, age and experience of pilots on general aviation accidents." *Accident Analysis & Prevention*, vol. 43(3), 2011, pp. 962-970.
- [8] W. C. Moon, K. E. Yoo, Y. C. Choi, "Air traffic volume and air traffic control human errors." *Journal of Transportation Technologies*, vol. 1(03), 47, 2011.
- [9] C. V. Oster, J.S. Strong, C. Kurt Zorn, "Investigation of accidents related to air traffic control." 51st Annual Transportation Research Forum 2010, pp. 853–872, United States, 2010.
- [10] S. Uslu, K. Dönmez, "Investigation of accidents related to air traffic control," Mehmet Akif Ersoy University Journal of Social Sciences Institute, vol. 8, 2017, pp: 271–287.
- [11] S. Z. Liang Cheng, R. M. Arnaldo Valdés, V. F. Gómez Comendador, F. J. Sáez Nieto, "Detection of Common Causes between Air Traffic Serious and Major Incidents in Applying the Convolution Operator to Heinrich Pyramid Theory." *Entropy*, vol. 21(12), 2019, 1166.
- [12] W. Kaleta, J. Skorupski, "A fuzzy inference approach to analysis of LPV-200 procedures influence on air traffic safety." *Transportation Research Part C: Emerging Technologies*, vol. 106, 2019, pp. 264-280.
- [13] G. W. Van Es, "A Review of Civil Aviation Accidents Air Traffic Management Related Accidents: 1980-1999." In 4th International Air Traffic Management R&D Seminar New-Mexico, 2001.
- [14] W. K. Lee, "Risk assessment modeling in aviation safety management." *Journal of Air Transport Management*, vol. 12(5), 2006, pp. 267-273.
- [15] M. Lower, J. Magott, J. Skorupski, "Analysis of Air Traffic Incidents using event trees with fuzzy probabilities." *Fuzzy sets and systems*, vol. 293, 2016, pp. 50-79.
- [16] Z. H. Ming, Q. L. Zhong, M. Zhang, S. C. Han, "New analytic calculating method for Performance-Based Navigation airway capacity." *Advances in Transportation Studies*, vol. 1, 2015, pp. 85-90.
- [17] S. H. Stroeve, H. A. Blom, G. B. Bakker, "Systemic accident risk assessment in air traffic by Monte Carlo simulation." *Safety Science*, vol. 47(2), 2009, pp. 238-249.
- [18] P. Brooker, "Air Traffic Management accident risk. Part 1: The limits of realistic modelling." *Safety science*, vol. 44(5), 2006, pp. 419-450.
- [19] H. İncekaş, Increasing Flight Safety and Preventing Accident Crimes Android Based Checklist, Master Thesis, Department of Computer Engineering, İzmir Katip Çelebi University, İzmir, Turkey, 2017.
- [20] "Number of flights performed by the global airline industry from 2004 to 2020." <https://www.statista.com/statistics/564769/airline-industry-number-of-flights/>, accessed 25 June 2020
- [21] "Airline accident statistics. " <https://aviation-safety.net/statistics/period/stats.php>, accessed 25 June 2020
- [22] "Waikato Environment for Knowledge Analysis (WEKA), " <https://tr.wikipedia.org/wiki/Weka>, accessed 18 June 2020
- [23] L. Yuh-Jye, Sequential minimal optimization. Data Science and Machine Intelligence Lab National Chiao Tung University, 2017.
- [24] "Sequential minimal optimization." https://en.wikipedia.org/wiki/Sequential_minimal_optimization, accessed 18 June 2020
- [25] K. Atmaca. "Naive Bayesian algorithm." <https://kenanatmaca.com/naive-bayesian-algoritmasi/>, accessed 3 June 2020
- [26] T. Pala, A.Y. Camurcu, "Design of Decision Support System in the Metastatic Colorectal Cancer Data Set and Its Application." *Balkan Journal of Electrical and Computer Engineering*, vol. 4(1), 2016, pp. 12-16.
- [27] G.S. Eraldemir, M.T. Arslan, E. Yildirim, "Comparison of random forest and j48 decision tree." *Classifiers Using HHT Based Features in EEG*, 2017, pp: 1250–1256.
- [28] A. M. Hormann, "Programs for machine learning." Part II. *Information and Control*, vol. 7(1), 1964, pp. 55-77.
- [29] K. R. Pradeep, N. C. Naveen, "Predictive analysis of diabetes using J48 algorithm of classification techniques." In 2016 2nd International Conference on Contemporary Computing and Informatics. pp. 347-352, 2016.
- [30] S. Aljawameh, M. B. Yassein, M. Aljundi, "An enhanced J48 classification algorithm for the anomaly intrusion detection systems." *Cluster Computing*, vol. 22(5), 2019, pp.10549-10565.
- [31] T. C. Smith, E. Frank, "Introducing machine learning concepts with WEKA." In *Statistical genomics*. pp. 353-378. Humana Press, New York, Usa, 2016.
- [32] R. R. Bouckaert, E. Frank, M. A. Hall, G. Holmes, B. Pfahringer, P. Reutemann, I. H. Witten, WEKA---Experiences with a Java Open-Source Project. *The Journal of Machine Learning Research*, 11, 2010, pp. 2533-2541.
- [33] Z. H. Zhou, "Learnware: on the future of machine learning." *Frontiers Comput. Sci.*, vol. 10(4), 2016, pp. 589-590.
- [34] G. Kayhan, E. Ergün, "Medicinal and Aromatic Plants Identification Using Machine Learning Methods". *Balkan Journal of Electrical and Computer Engineering*, vol. 8(1), 2020, pp. 81-87.
- [35] "What is Root Mean Square Error (RMSE)?" <https://www.statisticshowto.com/probability-and-statistics/regression-analysis/rmse-root-mean-square-error/>, accessed 11 June 2020
- [36] A. Saxena, M. K. Jat, "Analysing performance of SLEUTH model calibration using brute force and genetic algorithm-based methods." *Geocarto International*, vol. 35(3), 2020, pp. 256-279.
- [37] M. Zhang, "Intelligent route scheduling method for non-full-loaded vehicle based on unbalanced data mining." *Advances in Transportation Studies*, 1, 2019, pp. 73-83.
- [38] C. Sammut, G.I. Webb, Mean Absolute Error. *Encyclopedia of Machine Learning and Data Mining*, Springer US, Boston, MA, p. 652. 2010.

BIOGRAPHIES



EMRE KUŞKAN Erzurum, in 1991. He received the B.S. and degrees in civil engineering from the Yıldız Technical University, in 2015 and the M.S. degree in civil engineering from Erzurum Technical University, Erzurum, in 2019. He is currently Ph.D. student at Erzurum Technical University, Erzurum, since 2018 where he works as a research assistant.



MOHAMMAD ALI SAHRAEI received the B.S. in 2006. He received the M.S. and Ph. D. degrees in civil engineering from Universiti Teknologi Malaysia, in 2012 and 2019. He is currently Postdoctoral Researcher at Erzurum Technical University, Erzurum.



MUHAMMED YAŞIN ÇODUR Erzurum, in 1980. He received the B.S. in 2003. He received the M.S. and Ph. D. degrees in civil engineering from Atatürk University, Erzurum, in 2007 and 2012. Currently he is working as an Associate Professor in Erzurum Technical University, Erzurum. His primary research interests include transportation planning, traffic and transportation networks.

Application of Binary Genetic Algorithm for Holographic Vascular Mimicking Phantom Reconstruction

Tugba Ozge Onur and Gulhan Ustabas Kaya

Abstract—Since medical imaging is one of the essential methods for the diagnosis and treatment of several diseases, the characterization and calibration of medical imaging systems with low-cost equipment is the most crucial issue. In this context, tissue-mimicking phantoms have long been used for this purpose. The advantage of phantoms is that, in addition to the desired size and internal properties, they can be produced in a way that best carries the characteristic properties of tissue models and can be standardized so that they can be used in imaging environments. For this reason, it is important to make low-cost phantom designs produced from materials that are easy to shape and available and to ensure that they can be imaged with high quality.


In this study, digital holography and binary genetic algorithm (BGA) were used to reconstruct the images of phantoms that mimic the human vascular system produced at a low cost. The obtained results showed that BGA can be used as an alternative to the reconstruction methods commonly used in digital holography. Since BGA provides an alternative solution to obtain the image with high resolution in the reconstruction process without any image processing algorithm, it enables the diagnosis of diseases related to thin vascular structures in real-time with a reliable and non-contact method.

Index Terms—binary genetic algorithm, lateral shearing, reconstruction, thin vessel structures, tissue-mimicking phantom.


I. INTRODUCTION

PHANTOMS ARE artificial tissues that are used as an alternative to human tissue and that allow the diagnosis of disease to be made non-contact without taking samples from patients. Phantoms are used for many different purposes in the medical sense such as identification of cancerous cells or detection of tissue changes in organs, examination of vascular structures, and so on [1-3].

TUĞBA ÖZGE ONUR, is with Department of Electrical-Electronics Engineering, Zonguldak Bülent Ecevit University, Zonguldak, Turkey, (e-mail: tozge.ozdinc@beun.edu.tr).

 <https://orcid.org/0000-0002-8736-2615>

GÜLHAN USTABAŞ KAYA, is with Department of Electrical-Electronics Engineering, Zonguldak Bülent Ecevit University, Zonguldak, Turkey, (e-mail: gulhan.ustabas@beun.edu.tr).

 <https://orcid.org/0000-0002-5643-0531>

Manuscript received October 05, 2021; accepted Dec 21, 2021.
DOI: [10.17694/bajece.1004981](https://doi.org/10.17694/bajece.1004981)

Phantoms are the test materials that can depict the special characteristics of human tissues in medical imaging applications. Tissue mimicking phantoms have been utilized to qualify and calibrate medical imaging systems for a long time. In addition, phantoms are utilized for comparing the performances of medical imaging systems, training technicians using these systems, and developing new imaging systems and devices. The advantage of phantoms is that, in addition to the desired size and internal properties, they can be produced in a way that best carries the characteristic properties of tissue models and can be standardized so that they can be used in imaging environments. For this reason, it is important to be able to create low-cost phantoms made of materials that are easy to make, shape, and find.

The structure of the produced phantoms differs from each other about the area in which they will be used. For this reason, the materials used in the preparation of the phantom also change. For instance, polyacrylamide (PAA), which has solid elastic and optically transparent properties, was produced with acrylamide C3H5NO and water by Singh *et al.* The temperature fluctuations do not change in PAA [4]. Agar gel is prepared with water-gel structure by changing the temperature. The bonded molecule movement is characterized by agar gel [5]. Carrageenan gel that can be used for magnetic resonance imaging (MRI) phantoms is formed with polysaccharides of seaweed. It is also flexible like phantoms produced with agar gel [6]. The polyvinyl alcohol (PVA) based tissue contains a PVA dissolved in water. It has more structural stability in a longer period than the other tissue-mimicking phantom methods [7]. On the other hand, the phantom models as commercial rigid plastics, polyurethane gel, gelatin-alginate, elastomeric (rubber-like) materials, and so on have been formed so far [1, 8, 9].

Since the purpose of the use of phantoms is usually calibration and testing, different imaging methods are used for their scanning. Dabrowski *et al.* presented a real vessel phantom and scanned it via projection radiography, CT, US, and MR [10]. Vogt *et al.* used MRI, CT, and ultrasound for imaging breast-mimicking tissue phantoms [11]. He *et al.* generated a breast phantom via three-dimensional (3D) printing and they used CT and MRI imaging modalities [12]. Nisar *et al.* presented a vascular phantom and compared their model obtained by carrying out CAD software to the CT scan [13].

The imaging of the above-mentioned different phantom samples with digital holography has a very important place in

terms of resolution and image quality. In addition, the displayed data can be stored in very small areas and reconstructed in three dimensions [14]. The quantitative phase image of the micro sized samples can be captured directly and reconstructed numerically with the digital holographic microscopy technique [15-18]. There have been several types of research employing digital holography microscopy to characterize the biological cells in the current literature. For example, Anand *et al.*, Rappaz *et al.*, FalckMiniotis *et al.* have used digital holography microscopy to fingerprint the red blood cells, tumor cellars, and fibroblasts, respectively [19-21]. In addition, in recent years, it is utilized to recover the images [22]. Although there are various researches in the literature about examining and recovering living cells using digital holography microscopy, the studies about using digital holography microscopy for phantoms are limited. In other respects, common methods such as filtered back projection, iterative reconstruction algorithms, etc. have been employed generally in phantom image reconstruction applications [23]. In this study, we imaged the created phantoms by means of the lateral shearing digital holography technique and performed the reconstruction process with a binary genetic algorithm.

The organization of this paper is as follows: Section II explains the binary genetic algorithm in detail. The preparation of the phantoms and the detailed expression of the optical imaging system is given in Section III. Experimental results are presented in Section IV. Finally, this paper is concluded in Section V.

II. BINARY GENETIC ALGORITHM

The genetic reproduction, evaluation, and selection stages are the basis for GAs [24]. The population is initialized by the initialization step. The selection step enables the selection of the chromosomes that fit better for the formation of the mating pool. In this paper, a roulette wheel scheme is used [25]. The information shared between each pair of mating parents is performed by crossover schemes. The high value of probability is applied for crossover. The gene pairs are changed over a mutation operator and thus allow to obtain small variation for the fitness diversity in the population [26]. In this study, a binary mutation operator has been used for the simulations and the genes are produced randomly in the range of [0, 1]. The used algorithm for the binary mutation is detailed below,

If random number < mutation probability
 ↓
Flip the gen bit value and assign it to chromosome

In the last step, the fitness function is chosen to determine the error between the original and reconstructed images. Various measures can be used in this stage such as Root Mean Squared Error (RMSE), Mean Squared Error (MSE), and Mean Absolute Error (MAE) [27]. The GA operation can be terminated if the optimal solution is converged or the specified number of generations is reached. In this paper,

the MAE fitness function is used to reconstruct the image as defined in Eq. (1),

$$F = (1 + E_M)^{-1} \quad (1)$$

Here: E_M – is the MAE value between the reference and the computed reconstructions of BGA.

Since MAE is selected as the fitness function, the lower fitness function value means that the lowest difference between the reference and reconstructed one has the probability to be selected as parents for ensuing generations. In other words, the chromosomes that have lower misfits are preferred.

The steps of the BGA algorithm used in this study are presented in Algorithm 1.

Algorithm 1

Initialize the population (M, N)
 Evaluate the fitness function using Eq.(1)

While (end condition is not met) **do**
 Select the chromosomes used by [20]
 Crossover the parents and children
 Perform the mutation operator
 Evaluate the fitness function or a specified number of generations

Return the selection

where, M and N represent the numbers of chromosomes and genes, respectively.

In addition to selection, crossover, and mutation genetic operators, there have been some factors such as the size of the population, numbers of genes, bits, and generations, probabilities of crossover, and mutation which are set to make the algorithm work accurately. These parameters have a significant effect on the optimization performance of the algorithm. When the population size is large, the search area is explored more thoroughly with a longer computation time, while early convergence can be seen when the population size is very small. Also, the number of model parameters determines the number of genes. The increment in the number of parameters complicates the function by causing the model space size to be enlarged. Since the accuracy of the solution depends on the number of bits, the complexity and inappropriate values of the parameters make it difficult to explore the accurate solution. The crossover probability determines the new chromosomes to be obtained. Generally, a value between 0.8 and 0.95 is chosen as the crossover probability, since almost all parents are preferred to be replaced by their new chromosomes in the next generation. On the other hand, the mutation probability is often only 0.05% to 1% [28].

In this study, the population includes 100 individuals, and the probabilities of crossover and the mutation are constituted as 0.95 and 0.001, respectively. In addition, the crossover and mutation parameters are chosen as the same as the ones

presented in [24]. An example of these operators is given in Fig. 1.

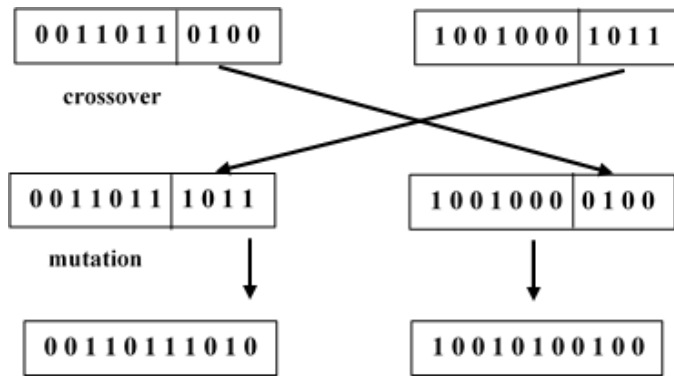


Fig.1. The crossover and mutation operators [25]

III. PHANTOM GENERATION AND IMAGING SETUP

A. Preparation of Phantom Samples

The phantoms used in this study are generated with agar gel [14]. To mimic the human vascular system, grass and white balloon filled with water are used for the generated phantoms. The grasses and balloons are placed randomly in the generated phantoms as presented in Fig. 4a and Fig. 5a, respectively.

B. Lateral Shearing Digital Holographic Microscopy System

In this study, the produced phantom samples showing vascular structures are visualized with a lateral shearing digital holographic microscopy (LSDHM) system [29]. The configuration of LSDHM is shown in Fig. 2.

With the system presented in Fig. 2, a coherent light source is needed to record the information of the object to be imaged (in other words, to create an interference pattern). In this study, this requirement is supplied by means of a He-Ne laser with a wavelength of 632nm. For the beam emerging from the light source to proceed without scattering, a uniform Gaussian distribution is created by first passing it through a spatial filter. It is then collimated with the aid of a lens (100 mm focal length). After the collimated beam runs through the object, it is amplified with a microscopic objective (MO) so that it can be seen more clearly. The magnification of the MO used is 6X (NA=0.75). In addition, transparent phantom samples produced as objects are used.

As the magnified object is reflected on the front and back surfaces of the shearing plate, two sheared object beams are formed. The factor causing this shearing is the thickness of the shear plate used. In this study, a 6mm thick plane ruled reflection grating (PRRG) glass plate was used as the shearing plate. Finally, the hologram (interference pattern) is modeled by the use of two sheared object beams on the image sensor which is a CMOS camera.

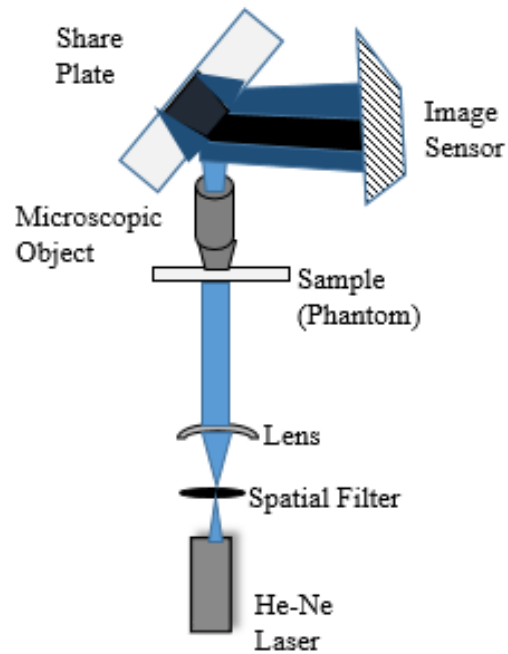


Fig. 2. The configuration of LSDHM

C. Field of View (FOV) and Resolution

Since the Field of View (FOV) is related to the visual angle of a lens, it affects the resolution and size of the captured image by the camera. If the measurement object moves away from the camera, the size of the region of the image increases. Accordingly, the captured individual pixels' mean resolution increase, too. That ensures the users see virtual content in enhanced veracity. Moreover, even though the angle of view is fixed, different sizes of FOV can be obtained when the lens is focused at different working distances. The full angle in terms of degrees related to the width of the sensor on which the lens will be used is defined as the angular field of view (AFOV) [30]. Therefore, AFOV is defined by the focal length of the lens as given in Eq. (2).

$$AFOV = 2 \tan^{-1} \left(\frac{H}{2f} \right) \quad (2)$$

Where: H –is the size of the sensor;
 f – is the focal length.

In addition, the relationship between the FOV and AFOV is defined in Eq. (3),

$$AFOV = 2 \tan^{-1} \left(\frac{FOV}{2xWD} \right) \quad (3)$$

Where: WD –is the working distance.

The sensor size relates to the X and Y pixels configuration. In this work, X and Y pixels are configured at 256x256. On the other hand, the focal length relates to the sensor size, the distance between the camera lens and object (WD), and the

recorded object size (B). The mathematical equation of focal length (f) is given in Eq. (4).

$$f(mm) = \frac{WD}{1 + \frac{B}{0,0014x\sqrt{X^2 + Y^2}}} \quad (4)$$

In this study, the sensor size is calculated as 0.506 mm, when the X and Y pixels are configured a 256x256, and the WD is taken as 80 mm. In addition, the size of the recorded balloon and grass-made phantoms are 5 mm and 0.2 mm respectively. The used MO has 6X magnification with $NA=0.75$. For this reason, the size of phantoms is assumed to be 3mm and 1.2mm. In this context, the calculated values of focal lengths (f) are 7.4mm and 23.75mm for the balloon and grass-made phantoms, respectively. Based on these (f) values, the AFOV values are obtained as 3.92 mm and 1.22 mm, respectively. Moreover, the FOV values are calculated as 5.48 mm and 1.70 mm for the balloon and grass-made phantoms, respectively.

IV. RESULTS AND DISCUSSION

In this paper, two different phantoms are generated by using grass and a balloon to mimic the human vascular system. The simulations are conducted on MATLAB 2018b program to do the performance review of the binary genetic algorithm in digital holographic image reconstruction. The original images of generated phantoms are acquired with the help of the experimental setup given in Fig. 2. After the holograms of the original images are obtained, for the reconstruction of digital holographic images of the generated phantoms, the BGA algorithm given in Algorithm 1 is used and the original phantom images are tried to be obtained via BGA. The values of the BGA parameters of this study are presented in Table I. The algorithm has stopped when the determined maximum iteration number has been attained.

TABLE I
THE USED QUANTITIES FOR BGA ALGORITHM

BGA Parameters	Determined values
Generations	5000
Probability crossover	0.95
Probability mutation	0.001
Elitism rate	20%

The originals, the binarized records, digital holographic, and BGA reconstructed phantom images are shown in Fig. 3 and Fig. 4 for the balloon and grass-made phantoms, respectively. The figures given with b and d in Fig. 3 and 4 correspond to the images of the recorded holograms and binarized reconstructed ones, respectively. The size of the original phantom images is 256x256. The best fit can be obtained by crossover and mutation operators with 5000 epochs and the reconstructed images have almost the same details as the original digital

holographic reconstructed ones as shown in Fig. 3 and Fig. 4. In addition, the fitness of the best elite graphs for the BGA can be seen in Fig. 3e and Fig. 4e with 5000 epochs.

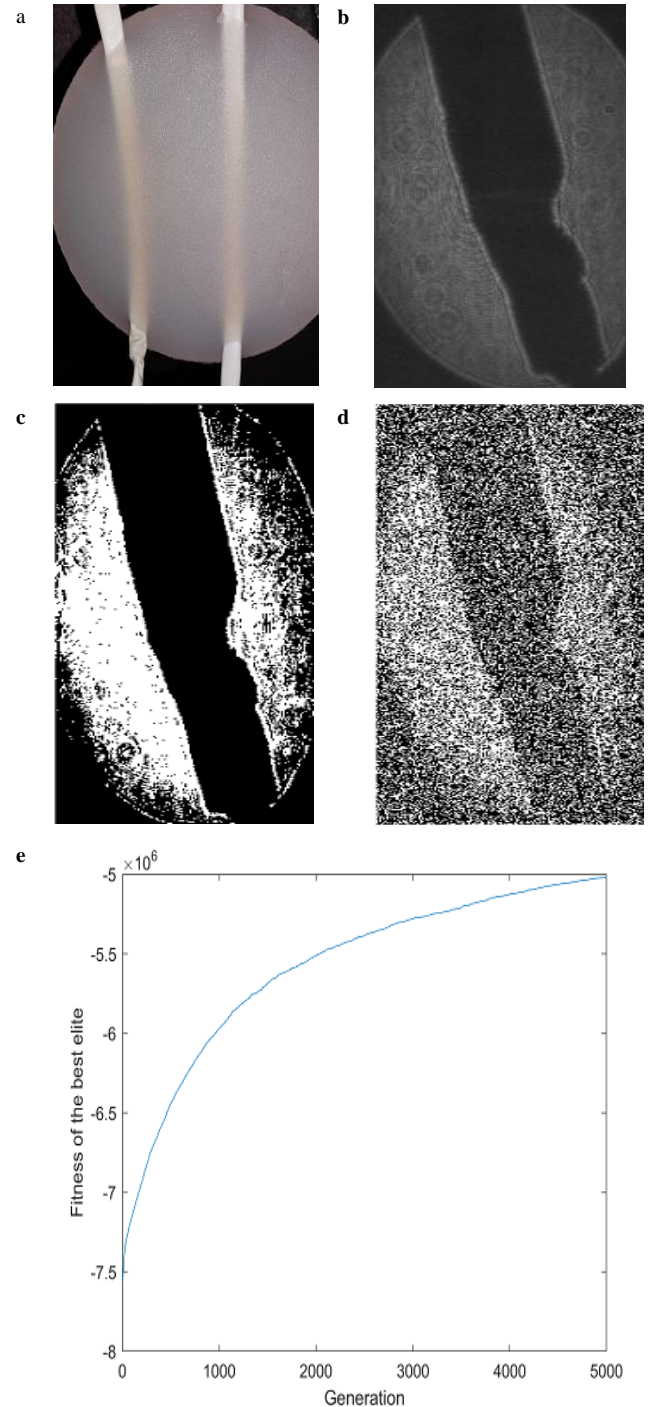


Fig. 3. Phantom with balloon a) Created phantom b) The image of recorded hologram c) Binarized recorded image d) Binarized reconstructed image e) Fitness graph

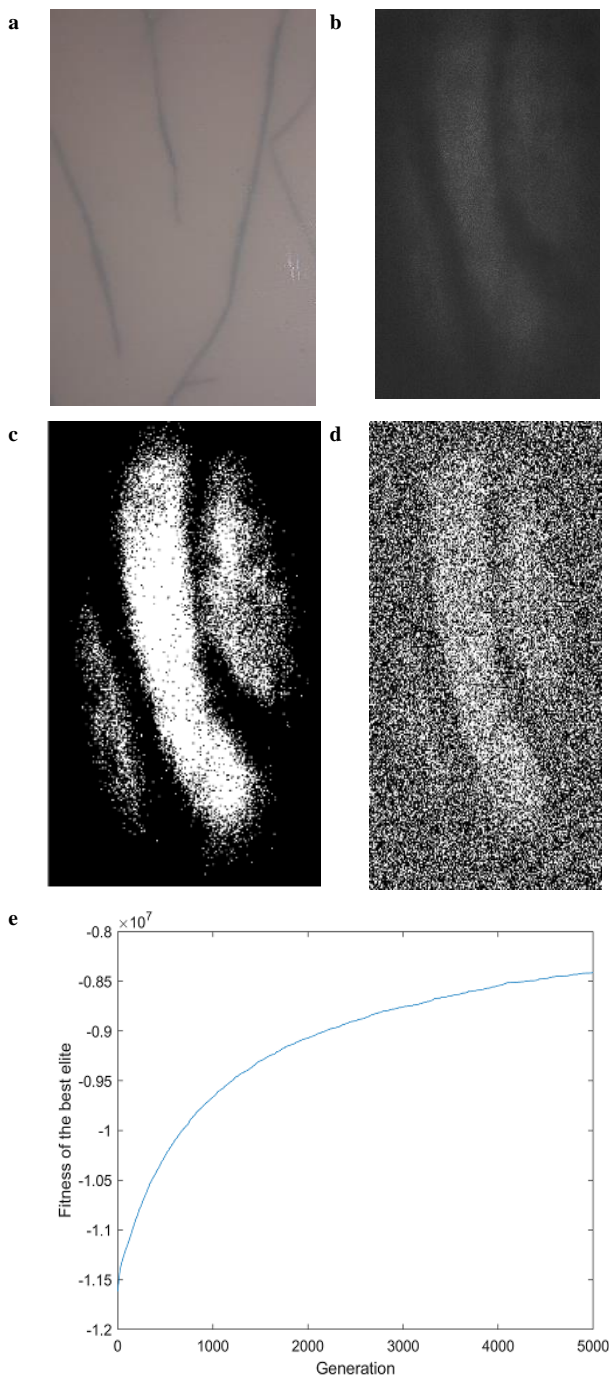


Fig. 4. Phantom with grass a) Created phantom b) The image of recorded hologram c) Binarized recorded image d) Binarized reconstructed image e) Fitness graph

In order to prove that the use of BGA for phantom image reconstruction in the LS-DHM system, our current study is compared with the existing applications in the literature given in Table II. For example, Günel and Kent propose a new approach, which utilizes genetic algorithms, to estimate a cross-sectional image by means of X-ray illumination for the reconstruction process. They also estimate the object parameters with some assumptions and the best results are obtained with the use of fuzzy genetic algorithms [31]. In this study, although the phantom image is reconstructed with GA

and FGA, the used phantom has not been obtained with a holographic system. However, in our proposed system, an artificial phantom is produced with a holographic system (LSDHM) that can be captured very small (in micro and nano-sized) objects, and a noise-free image can be obtained with the genetic algorithm. Uliana *et al.* investigated four methods such as the energy of the signal, integral of the modulus, peak-to-peak, and autocorrelation of the maximum value to estimate the PA signal amplitude. They aimed to form a thermal image by utilizing PA signal amplitude. The PA images of the tissue-mimicking phantom, which are fabricated with a gelatin/agar powder and submerged in a water tank, are used to perform the experiments. They notified that the usage of an evolutionary GA to optimize the parameters of thermal images is improved the error on average by 7.5% [32]. Although Uliana *et al.* obtained phantom images with PA, the system used is quite more complex than the LSDHM system. For this reason, the phantom acquisition is fatiguing. On the other hand, the usage of GA is for estimating the variations in PA signal amplitude, not for the reconstruction of the phantom image. Onur *et al.* proposed the use of the phase-shifting process within the Lateral shearing digital holographic microscopy. They aimed to increase the system stability, adjust the setup alignment and reduce the mechanical and unwanted vibrations. They also imaged the tissue-mimicking phantom manufactured using agar powder and distilled water [14]. Although Onur *et al.* used the LSDHM system for the production of mimicking phantoms for the first time, they used the phase-shifting technique for image reconstruction. However, this technique requires more than one image recording. To eliminate this problem and to reconstruct the images obtained with LSDHM in real-time, the usage of GA is proposed in our current work. Ziemczonok *et al.* performed a study that they visualized biological cell phantoms, which are formed with 3D-printed, via a holographic tomography microscope system. In the reconstruction process, the tomographic reconstruction algorithm is used. This algorithm bases on applying the Fourier diffraction theorem in the Rytov approximation [33]. Magliani *et al.* used genetic algorithms to find the diffusion parameters in several public image datasets. However, they obtained a better set of parameters and higher precision of the retrieval, GA is preferred for optimization not to for the reconstruction process. In addition, the phantom images are not imaged in their study [34]. Devi discussed the different medical imaging techniques by surveying the comprehensive survey. To evaluate the system, they also compared the soft computing methods based on a certain parameter. The performances of the fuzzy approach, machine learning, and genetic algorithm applications are investigated. Although the medical images are investigated, phantom images are not used in these methods [35]. Gouicem applied the Fuzzy penalty (FP) function for GA optimization in the field of image reconstruction. The real and synthetic image datasets are analyzed and these images are reconstructed from a few projections [36].

As mentioned above, there are many studies in the literature to obtain mimicking phantom images. However, in some of these studies, the stages of obtaining images are quite laborious. The systems used are quite complex. Therefore, there is a need

for optimization and parameter estimation using genetic algorithms.

TABLE II
EXISTING STUDIES DEALING WITH RECONSTRUCTION
TECHNIQUES AND PHANTOM IMAGES

Authors	Purpose	Techniques/Methods	Material
Günel and Kent (1998)[31]	To determine the object parameters and to estimate a cross-sectional image from an objects' X-Ray projections.	Reconstruction with genetic and fuzzy genetic algorithms	Head phantom
Uliana <i>et al.</i> (2018)[32]	To form a thermal image by estimating the variations in PA signal amplitude.	Imaging with Photoacoustic (PA)	A tissue-mimicking phantom fabricated using a gelatin and agar powder
Onur <i>et al.</i> (2021) [14]	To image the soft tissue-mimicking phantom for early diagnosis	Phase shifted-lateral shearing digital holographic microscopy	Soft tissue-mimicking phantoms
Ziemczonok <i>et al.</i> (2019)[33]	To visualize the 3D-printed biological cell phantoms by means of a Holographic tomography system	Tomographic reconstruction algorithm	3D-printed biological cell phantoms
Magliani <i>et al.</i> (2019)[34]	To optimize the Diffusion Parameters via genetic algorithms	Not defined	Public image datasets
Devi <i>et al.</i> (2021)[35]	To discuss the various biomedical imaging techniques with a comprehensive survey	Fuzzy logic, artificial neural network, genetic algorithm, machine learning	Several medical images
Gouicem <i>et al.</i> (2012)[36]	To reconstruct the computed tomography image by using the Fuzzy and GA optimization	Image Reconstruction GA, fuzzy inference	Computer Tomography (CT) image (real and synthetic image)

In addition, in some studies, genetic algorithm has been used for image recovery. However, the materials used in these

studies are synthetic data. In order to eliminate these problems, the phantoms are produced with the lateral shearing digital holography technique, which allows the imaging of micro and nano-sized particles. In addition, the usage of a genetic algorithm for the real-time reconstruction of images obtained from LSDHM has been suggested. To the best of the authors' knowledge, the experimentally produced phantoms have not been imaged with an optical system and recovered by a genetic algorithm. Thus, both the non-contact imaging of the vascular mimicking phantom images obtained with a holographic system and the recovery of these images in real-time with the genetic algorithm is provided.

V. CONCLUSIONS

In this paper, we presented the task of using digital holography and binarized genetic algorithm for imaging tissue-mimicking phantoms. The two different structured phantoms made of grass and balloon that are mimicking the human vascular system are evaluated for the experiments. The necessity of phantom usage arises from providing a more realistic imaging environment and ensuring that all kinds of operations are performed without the need for real tissue. Therefore, it is a medically important task to generate either different tissue mimicking structures or visualize or reconstruct them with higher quality. The results obtained from this study have shown that optimal solution can be obtained by using BGA and image reconstruction can be achieved using BGA as well as digital holography.

REFERENCES

- [1] P. Pasyar, S. Masjoodi, Z. Montazeriani, and B. Makkiabadi, "A digital viscoelastic liver phantom for investigation of elastographic measurements." *Computers in Biology and Medicine*, vol. 127, 2020, pp. 104078.
- [2] R. A. O. Jaime, R. H. Q. Basto, B. Lamien, H. R. B., Orlande, S., Eibner, and O. Fudym, "Fabrication methods of phantoms simulating optical and thermal properties." *Procedia Engineering*, vol. 59, 2013, pp. 30 – 36.
- [3] K. A. Abdullah, M. F. McEntee, W. Reed, and P. L. Kench, "Development of an organ specific insert phantom generated using a 3D printer for investigations of cardiac computed tomography protocols." *Journal of Medical Radiation Science*, vol. 65, no. 3, 2018, pp. 175-183.
- [4] S. Singh, and R. Repaka, "Numerical study to establish relationship between coagulation volume and target tip temperature during temperature-controlled radiofrequency ablation." *Electromagnetic Biology and Medicine*, vol. 37, 2018, pp. 13-22.
- [5] T. Kondo, M. Kitatuji, Y. Shikinami, K. Tuta, and H. Kanda, "New tissue mimicking materials for ultrasound phantoms." In *Proceedings of IEEE Ultrasonic Symposium*, vol. 3, 2005, pp. 1664-177.
- [6] E. In, H. Naguib, and M. Haider, "Mechanical stability analysis of carrageenan-based polymer gel for magnetic resonance imaging liver phantom with lesion particles." *Journal of Medical Imaging*, vol. 1, no. 3, 2014, pp. 035502.
- [7] M. Arif, A. Moelker, and T. van Walsum, "Needle tip visibility in 3D ultrasound images." *Cardio Vascular and Interventional Radiology*, vol. 41, 2018, pp. 145-152.
- [8] M. Kavitha, M. Ramasubba Reddy, and S. Suresh, "Modelling, design and development of tissue mimicking phantoms for ultrasound elastography." 2012 IEEE-EMBS Conference on Biomedical Engineering and Sciences, 2012, pp. 564-567.
- [9] Y. H. Kao, O. S. Luddington, S. R. Culleton, R. J. Francis, and J. A. Boucek, "A gelatin liver phantom of suspended 90Y resin microspheres to simulate the physiologic microsphere biodistribution of a

- postradioembolization liver." *Journal of Nuclear Medicine Technology*, vol. 42, no. 4, 2014, pp. 265-273.
- [10] W. Dabrowski, J. Dunmore-Buyze, R.N. Rankin, D.W. Holdsworth, and A.Fenster, "A real vessel phantom for imaging experimentation." *Medical Physics*, vol. 24, no. 5, 1997, pp. 687-693.
- [11] W.C. Vogt, C. Jia, K.A. Wear, B.S. Garra, and T.J. Pfefer, "Phantom-based image quality test methods for photoacoustic imaging systems." *Journal of Biomedical Optics*, vol. 22, no. 9, 2017, pp.095002-1-14.
- [12] Y. He, Y. Liu, B.A. Dyer, J.M. Boone, S. Liu, T. Chen, F. Zheng, Y. Zhu, Y. Sun, Y. Rong, and J. Qui, "3D-printed breast phantom for multi-purpose and multi-modality imaging." *Quantative Imaging In Medicine And Surgery*, vol. 9, no. 1, 2019.
- [13] H. Nisar, J. Moore, R. Piazza, E. Maneas, E.C.S. Chen, and T.M. Peters, "A simple, realistic walled phantom for intravascular and intracardiac applications." *International Journal of Computer Assisted Radiology and Surgery*, vol. 15, 2020, pp. 1513-1523.
- [14] T.O., Onur, G. Ustabas Kaya, and C. Kaya, "Phase shifted-lateral shearing digital holographic microscopy imaging for early diagnosis of cysts in soft tissue-mimicking phantom." *Applied Physics B*, vol. 127, no. 61, 2021.
- [15] A. Anand, P. Vora, S. Mahajan, V. Trivedi, V. Chhaniwal, A. Singh, L. Leitgeb, and B. Javidi, "Compact, common path quantitative phase microscopic techniques for imaging cell dynamics." *Pramana*, vol.82, no. 1, 2014, pp. 71-78.
- [16] A. S. G. Singh, A. Anand, R. A. Leitgeb, and B. Javidi, "Lateral shearing digital holographic imaging of small biological specimens." *Optics Express*, vol. 20, 2012, pp. 23617-23622.
- [17] Y. Park, C. Depeursinge, and G. Popescu, "Quantitative phase imaging in biomedicine." *Nature Photonics*, vol. 12, no. 10, 2018, pp. 578-589.
- [18] S. Devinder, A. Lal, T. R. Dastidar, and S. K. Dubey, "Quantitative analysis of numerically focused red blood cells using subdivided two-beam interference (STBI) based lateral-shearing digital holographic." *arXiv:1909.03454 [eess.IV]*, 2019.
- [19] A. Anand, V. K.Chhaniwal, N. R.Patel, and B. Javidi, "Automatic identification of malaria-infected RBC with digital holographic microscopy using correlation algorithms." *IEEE Photonics Journal*, vol. 4, no. 5, 2012, pp. 1456-1464.
- [20] B. Rappaz, B. Breton, E. Shaffer, and G. Turcatti, "Digital holographic microscopy: a quantitative label-free microscopy technique for phenotypic screening." *Combinatorial Chemistry & High Throughput Screening*, vol. 17, no. 1, 2014, pp. 80-88.
- [21] M. Falck Miniotis, A. Mukwaya, and A. Gjørloff Wingren, "Digital holographic microscopy for non-invasive monitoring of cell cycle arrest in L929 cells." *PLoS One*, vol.9, no. 9, 2014, pp. e106546.
- [22] X. Quan, O. Matoba, and Y. Awatsuji, "Image recovery from defocused 2D fluorescent images in multimodal digital holographic microscopy." *Optics Letters*, vol. 42, 2017, pp. 1796-1799.
- [23] N. Miyaji, K. Miwa, A. Tokiwa, H. Ichikawa, T. Terauchi, M. Koizumi, and M. Onoguchi, "Phantom and clinical evaluation of bone SPECT/CT image reconstruction with xSPECT algorithm." *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 10, no.71, 2020, pp. 2-12.
- [24] D. E. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning*, AddisonWesley, Reading Massachusetts, 1989.
- [25] M. Mitchell, *An Introduction to Genetic Algorithms*, MIT Press, Cambridge, Massachusetts, 1999.
- [26] A. Hussain, and Y. S. Muhammad, "Trade-off between exploration and exploitation with genetic algorithm using a novel selection operator." *Complex & Intelligent Systems*, vol. 6, 2020, pp. 1-14.
- [27] M. K. Ahamad, "Comparative analysis the fitness function of k-means and kernel fisher's discriminant analysis (KFDA) with genetic algorithm." *European Journal of Molecular & Clinical Medicine*, vol. 7, no. 11, 2020, pp. 6231-6241.
- [28] A. Hassanat, K. Almohammadi, E. Alkafaween, E. Abunawas, A. Hammouri, and V. B. S. Prasath, "Choosing mutation and crossover ratios for genetic algorithms- a review with a newdynamic approach." *Information*, vol. 10, no. 390, 2019, pp. 2-36.
- [29] G. Ustabas Kaya, "Imaging of transparent objects with phase shifting-lateral shearing digital holographic microscopy." *IEEE, 2020 28th Signal Processing and Communications Applications Conference (SIU)*, 2020.
- [30] J.E. Greivenkamp, *FieldGuidetoGeometricalOptics*, SPIE Press, Bellingham, WA, 2004.
- [31] T. Günel, and S. Kent, "Genetic approach for the determination of object parameters from X Ray projections." *Turkish Journal of Electrical Engineering and Computer Science*, vol.6, no.3, 1998, pp. 277-286.
- [32] J. H. Uliana, D. R. T. Sampaio, A. A. O. Carneiro, and T. Z. Pavan, "Photoacoustic-based thermal image formation and optimization using an evolutionary genetic algorithm." *Biomed Research International*, vol. 34, no. 2, 2018, pp. 147-156.
- [33] M. Ziemczonok, A. Kuś, P. Wasylczyk, and M. Kujawińska, "3D-printed biological cell phantom for testing 3D quantitative phase imaging systems." *Scientific Reports*, vol. 9, no. 1, 2019, pp. 1-9.
- [34] F. Magliani, L. Sani, S. Cagnoni, and A. Prati, "Genetic algorithms for the optimization of diffusion parameters in content-based image retrieval." *arXiv:1908.06896v1 [cs.CV]* 19 Aug 2019.
- [35] M. Devi, S. Singh, S. Tiwari, S. C. Patel, and M. T. A. Ayana, "Survey of soft computing approaches in biomedical imaging." *Journal of Healthcare Engineering*, vol. 2021, 2021, pp. 1-15.
- [36] A. M. T. Gouicem, K. Benmahammed, R. Draï, M. Yahi, and A. Taleb-ahmed, "Multi-objective GA optimization of fuzzy penalty for image reconstruction from projections in X-ray tomography." *Digital Signal Processing*, vol. 22, no. 3, 2012, pp. 486-496.

BIOGRAPHIES



TUĞBA ÖZGE ONUR is an Assistant Professor in the department of Electrical Electronics Engineering at Zonguldak Bülent Ecevit University. She received her MSc and PhD degrees, with highest honors, in Electrical-Electronics Engineering from Zonguldak Bülent Ecevit University in 2008 and 2016, respectively. Her doctoral thesis study was partly carried out at Luleå University of Technology (Luleå, Sweden) and University of Wisconsin-Madison (Madison, Wisconsin, USA). Her research interests include signal and image processing, ultrasound signal processing, system identification and nonlinear systems. She has published papers in the field of signal processing applications.



GÜLHAN USTABAŞ KAYA is a Research Assistant Dr. in the department of of Electrical-Electronics Engineering at Zonguldak Bülent Ecevit University. She received her MSc and PhD degrees in Electrical-Electronics Engineering from Zonguldak Bülent Ecevit University in 2013 and 2019, respectively. Her interest fields are in the development of novel techniques and applications in digital holography, interferometry, microscopy, photonics, signal and image processing, data mining and machine learning. She has published papers in the field of optics, signal processing and machine learning applications. She is also an OSA member.

An Intelligent Machine Condition Monitoring Model for Servo Systems

Hayri Mutlu, Mustafa Caner Akuner, and Gazi Akgun

Abstract — The installation of industrial servo systems and the determination of control parameters are limited to the skills and knowledge of the commissioner. In addition, commissioned systems are often not re-optimized if environmental influences or loads change. The goal of this research is to create an artificial neural network (ANN) model for servo systems that will keep the servo system's proportional, integral, and derivative (PID) parameters working optimally. For this process, a machine condition monitoring algorithm developed with the ANN technique, which uses the data such as actual current, torque, power, position to be obtained from the servo system on an industrial controller, for the control and rearrangement of the parameters.


Index Terms — Servo System, Artificial Neural Network, PLC, ProfiNET

I. INTRODUCTION


SERVO MOTORS are preferred in the industry because of their fast and precise positioning features [1]. Servo motors are controlled by servo drives that contain many adjustable parameters that will significantly affect their response [2]. The whole formed together with the servo motor and the driver used for its control is described as a servo system. Servo systems are time-varying and non-linear control structures [3] [4].

Servo systems must be installed in accordance with the mechanics and load characteristics they drive. PID controllers are widely used in industrial settings. The basic concept of PID controller is to regulate error, calculated with desired and actual state set to zero.


HAYRI MUTLU, is with Department of Mechatronics Engineering University of Marmara University, Istanbul, Turkey, (e-mail: hayri.mutlu@gmail.com).

 <https://orcid.org/0000-0003-0584-3827>

MUSTAFA CANER AKUNER, is with Department of Mechatronics Engineering University of Marmara University, Istanbul, Turkey, (e-mail: caner.akuner@gmail.com).

 <https://orcid.org/0000-0001-8397-3454>

GAZI AKGUN, is with Department of Mechatronics Engineering University of Marmara University, Istanbul, Turkey, (e-mail: gazi.akgun@marmara.edu.tr).

 <https://orcid.org/0000-0002-8154-5883>

Manuscript received Nov 04, 2021; accepted Dec 17, 2021.
DOI: [10.17694/bajece.1018947](https://doi.org/10.17694/bajece.1018947)

To compensate the system response, the error proportional gain (K_p), as well as derivative (K_d) and integral gains (K_i), are used for calculate the control signal (u_c) as seen in Eq.1.

$$u_c = K_p e + K_d \frac{de}{dt} + K_i \int_0^t e(t) dt \quad (1)$$

One method for calculate control gains is to manually determine the parameters by an expert, and the other is to use auto-tune features in the servo drives if they are supported in the industrial applications. The commissioning of industrial servo systems and the determination of their parameters are limited to the skills and knowledge of the technical personnel. Research for the auto-tune feature has also shown that it does not fully design gain parameters for different servo equipment [2]. It simply assigns relatively safe and rigid parameter values based on the inertia imposed on the motor, which can sacrifice some of the controller performance. Therefore, the gain parameters still need to be manually fine-tuned through an expert to ensure the response meets the needs of the application. In general, there is no specific method of adjusting the gain. It depends on the experience of the specialist and the adjustment procedure is also time consuming [2]. Furthermore, when environmental influences or loads change, commissioned systems are frequently not re-optimized.

It is aimed to develop an adaptive machine condition monitoring algorithm that will enable these systems to work most efficiently under all conditions. Reduce the need for expert personnel in the installation step of the systems, and provide more accurate PID parameters value assignment by detecting the errors.

The adaptive PID controller developed with artificial neural networks instead of traditional PID tuning methods and algorithms produces better results in dynamic response performance in nonlinear systems in the industry, according to simulation and experiments [5], [6]. As a result, the machine condition monitoring algorithm is built using artificial neural networks, which is a more innovative model for evaluating and recalculating the current and speed PID gain parameters in the servo driver.

Current, torque and vibration data are generally preferred for different types of motors to feed artificial neural networks [7] [8]. However, data obtained exclusively from servo drives is

preferred in this study so that the relevant algorithm could be applied to existing devices without the need for a new sensor or equipment. In order to obtain the training data to be used in the training of the ANN, an experimental setup is designed, which will enable it to be obtained via the servo driver and apply different load characteristics. After the obtained data are processed, the training process is completed with the Matlab Neural Net fitting tool and using the Levenberg-Marquardt method [9] [10].

Programmable Logical Controller (PLC) is often used as a real-time controllers of automation systems and applications due to its flexibility, reliability, and relatively low-cost advantage in controlling complex systems [11],[12]. They are primarily used in the control of both machinery and industrial processes as an essential component of industrial automation systems. Rather than running the developed algorithm on a computer or a different controller, it is hoped that it will be integrated into PLCs and applied to existing systems without requiring new hardware. PLCs are not developed to run artificial intelligence software, but it is claimed that PLCs run ANN algorithms and meet the performance expectation [10] [13]. As a result, the developed ANN algorithm has been converted into a program that will be executed by the PLC.

PLC should communicate with the servo driver for acquisition of data to feed ANN algorithm. Also, the cross data to be used to feed the ANN must be acquired with simultaneous and deterministic time intervals. ProfiNet IRT real-time communication protocol is preferred as the protocol to meet this requirement. When compared with similar protocols in motion control applications in the studies examined, it has been observed that it stands out compared to other protocols, especially thanks to its low latency and deterministic structure [14] [15] [16]. In the consideration of all these information, a PLC that supports ProfiNet IRT protocol, and a servo system are recommended as hardware for the intelligent machine condition monitoring algorithm to be implemented for the problem that is aimed to be solved.

Ziegler Nichols Z-N method, which is widely used in the industry, is preferred to adjust the PID controllers to test and compare the performance of the developed system [17].

II. METHOD

The servo system is first commissioned and made operational in this study, after which experiments are carried out with various load characteristics and PID parameters to obtain data. As shown in Fig.1, the obtained data are used in ANN training. The performance of the ANN is tested by experiments obtained with different loads that are not used in the training process. The ANN model is redesigned as a Simulink program. PLC codes are generated from the Simulink program. ANN algorithm has been added to this program on top of data acquisition and parameter writing functions from the servo system. Then, the tests are repeated with the load characteristics not used in the training in the full experimental setup, and the results are compared with the classical methods.

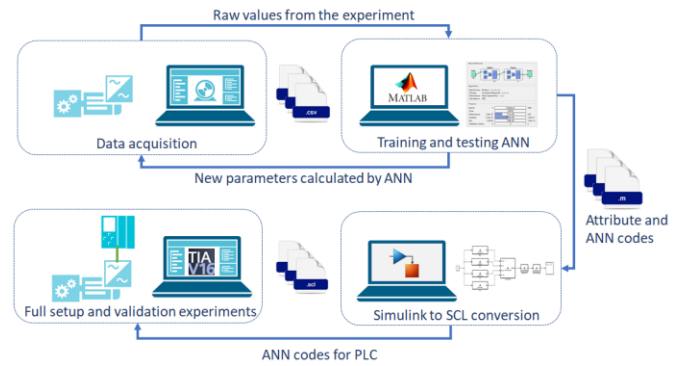


Fig.1. Flow Chart

A. Servo Systems

Servo systems can be defined as the sum of a synchronous motor with a feedback and an electronic power and control units capable of driving this motor in closed-loop servo mode.

Motor drivers consist of two main structures, a control unit that enables the motors to be controlled with different control techniques, and a power unit that will provide the electrical energy that will enable the motor to move.

We can define it as control units for structures that control the outputs of the power unit with v/f, vector, and servo control characteristics, which are mostly preferred in the industry. These control units may have input terminals and/or communication interfaces that allow them to receive motor movement orders directly from their own control unit as well as from an upper control unit.

As seen in the Fig.2, in the servo control type, it consists of 3 control loops, which enable to control the position, speed and current of the motor. The motion of the motor is controlled by cascade PI control scheme in the servo driver we have chosen for this control process.

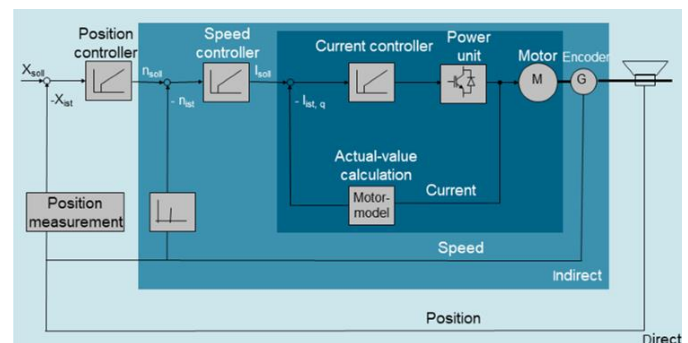


Fig.2. Servo drive control diagram

B. Artificial Neural Networks

Algorithms that imitate the work of the nerves in the brain are called artificial neural networks. These networks consist of artificial neurons connected to each other by synapse-like structures. The connection between each neuron carries information from that neuron to the other neuron. The neuron that receives the information performs the necessary operation and transfers this information to the next neuron [18]. It is developed by Frank ROSENBLATT and an artificial nerve called "perceptron" connects the inputs from the external environment or other cells to the cell with weights. Then It

calculates the sum of the weighted inputs and the bias parameter, which is the criterion of how easily the cell can be stimulated. The output of the artificial nerve is calculated by activation function with v_k , sum of the weighted input. [19].

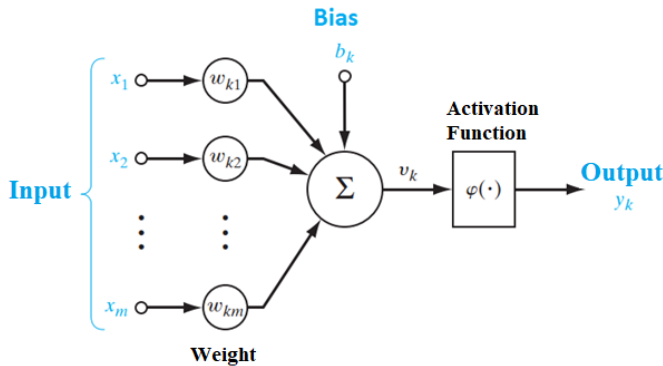


Fig.3. Perceptron model

1) Training Artificial Neural Networks

The process of determining the weight values and bias coefficients in the structure of each neural network that makes up the ANN is called training the ANN. Calculating the local minimum point in the cost function of minimized error between actual value and input data set is required for this process. There are several different techniques used in calculating. Some of the techniques we can use include Levenberg-Marquardt (LM), Bayesian Regularization, and Scaled Conjugate Gradient. The LM algorithm, which is another type of Newton's algorithm, is used between ANN and feedforward back propagation algorithms. The LM algorithm is an advantageous one in that it converges quickly, contains few parameters, and operates only with first-order partial derivatives [20]. The LM algorithm is defined in the Eq.2 below.

$$\Delta w(k) = -(J_k^T J_k + \mu_k I)^{-1} J_k^T e_k \quad (2)$$

$$W(k+1) = W(k) + \Delta w(k)$$

Where W is the weight vector, I is the unit matrix, μ is the combination coefficient. As stated in Eq. 4, the error matrix and Eq. 3 represent the detailed Jacobian matrix and consist of first order derivatives of the error matrix according to the weights [20].

$$J = (PxM)xN \quad (3)$$

Where P is the number of training samples, M is the number of outputs, and N is the number of weights (the number of hidden layer neurons).

$$e = (PxM)xI \quad (4)$$

Where P is the number of training samples, M is the number of outputs, and I is the unit matrix and the e is error vector

2) Creation of the Data Set

A data set is the labeled and ready-to-use data obtained from the real world or simulations to train ANN. In addition, the values that show some of the specially defined features of the data set, which are mostly used as vectors, are called the features of the data [21].

As we mentioned before, the experimental data is exported from Starter, which is the Siemens S120 servo drive commissioning software, is labeled and transferred to MATLAB. In addition, the normalization of the experimental data is performed. After this process, a MATLAB function is used again to generate the features of the experimental data. In this part root mean square (RMS), integrated absolute value (IAV), mean absolute value (MAV) functions of the data, which are seen in equations 5 to 8, respectively, are used.

$$x_{rms} = \sqrt{\frac{\sum_{i=1}^n x_i^2}{n}} \quad (5)$$

$$x_{wl} = \sum_{i=1}^n |x_i - x_{i-1}| \quad (6)$$

$$x_{iav} = \sum_{i=1}^n |x_i| \quad (7)$$

$$x_{mav} = \frac{\sum_{i=1}^n |x_i|}{n} \quad (8)$$

In Fig.4, the data obtained from the experiment performed after adjusting the PID values in the no-load condition of the servo system are shown graphically.

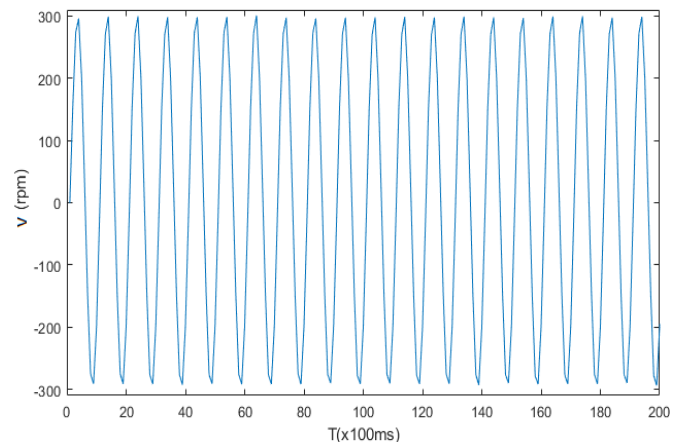


Fig.4. Example of raw speed data read from the drive

The raw data obtained are normalized before being used to train the ANN, and then the RMS values are calculated by shifting 32 samples in a sampling window consisting of 256 values. Windowing functions are functions that multiply signal segments. With these functions, it is ensured that the center parts of the signal window parts are highlighted. The graph of this

feature data used for training of ANN is also shown in Fig.5.

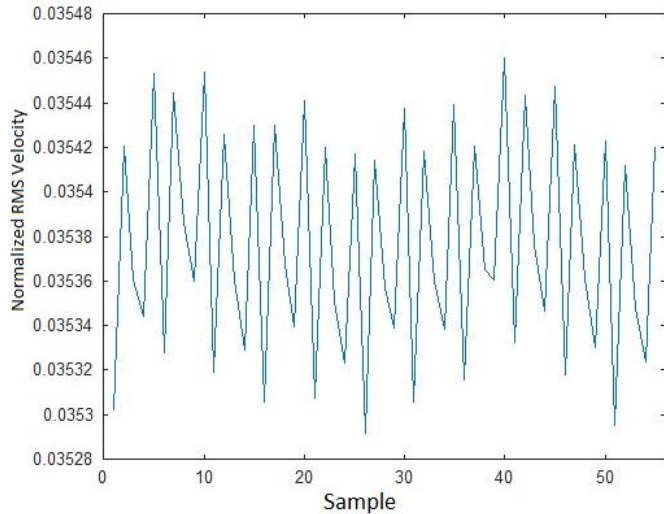


Fig.5. Example of RMS attribute calculated velocity data

3) Creating ANNs with MATLAB

In the training process of ANN, Neural Net Fitting tool integrated into MATLAB is used to solve the data fitting problem by using a two-layer feed-forward network. This tool is an application component that allows us to separate the training data, validation data and test data, train according to the selected algorithm, validate the training, define the network architecture, and train the network. In addition, this tool produces many outputs that allows us to evaluate training performance, such as histogram of errors and regression analysis. Finally, it can generate it as a MATLAB function in the completed ANN algorithm. The ability to produce an integrated solution in ANN training has enabled us to prefer this tool in ANN training. With the regression, performance analysis and experiments using histograms, which can be seen in Fig.6 , the training validation and test data rates and the number of hidden layers has been optimized.

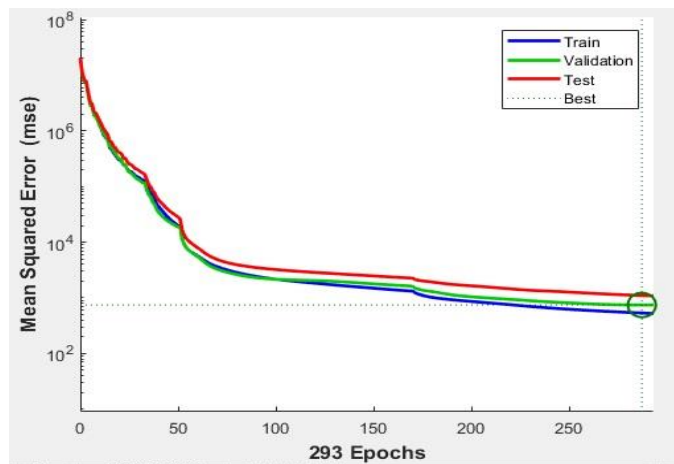


Fig.6. Training performance graph

Furthermore, the generated ANN is tested and verified using a variety of experiments that are not included in the training data.

4) Generating Simulink Code

To run the obtained ANN model on the PLC, the algorithms used in MATLAB/Simulink must be converted to a coding language supported by PLCs for feature calculation, creation of the data array that will feed the ANN, and singularization of the result. For this process, the PLC code generator function available in Simulink is used. Then, MATLAB functions are transferred to the Simulink and the block structure as seen in Fig.7 is established.

In this structure, X refers to the data taken from the real world. In the second part, the feature calculation is made and the following *ConcatMatrix* is combined with the ANN feed data array with the calculated features. In the third operation, ANN is run and recalculated PID values are found, and these data are sent to *OutConasrator* block to make a single data group to be uploaded to the Servo system.

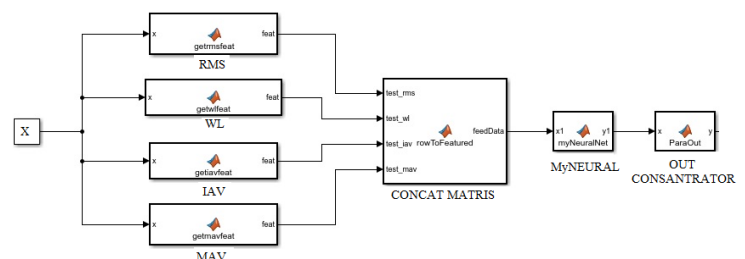


Fig.7. Simulink program

Then, each function is converted one by one into the SCL language specified in PLC programming standards.

C. Industrial Control Systems

PLCs, one of the control systems used in industrial environments, and ProfiNet, one of the communication protocols developed to meet industry needs, are preferred as stated in the introduction of this study.

1) PLC

Programmable logic controllers are devices designed to operate in a wide temperature band, in industrial environments with vibration and electromagnetic noise, and have a wide range of uses from machine automation to process automation due to their stable working structure. The two main parameters that determine the cycle rate of these devices are the program memory unit size and one command processing speed. These parameters are designed to meet the needs of the specific area where PLCs will be used in industry. There are compact models with input and output terminal units, as well as fully modular PLCs.

In this study, Siemens S7-1500 series product is preferred. This device is a modular PLC, designed for the control of mid-to-high-end machines with 300KB programming memory and 40 ns bit-instruction speed. The MATLAB Simulink PLC code generator tool supports this device. It also supports the ProfiNet IRT feature without requiring any additional hardware.

These controllers are used in industry generally with a single core CPU. PLCs with this design cannot perform multitasking at the single time. Predefined interrupts are used in operating systems when this type of multitasking is desired. In cases where

it is desired to perform an operation with simultaneous time intervals, PLCs’ that designed with cyclic interrupts can be chosen.

2) ProfiNet

ProfNet is a real-time communication protocol that uses a standard Ethernet interface and communicates at Level 2 from the OSI layers. ProfNet can also be defined as a set of protocols consisting of different sub-protocols customized according to application needs. ProfEnergy, ProfiSafe, Profinet IO can be given as an example.

ProfNet IRT, on the other hand, is defined as an advanced type of real-time communication protocol that guarantees communication data to be sent and received within a predetermined synchronization time. This customized protocol is preferred especially in motion control operations because of its speed and prevention of data loss. The most important thing to remember about using this protocol is that it requires ProfNet IRT support on both the controller and the devices to be controlled, as well as all network elements that connect these two types.

In this study, the IRT synchronization time is determined as “1 ms” and the configuration is made as specified in the network settings of the PLC, to ensure that the number of devices we communicate with is small and to provide the best possible speed of communication.

The ProfNet Siemens standard telegram 111, which will be used in communication with the servo driver, is used to send motion orders to the motor and to upload the calculated PID parameters to the driver. However, to transfer the actual measurement values required to feed the ANN to the PLC, a user-defined 8-word data has been added to the tail of the transmit section of this telegram. Fig.8 shows the telegram configuration made in the Starter software.

Telegram 111	PROfinet e				
		r2089[0] : CO: Send binector	ZSW1	0000	hex
		r2089[3] : CO: Send binector	POS_ZSW1	0000	hex
		r2089[4] : CO: Send binector	POS_ZSW2	0000	hex
		r2089[1] : CO: Send binector	ZSW2	0000	hex
		r2089[2] : CO: Send binector	MELDW	0000	hex
		r2521[0] : CO: LR position ac	XIST_A	0000_0000	hex
		r63 : CO: Actual speed smot	NIST_B	0000_0000	hex
		r2131 : CO: Actual fault code	FAULT_CODE	0000	hex
		r2132 : CO: Actual alarm coc	WARN_CODE	0000	hex
		r68 : CO: Absolute current ac	user	0000	hex
		r61[0] : CO: Actual speed un	user-defined	0000	hex
		r70 : CO: Actual DC link volte	user-defined	0000	hex
		r72 : CO: Output voltaqe	user-defined	0000	hex
		r80 : CO: Torque actual valu	user-defined	0000	hex
		r82[0] : CO: Active power act	user-defined	0000	hex
		r84 : CO: Flux actual value	user-defined	0000	hex
		r482[0] : CO: Encoder actual	user-defined	0000_0000	hex

Fig.8. ProfiNet telegram configuration

III. EXPERIMENTAL SETUP

An experimental setup has been designed in which different load characteristics can be applied to train the proposed ANN. For this purpose two screws are added to a composite wheel that we attached to the motor shaft, and various weights are placed on these screws, and experiments are conducted. The experiments are started with the condition where the motor is completely unloaded, respectively. The motor tune process has

been completed under these conditions. After this process, data record for 3.3 seconds with an input signal as sinusoidal velocity-position function depending on the time that we had previously determined with the same physical conditions. In this data, 8 values to be used to train the ANN are cross sampled with 100 millisecond time intervals. After this step, the related experiment is repeated by changing the speed and current PID parameters, respectively. Then the experiments are continued with a balanced load placement. The same test procedure is applied and repeated for the new physical conditions. Next step, an unbalanced load is created by attaching a single weight, and the relevant samples are taken by repeating the same test procedures. Finally, to create a variable momentum, the load is placed on the screw with a sliding motion, and the experimental procedures are repeated, and the data to be train ANN is collected and recorded. In addition, to better feed the ANN, the tuned values obtained in the experiments are used in cross conditions, and training data are obtained. After obtaining the relevant data, the ANN model is trained.

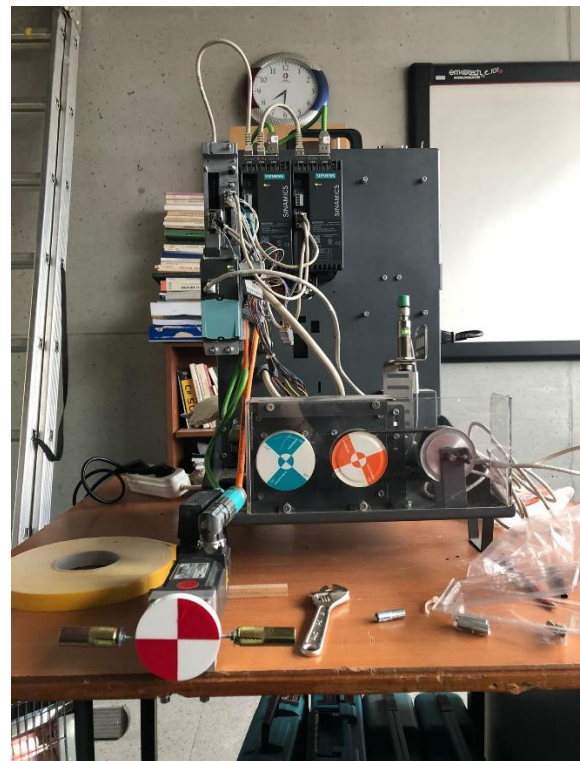


Fig.9. Experimental setup for ANN training

As seen in the flow chart in Fig.10, during the training of the ANN, the experiments are repeated by changing the test weight to create an independent data for the tests, both by attaching two weights for the balanced load and by attaching a single weight for the unbalanced load. The values obtained from the ANN are compared with these values, and the performance of the ANN is decided, and the most suitable ANN is decided by testing the experiments in a real experimental setup with reduced weight for the 3 most suitable ANN designs. The traction current of the motor and the tracking error in the speed data are used and the

most successful ANN model parameters are embedded to the industrial controller.

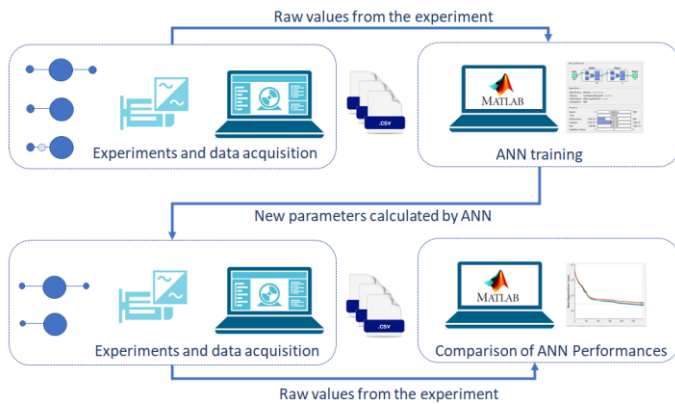


Fig.10. Flow chart for ANN training

Finally, in our experimental setup, the servo system is arranged to communicate with ProfiNet IRT via PLC to test the proposed system. The PLC program is rearranged to send the motion command to the servo system. In addition, the actual measurement values that ANN will use are read and the results produced by ANN are write to the relevant PID parameters.



Fig.11. Components of experimental setup

Samples are taken in full experimental setup with a lighter and balanced loaded weight that is not used in ANN trainings in the newly arranged experimental set. The results of the experiments are compared in the discussion, which is the next section of this study.

IV. DISCUSSIONS

Experiments are repeated using Z-N tuning techniques, which are commonly used in the industry, to compare the results of the previous experiments with the results of the most recent experimental setup. First and foremost, it should be noted that the Z-N technique is an experimental technique that begins with observational data and is then constrained by the knowledge and abilities of the person who commissioned the system for optimization. As a result, following the Z-N process, the current and speed values are monitored online, the driver is re-adjusted using manual optimization techniques, and the experiment is

repeated. As shown in Figure 12, three different results are obtained, with which we can compare the data. First, we look at the feedback data obtained with the speed control set value and the mean squared error (MSE) data generated in the three systems. As seen in Figure 12 and Table 1, the follow-up errors in the values obtained with the ANN technique as 2010 which is much less than Z-N techniques. This shows that with these values calculated by ANN and their continuous control and regulation, better results can be obtained completely independent of the capabilities of the person who commissioned them.

TABLE I
SPEED MSE COMPARISON TABLE

	Z-N	Optimized	ANN
MSE Speed	171978	23549	2010

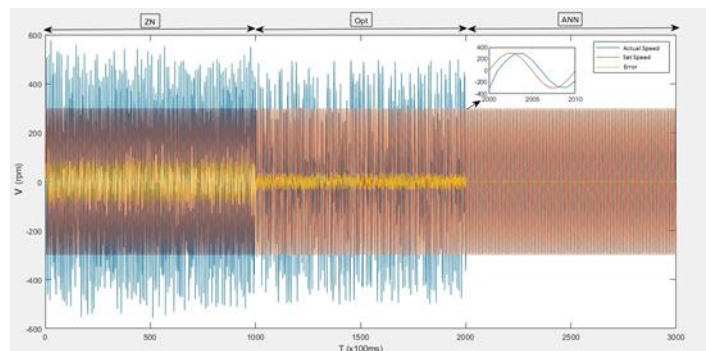


Fig.12. Velocity-Time comparison trend

Another control data is the current of the motor. In the motor driven by using the parameters obtained with the Z-N technique, the motor current was measured as 3.405 amperes on average. By using manual optimization techniques, the motor current has been reduced to an average of 2.952 amps. In the experiment performed using the parameters obtained from the ANN technique, the motor current was measured as 0.048 ampere on average. As seen in Figure 13, current is much less and stable in the same motion characteristic. We can say that the motors controlled by the values calculated with the ANN work more efficiently by looking at the results of the relevant experiments.

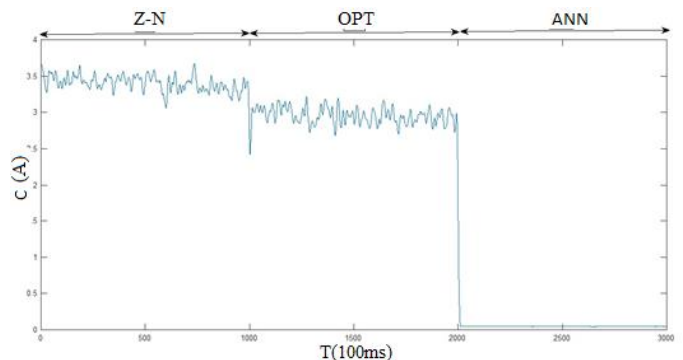


Fig.13. Current-Time comparison trend

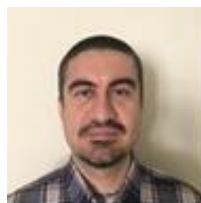
V. CONCLUSION

Given these facts, we can conclude that the ANN technique can be used to calculate the PID coefficients of servo systems and that these ANN algorithms can be successfully implemented within certain limits in the control objects known as PLCs. Furthermore, one of the most important features of this system is that it collects data from the system at sampling intervals of every 3.3 seconds and updates the relevant parameters with the new value calculated by the ANN algorithm based on the new situation it observes. It demonstrates that the system has an adaptive structure, and that the proposed intelligent method can be used to determine the optimal control parameters, especially in systems where power and torque demand can change over time.

REFERENCES

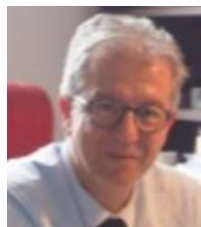
- [1] «Jaen-Cuellar, A. Y., de J. Romero-Troncoso, R., Morales-Velazquez, L., & Osornio-Rios, R. A. (2013). PID-controller tuning optimization with genetic algorithms in servo systems. *International Journal of Advanced Robotic Systems*, 10(9), 324.»
- [2] «Chen, C. W., Chang, L. K., Liao, Y. T., Chung, C. H., Su, W. C., Chen, K. S., & Tsai, M. C. (2020, November). Tuning of Servo Drive Controller Based on Boosted Tree Model and Particle Swarm Optimization. In 2020 23rd International Conference on Electricals.»
- [3] «Lin, Q. S., Yao, Y. F., & Wang, J. X. (2010, November). Simulation and application of neural network PID auto-tuning controller in servo-system. In 2010 2nd International Workshop on Database Technology and Applications (pp. 1-4). IEEE.»
- [4] Firoozian, R. (2014). *Servo motors and industrial control theory*. Springer..
- [5] «Jingjing, X., & Jiaoyu, L. (2013, May). Neural network PID controller auto-tuning design and application. In 2013 25th Chinese Control and Decision Conference (CCDC) (pp. 1370-1375). IEEE.»
- [6] «Aftab, M. S., & Shafiq, M. (2015, February). Adaptive PID controller based on Lyapunov function neural network for time delay temperature control. In 2015 IEEE 8th GCC Conference & Exhibition (pp. 1-6). IEEE.»
- [7] «Kumar, S., Mukherjee, D., Guchhait, P. K., Banerjee, R., Srivastava, A. K., Vishwakarma, D. N., & Saket, R. K. (2019). A comprehensive review of condition based prognostic maintenance (CBPM) for induction motor. *Ieee Access*, 7, 90690-90704.»
- [8] «Samhour, M., Al-Ghandour, A., Ali, S. A., Hinti, I., & Massad, W. (2009). An intelligent machine condition monitoring system using time-based analysis: neuro-fuzzy versus neural network. *Jordan Journal of Mechanical and Industrial Engineering*, 3(4), 294-».
- [9] Watson, G. A. (Ed.). (2006). *Numerical analysis: proceedings of the Biennial Conference held at Dundee, June 28-July 1, 1977 (Vol. 630)*. Springer..
- [10] «Duymazlar, O., Engin, M., & Engin, D. (2020, June). Embedded Artificial Neural Network on PLCs to Predict Nonlinear System Responses. In 2020 9th Mediterranean Conference on Embedded Computing (MECO) (pp. 1-4). IEEE.»
- [11] «Jung, I. S., Mulman, B. M., Thapa, D., Koo, L. J., Bae, J. H., Hong, S. H., ... & Wang, G. N. (2008, October). PLC control logic error monitoring and prediction using Neural Network. In 2008 Fourth International Conference on Natural Computation (Vol. 2,».
- [12] «Canedo, A., Ludwig, H., & Al Faruque, M. A. (2014). High communication throughput and low scan cycle time with multi/many-core programmable logic controllers. *IEEE Embedded Systems Letters*, 6(2), 21-24.»
- [13] «Li, J., & Gómez-Espinosa, A. (2018, November). Improving PID control based on neural network. In 2018 International Conference on Mechatronics, Electronics and Automotive Engineering (ICMEAE) (pp. 186-191). IEEE.»
- [14] «Dias, A. L., Sestito, G. S., & Brandao, D. (2017). Performance analysis of profibus dp and profinet in a motion control application. *Journal of Control, Automation and Electrical Systems*, 28(1), 86-93.»
- [15] «Fontanelli, D., Macii, D., Rinaldi, S., Ferrari, P., & Flammini, A. (2013, May). Performance analysis of a clock state estimator for PROFINET IO IRT synchronization. In 2013 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)».
- [16] «Wu, X., Xie, L., & Lim, F. (2014, October). Network delay analysis of EtherCAT and PROFINET IRT protocols. In IECON 2014-40th Annual Conference of the IEEE Industrial Electronics Society (pp. 2597-2603). IEEE.»
- [17] «Liu, L., Shan, L., Yan, J., Liu, C., & Dai, Y. (2018, June). An improved BFO algorithm for optimising the PID parameters of servo system. In 2018 Chinese Control And Decision Conference (CCDC) (pp. 3831-3836). IEEE.»
- [18] «Alan, M. (2020) Biosignal classification and disease prediction with deep learning. Master Thesis, Marmara Universities institute for graduate studies in pure and applied sciences, Istanbul, Turkey.»
- [19] Haykin, S. S. (2009). *Neural networks and learning machines/Simon Haykin..*
- [20] «OKKAN, U., SERBEŞ, Z. A., & GEDİK, N. (2018). MATLAB ile Levenberg-Marquardt algoritması tabanlı YSA uygulaması: Aylık yağış-akış modellenmesi. *Dicle Üniversitesi Mühendislik Fakültesi Mühendislik Dergisi*, 9(1), 351-362.»
- [21] «Akgun, G. (2015) Data driven predictive control of exoskeleton for hand rehabilitation. Master Thesis, Marmara Universities institute for graduate studies in pure and applied sciences, Istanbul, Turkey.»

BIOGRAPHIES



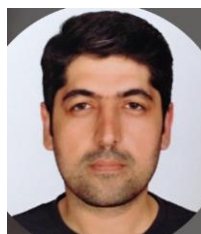
Siemens Turkey.

HAYRI MUTLU is born in 1983, in Haskova, Bulgaria. He received his BSc. in Mechatronics Education from Marmara University in 2007. Currently, he is continuing his MSc. study at Marmara University. He instructs training about PLC, HMI, and Industrial Networking at



University.

MUSTAFA CANER AKUNER is born in 1968, in Samsun, Turkey. He received his BSc. in Electrical Education from Marmara University in 1986. His M.S. and Ph.D. degrees from Marmara University 1993 and 1999, respectively. He is currently head of Mechatronics Engineering department in Marmara



GAZI AKGUN is born in 1982, in Denizli, Turkey. He received his BSc. in Electrical Education from Dicle University in 2004. His M.S. and Ph.D. degrees from Marmara University 2015 and 2019, respectively. He is currently assistant professor at Mechatronics Engineering department in Marmara University.

Broadband Low Reflection Surfaces with Silicon Nano-pillar Square Arrays for Energy Harvesting

Turgut Tut

Abstract—In this work, optimization of the nanopillar arrays and thin films coated on silicon substrate has been investigated in order to minimize the optical reflection loss from the silicon substrate surface. Nano-pillars' height, incline angle, array properties are systematically optimized. Full field Finite Difference Time Domain method is used to simulate EM fields and calculate the reflection loss from the modified nanostructured substrate surfaces in 400nm-1100nm spectral range. Simulation optimization recipe for the nanopillar structure is clearly presented and it is not only useful for square arrays but for regular arrays of nano-pillars in general.

Index Terms—Nano-pillars, quantum efficiency, anti-reflection, thin films, nanostructures, truncated nano-cones, solar cells, solar energy.

I. INTRODUCTION

GENERALLY CRYSTALLINE silicon based photonic devices such as photodiodes, solar cells, phototransistors need low surface reflectance over a wide spectrum of light in order to have high external quantum efficiency and energy harvesting through outer circuits. One of the conventional methods to reduce surface reflection is to use single layer dielectric at specific wavelength but this does not reduce reflectivity for broadband. SiO_2 , SiN_x , TiO_2 , Al_2O_3 dielectric materials are the most popular thin film materials for antireflection coatings. When double dielectric layers are used for this purpose, there are two reflection minimums in the reflection spectrum. Unfortunately, in order to achieve a broadband low reflection, multi-layer band pass filter should be used, but this requires many thin film layers, and this increases the production costs. However, this is not an optimum solution to the broadband low reflection problem. Surface modification is needed in order to obtain broadband low reflection from the optoelectronic substrate surfaces. This modification can be achieved by using micro [1] or nano-size [2-12] structures over the device surfaces. In order to build such small features, researchers used mainly wet etching and dry etching methods. For example, industrial pyramidal surface texturing for crystalline silicon uses anisotropic wet-

etching method. In other methods, dry etching can be used with ionized gases in the plasma vacuum chambers. For some studies both etching methods are used in combination.

Some researchers studied some degree of randomness [9-13] in the periodic structures to decrease the reflection further. However, for large area applications this can have some advantages as well as disadvantages as to have some standard way of production.


Structure sizes show differences in these applications. Some applications use micrometer size pillars [1] and some applications use nano-meter size pillars. For applications that need thin silicon film layers, nanometer size structures have to be used in order to achieve low reflection in broadband. If the pillars are tall (more than micron) then this can lead to quantum efficiency loss due to charge recombination in the pillar structure and the generated electron hole pairs cannot be extracted from the devices for energy harvesting. Therefore, the height of the pillars should be optimized to get low reflectivity. When the height of the pillars is too short then the coupling of incident light into the silicon substrate is low.

To summarize, recently several types of surface nanostructures have been investigated. Moth-eye nanopyramidal pillars which use graded index refractive index to reduce reflection. Nanowires and nanocylinders that use Mie resonances (scattering) which result in overall decrease in reflection. Metallic nanoparticles use their plasmonic effect to direct electromagnetic energy through the substrate effectively and decrease the reflection from the surface.

Optimization studies of the nanostructured surfaces in optoelectronic devices are still popular since the energy sector is very crucial for the economy of the countries. Even one percent increase in energy efficiency in solar cells means millions of dollars saving in countries' economies. In order to obtain high efficiency in solar cells, low reflection from the surfaces irrespective of incident angles in wide wavelength spectrum is needed.

In this work, we optimized nano-pillars filling ratio, pillar height, apex angle, and calculated the weighted average reflection with respect to wavelength to get the overall efficiency improvement for the optoelectronic devices. We focused on nano-pillar arrays of square type.

TURGUT TUT, is with Department of Nanotechnology Engineering, Abdullah Gul University, Kayseri, Turkey, (e-mail: turgut.tut@agu.edu.tr).

 <https://orcid.org/0000-0002-4589-201X>

Manuscript received May 26, 2021; accepted Jan 27, 2022.
DOI: [10.17694/bajece.943854](https://doi.org/10.17694/bajece.943854)

II. METHODS AND SIMULATION RESULTS

In this letter, we present the optimization of the nano-pillar square arrays in order to achieve broadband low reflection surfaces. When compared to a bare planar crystalline silicon wafer surface, the modified surface has significant reduction in reflection in 400nm-1100nm range.

We used Finite Difference Time Domain (FDTD) method to simulate photonic nanostructures. We placed monitors just below the nano-pillars to calculate the transmission to the bulk substrate. We also placed a monitor above the structure to simulate the EM fields and calculate the reflected power. We also calculate the light absorbed in the nanostructure. The optical constants of crystalline structure and thin films are taken from [14]. We first optimized the filling ratio of the nano-pillar arrays, then height to minimize the reflection. We also varied the pillar wall angle to create truncated cones to search the effect of cone structure to the reflection properties. We used dielectric anti-reflection coatings with SiO₂ thin film to decrease the reflection further. These nano-cone truncated pillar structures can be fabricated using e-beam lithography and dry etching techniques. Due to the nature of etching processes, the vertical pillars can be obtained under special physical and chemical conditions [17, 18]. It is important to have vertical pillars to get optimum reflection from such structures since as the pillar angle gets lower, the reflection start to increase Fig.7. For large area applications, nano-imprint lithography can be used with a previously prepared template using nano-fabrication methods. We also realized that to increase the efficiency of photonic devices, we have to increase the optical power absorbed in the bulk of the active silicon layer. The absorbed light in the nanostructures part could be lost due to recombination and cannot contribute to the photocurrent with large percentage. Therefore, we have to maximize the absorption not in the nanostructures but the optical absorption in the bulk silicon active device region.

Solar irradiance differs with wavelength. We can use the total reflection of light from the surface weighted with solar irradiance. This way the performance of the low reflection loss performance can be quantified better. ASTM Air Mass 1.5 direct solar irradiance is used for the calculation. In this formulation, irradiance is multiplied with the wavelength and the integral is taken with wavelength. The numerator is composed of Irradiance multiplied with wavelength and reflection. For a random polarization condition, both TE and TM polarizations are calculated, and the average is taken to get more realistic result. We used the formula given in Equation (1) in calculating the weighted averaged reflection in 400nm-1100nm wavelength range.

The novelty of our study is the following: 1) Instead of sweeping mode of the simulation, we optimized the parameters in sequence and each time, the optimization is done with respect to one parameter. This saves computational time. 2) We checked whether the angle of incline effects the reflection properties. 3) The optimization is done in terms of

filling ratio, height, pillar angle, and dielectric anti-reflection coating thickness. 4) We obtained the lowest averaged reflection loss for regular square array of silicon nanopillars.

As a first step, we fixed the diameter of the pillars as $d=190\text{nm}$ and pillar height as $h=120\text{nm}$ and varied the ratio of pillar diameter to period of pillar a which is d/a and varied that from 0.4 to 1.0. Minimum weighted averaged reflection is obtained as 3.91 percent at $d/a=0.7$ as depicted in Fig. 2. We used a to denote the period of the square nano-pillar array. Since we are dealing with the thin film device applications, we take the height range 60nm-190nm since larger and smaller height nano-pillar arrays have high reflectivity in the interested spectral region. As depicted in Fig.3, and Fig.4, we varied the height of the pillars and we found that 120nm is the optimum height especially for low reflection in 400nm-1100nm range. For large pillar heights, the reflectivity increases especially for UV region in the spectrum. For shorter pillars, the reflection increases especially for wavelengths larger than 600nm. The weighted averaged reflection calculations for the whole set of height range can be seen in Fig.4. Therefore, we fixed the pillar height at 120nm.

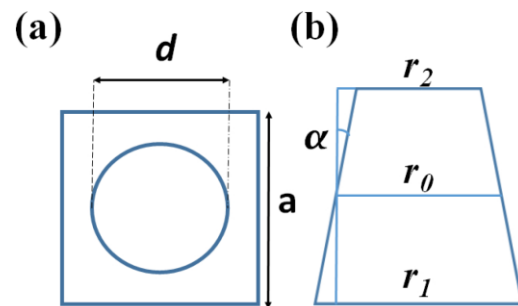


Fig.1. (a) Unit cell used in the simulation of square nanopillars, (b) Cross section of truncated pillar structure to define the pillar angle

$$\text{Weighted Average Reflection} = \frac{\int_{400}^{1100} \frac{R_{TE}(\lambda) + R_{TM}(\lambda)}{2} I(\lambda) \lambda d\lambda}{\int_{400}^{1100} I(\lambda) \lambda d\lambda} \quad (1)$$

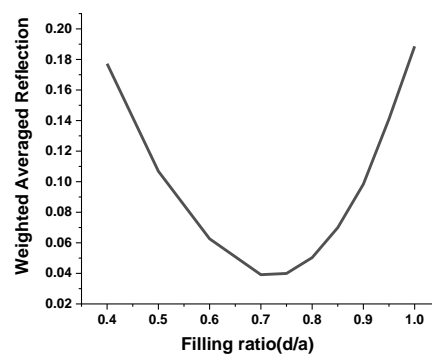


Fig.2. Weighted averaged power versus filling ratio of the pillars

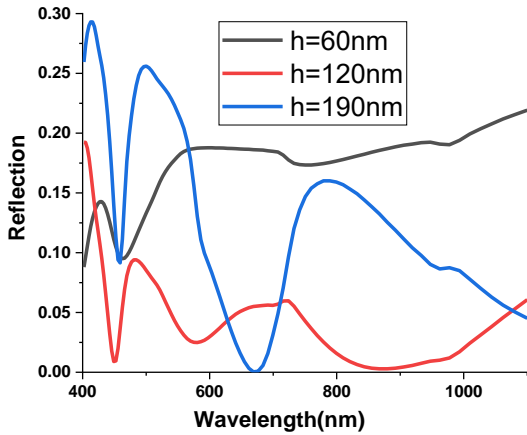


Fig.3. Height optimization of the pillars

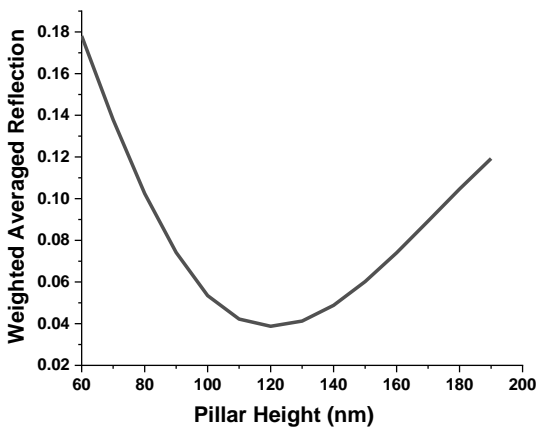


Fig.4. Weighted Averaged Reflection versus Pillar height

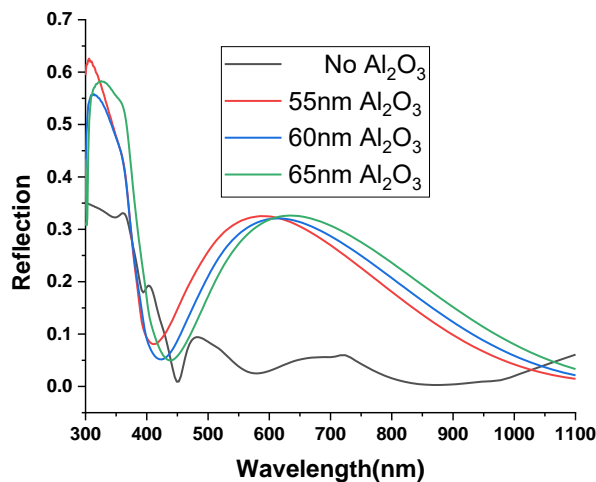


Fig.5. Anti-reflection dielectric Al₂O₃ thin film coating on the nanopillar array

We further examine the anti-reflection dielectric film Al₂O₃ coating. In figure 5, 55nm, 60nm, and 65 nm thick Al₂O₃ thin film added at top of the nano-pillar structures. As can be noticed, the reflection is minimum and small in about 50nm range around the minimum [15]. At other spectral wavelength range 400nm-1100nm, the reflection is even larger than we obtained when only nano pillar arrays are used Fig.5.

Therefore, instead of using the concept of antireflection minimization at one point, the analysis of a range of thin film thickness is needed using FDTD simulations.

As depicted in Fig.1(b), the truncated cone architecture is used to decrease further the reflection from the surface. The angle at which the pillar walls make with the vertical varied between 0-26 degrees. As can be seen in Fig.6, for wavelengths below 550nm, the reflection decreases with pillar slope however above 550nm, the reflection increases considerably. Truncated nanopillars apex angle can be achieved with optimizing the dry etching conditions, e.g., chamber pressure, temperature, gas flow rates. When the weighted average reflection is calculated for different pillar slopes (delta r is the difference of the bottom and top truncated nanocone pillar diameters) as shown in Fig.7. The minimum weighted averaged reflection occurs for vertical pillars therefore it is very important to obtain vertical pillars in square nanopillar arrays to obtain minimum reflection surfaces. For wavelengths smaller than 550nm, truncated nanocone arrays have lower reflections. However, for wavelengths larger than 550nm, they have higher reflections. For specific range of wavelengths, this property can be useful.

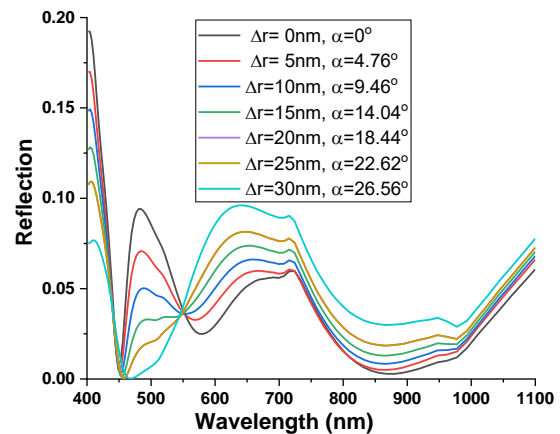


Fig.6. Reflection optimization with cone angle alpha.

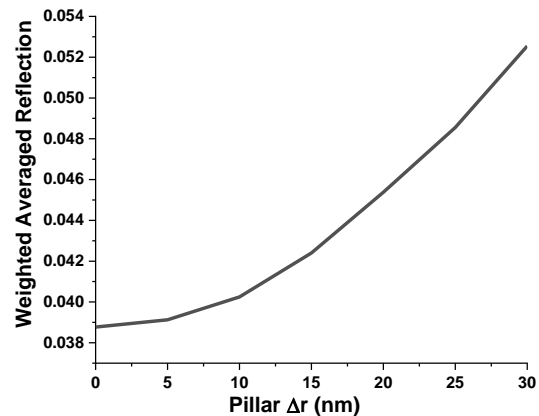


Fig.7. Total weighted reflection with solar irradiance on earth surface versus pillar angle.

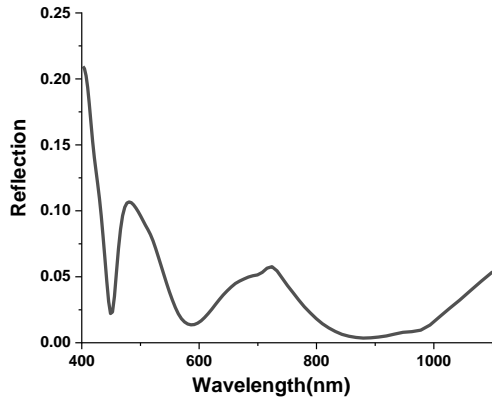


Fig.8. The optimum structure reflection versus wavelength with 5nm SiO₂ thin film.

It is also noticeable that the reflection is relatively high for the UV spectral region below 450nm wavelength. This is due to the fact that at small wavelengths the light rays start to ignore the nano-pillar array structure and get reflected from the bare silicon substrate surface. Therefore, reflection loss is high for UV region. Dielectric anti-reflection coating is also investigated by depositing SiO₂ layer on top of the all structure with thicknesses varied from 5nm to 70nm. The minimum reflection is obtained with weighted averaged reflection of 3.75 percent with 5nm thickness in 400nm-1000nm range which is one of the lowest reflection losses reported in the literature. The reflection spectra is depicted in Fig.8. When we compare the recent results from other groups, the lowest value for the average reflection from regular nanopillar square arrays is 4 % Table 1. This averaged reflection loss can be avoided further if compound nanopillars with a scattering tip on top of the pillar are used so that light can be absorbed in the substrate not in the pillar structures. The absorption in the UV region of the spectrum at the nano-pillar array structure is large; this is also a big issue if there is recombination in the nanostructure for energy extraction from an active device. We will pursue these issues in future work.

TABLE I
COMPARISON OF RECENT RESULTS FOR THE AVERAGE REFLECTIONS FROM REGULAR SQUARE NANOPILLAR ARRAYS

Author	Year	R (Average)
J. Li et al. [19]	2009	>10 %
C. Lin et al. [20]	2011	4.64 %
J. Proust et al. [21]	2016	4 %
J. Kim et al. [22]	2021	7.1%

III. CONCLUSION

The reflection properties of the nanostructured silicon surfaces have been investigated and understood physically through weighted average reflection as a function of wavelength. We see that using truncated nano-cones has potential to reduce further the total weighted reflection loss from the surface of the photonic devices surfaces for specific wavelength range. For broadband low reflection, vertical nano-pillars should be

used for square pillar arrays. We developed a simulation recipe to achieve minimum weighted average reflection architecture of the structure with respect to pillar height, filling ratio, incline angle, dielectric anti-reflection coating. Depositing 5nm SiO₂ thin film, one of the minimum reflections is obtained for square array type nano-pillar structures in the literature, which is 3.75 %. In order to optimize nanopillar structure further, adding a scattering tip structure on top of the pillars will be investigated as future study.

These modified nanostructured surfaces concept is useful for photonic device applications having thin film silicon or active absorbing semiconducting material for optical energy harvesting through low surface reflection over a broadband optical spectrum [16]. In broad sense, thin film solar cells, photodetectors, phototransistors applications are the potential applications of these nanostructured surfaces with low reflection.

ACKNOWLEDGMENT

This work is funded by TÜBİTAK 1001 project under project number 219M280. Computing resources were supported by AGÜ (Abdullah Gul University). I would like to thank my colleagues Evren Mutlugün and Mehmet Şahin for valuable discussions on the subject.

REFERENCES

- [1] P. Campbell, M. A. Green. "Light trapping properties of pyramidally textured surfaces." *J. Appl. Phys.* Vol. 62. No.1, 1987, pp 243-249.
- [2] S. Chattopadhyay, Y.F. Huang, Y.J. Jen, A. Ganguly, K.H. Chen, L.C. Chen. "Anti-reflecting and photonic nanostructures." *Mater. Sci. Eng. Rep.*, Vol.69. No.1-3, 2010, pp 1-35.
- [3] P. Lalanne and G. M. Morris, "Antireflection behavior of silicon subwavelength periodic structures for visible light." *Nanotechnology* Vol. 8. No.2, 1997, pp 53–56.
- [4] Y. Kanamori, M. Sasaki, and K. Hane, "Broadband antireflection gratings fabricated upon silicon substrates." *Opt. Lett.* Vol.24. No.20, 1999, pp 1422–1424.
- [5] K. Hadobás, S. Kirsch, A. Carl, M. Acet, and E. F. Wassermann, "Reflection properties of nanostructure-arrayed silicon surfaces." *Nanotechnology* Vol.11. No.3, 2000, pp 161–164.
- [6] H. Sai, H. Fujii, K. Arafune, Y. Ohshita, M. Yamaguchi, Y. Kanamori, and H. Yugami, "Antireflective subwavelength structures on crystalline Si fabricated using directly formed anodic porous alumina masks." *Appl. Phys. Lett.* Vol.88, No.20, 2006, pp 201116.
- [7] S. A. Boden and D. M. Bagnall, "Tunable reflection minima of nanostructured antireflective surfaces." *Appl. Phys. Lett.* Vol.93. No. 13, 2008, pp 133108.
- [8] Y.-H. Pai, Y.-C. Lin, J.-L. Tsai, and G.-R. Lin, "Nonlinear dependence between the surface reflectance and the duty-cycle of semiconductor nanorod array." *Opt. Express* Vol.19. No. 3, 2011, pp 1680–1690.
- [9] Y.-F. Huang, S. Chattopadhyay, Y.-J. Jen, C.-Y. Peng, T.-A. Liu, Y.-K. Hsu, C.-L. Pan, H.-C. Lo, C.-H. Hsu, Y.-H. Chang, C.-S. Lee, K.-H. Chen, and L.-C. Chen, "Improved broadband and quasi-omnidirectional anti-reflection properties with biomimetic silicon nanostructures." *Nat. Nanotechnol.* Vol.2. No.12, 2007, pp 770–774.
- [10] H. Sai, Y. Kanamori, K. Arafune, Y. Ohshita, and M. Yamaguchi, "Light trapping effect of submicron surface textures in crystalline Si solar cells." *Prog. Photovolt. Res. Appl.* Vol.15. No.5, 2007, pp 415–423.
- [11] S. A. Boden and D. M. Bagnall, "Optimization of moth-eye antireflection schemes for silicon solar cells." *Prog. Photovolt. Res. Appl.* Vol.18. No.3, 2010, pp 195–203.

- [12] H. Sai, H. Fujii, K. Arafune, Y. Ohshita, Y. Kanamori, H. Yugami, and M. Yamaguchi, "Wide-angle antireflection effect of subwavelength structures for solar cells." *Jpn. J. Appl. Phys.* Vol. No. 46-6A, 2007, pp 3333–3336.
- [13] P. Seliger, M. Mahvash, C. Wang, and A. F. J. Levi, "Optimization of aperiodic dielectric structures." *J. Appl. Phys.* Vol. 100. No.3, 2006, pp 034310–034316.
- [14] D. F. Edwards, "Silicon (Si)," *Handbook of Optical Constants of Solids*, E.D.Palik, ed. (Academic, Orlando, Fla., (1985).
- [15] B. L. Sopori and R. A. Pryor, "Design of antireflection coatings for textured silicon solar cells." *Sol. Cells* Vol.8. No.3, 1983, pp 249–261.
- [16] D. Shir, J. Yoon, D. Chanda, J.-H. Ryu, and J. A. Rogers, "Performance of ultrathin silicon solar microcells with nanostructures of relief formed by soft imprint lithography for broad band absorption enhancement." *Nano Lett.* Vol.10. No.8, 2010, pp 3041–3046.
- [17] H. Park, Y. Dan, K. Seo, Y. J. Yu, P. K. Duane, M. Wober, and K. B. Crozier, "Filter-Free Image Sensor Pixels Comprising Silicon Nanowires with Selective Color Absorption." *Nano Lett.* Vol.14. No.4, 2014, pp 1804–1809.
- [18] T. Tut, Y. Dan, P. Duane, Y. Yu, M. Wober, and K. B. Crozier, "Vertical waveguides integrated with silicon photodetectors: Towards high efficiency and low cross-talk image sensors.", *Appl. Phys. Lett.* Vol.100., 2012, pp 043504.
- [19] J. Li, H.u Yu, S. M. Wong, G. Zhang, X. Sun, P. G. Lo, and D. Kwong, "Si nanopillar array optimization on Si thin films for solar energy harvesting.", *Appl. Phys. Lett.* Vol. 95, 2009, pp 033102.
- [20] C. Lin, N. Huang, and M. L. Povinelli, "Effect of aperiodicity on the broadband reflection of silicon nanorod structures for photovoltaics.", *Opt. Express* Vol. 20, No.1, 2011, pp 125-132.
- [21] J. Proust, A. Fehrembach, F. Bedu, I. Ozerov, N. Bonod, "Optimized 2D array of thin silicon pillars for efficient antireflective coatings in the visible spectrum.", *Sci. Rep.* Vol. 6, 2016, pp 24947.
- [22] J. Kim, S. You, C.Kim, "Surface Texturing of Si with Periodically Arrayed Oblique Nanopillars to Achieve Antireflection.", *Materials* Vol. 14, 2021, pp380.

BIOGRAPHIES



TURGUT TUT was born in Mersin, Turkey, in 1978. He received his B.S. and M.S. degrees in physics from Bilkent University, in Ankara, Turkey, in 2004 and the Ph.D. degree in physics from Bilkent University, in Turkey, in 2008. From 2008 to 2011, he was a Postdoc researcher at Harvard University. He has been working at

Abdullah Gül University as Assistant Professor since 2018. He is the author of more than 20 articles, and one patent. His research interests include micro/nano-pillars, photodetectors, micro/nano-lithography techniques, thin film transistors, nanophotonics, numerical simulations of quantum systems and photonic devices.

Hybrid 3D Convolution and 2D Depthwise Separable Convolution Neural Network for Hyperspectral Image Classification

Hüseyin Firat*, Mehmet Emin Asker and Davut Hanbay

Abstract— Convolutional neural networks (CNNs) are one of deep learning methods that are often used to solve the problem of hyperspectral image classification (HSIC). CNN has a strong feature learning ability that can ensure more distinctive features for higher quality HSIC. The traditional CNN-based methods mainly use the 2D CNN for HSIC. However, with 2D CNN, only spatial features are extracted in HSI. Good feature maps cannot be extracted from spectral dimensions with the use of 2D CNN alone. By using 3D CNN, spatial-spectral features are extracted simultaneously. However, 3D CNN is computationally complex. In this study, a hybrid CNN method, which is a combination of 3D CNN and 2D CNN, is improved to solve the two problems described above. Using hybrid CNN decreases the complexity of the method compared to using only 3D CNN and can perform well against a limited number of training samples. On the other hand, in Hybrid CNN, depthwise separable convolution (DSC) is used, which decreases computational cost, prevents overfitting and enables more spatial feature extraction. By adding DSC to the developed hybrid CNN, a hybrid depthwise separable convolutional neural network is obtained. Extensive applications on frequently used HSI benchmark datasets show that the classification performance of the proposed network is better than compared methods.

Index Terms—3D Convolutional Neural Network, Depthwise Separable Convolution, Deep Learning, Hyperspectral Image Classification


I. INTRODUCTION

HYPERSPECTRAL IMAGES (HSIs) consist of tens or even hundreds of continuous narrow spectral bands with high spectral resolution, which can ensure abundant spatial-spectral feature information [1].


HÜSEYİN FIRAT, is with Vocational School of Technical Sciences of Dicle University, Diyarbakır, Turkey, (e-mail: huseyin.firat@dicle.edu.tr).

 <https://orcid.org/0000-0002-1257-8518>

MEHMET EMİN ASKER, is with Vocational School of Technical Sciences of Dicle University, Diyarbakır, Turkey, (e-mail: measker@dicle.edu.tr).

 <https://orcid.org/0000-0003-4585-4168>

DAVUT HANBAY, is with Department of Computer Engineering of Inonu University, Malatya, Turkey, (e-mail: davut.hanbay@inonu.edu.tr).

 <https://orcid.org/0000-0003-2271-7865>

Manuscript received December 21, 2021; accepted Jan 21, 2022.

DOI: [10.17694/bajece.1039029](https://doi.org/10.17694/bajece.1039029)

Since HSI has different spectral feature informations, it is commonly used in many fields such as agriculture, mining, astronomy, object tracking, military exploration, environmental monitoring, and vegetation [2-5]. In these applications, HSI is a major challenge for any classification method, as it includes multiple land cover classes resulting in high inter-class similarity and within-class variability. In recent years, many classification methods have been proposed to improve HSI classification (HSIC) performance. Since HSI contains rich spectral feature information, traditional classifiers such as support vector machines (SVM) [4], logistic regression [5], maximum likelihood [6], random forest [7] and k-nearest neighbors [5] have been proposed for HSIC. However, since these classification methods are based on spectral feature information, they cannot take full advantage of the important spatial feature information of HSIs. In this case, classification performance decreases. Nevertheless, traditional HSIC methods are based on handcrafted features with limited representation ability that do not fit well with the classification task [8].

In recent years, deep learning-based techniques have made great advances in many computer vision tasks, such as object detection [9], natural language processing [10], semantic segmentation [11-12], and image classification [13-14], because they can automatically extract robust and distinctive features from the original data in a hierarchical manner. Progress in deep learning techniques takes HSIC to a novel level. The training phase of deep learning techniques automatically extracts the features and uses these extracted features in the classification phase. Stacked autoencoder (SAE), a typical deep learning method, can extract spatial-spectral feature information. Later, it can combine this information for HSIC [15]. Chen et al. [16] applied the concept of deep learning to the field of HSIC at first. For spatial-spectral feature extraction and classification, they developed a method in which PCA, SAE and LR are used together. Tao et al. [17] proposed SAE based LR in which SAE and LR were used together to extract deep spectral features. Deep belief network (DBN) based classification is other deep learning methods used for HSI analysis. It has been proposed to combine the spatial feature information of HSI with spectral feature information [18]. Jia et al. [19] proposed a method in which PCA, hierarchical learning-based feature extraction and LR were used together for HSI analysis based on DBN. Li et al. [20] proposed a new HSIC method using optimal DBN and texture feature enhancement. While deep features can be extracted hierarchically in a layer-

based training process using SAE and DBN, training samples consisting of image patches to be input to these models need to be flattened to one dimension. In this case, spatial features in HSI cannot be fully used. Also, SAE and DBN are unsupervised learning algorithms and do not directly use informations of class while learning features [21]. Convolutional neural networks (CNN) are another deep learning method used for HSIC [22]. CNN has a strong feature learning ability that can ensure more distinctive features for higher quality HSIC [23]. The more distinctive the learned features are, the easier it will be to solve the classification problems [24]. This situation greatly improves the classification accuracy. Therefore, CNN-based methods used in HSIC are one of the most effective methods used to extract deeper spatial, spectral or spatial-spectral feature information. CNN-based methods are increasingly being applied as they effectively improve the HSIC performance. HSI is 3D data with two spatial dimensions (width, height) and spectral dimension (channel/depth). In CNN-based HSIC, mainly 2D CNN or 3D CNN is used for hierarchical feature extraction [25]. Chen et al. [26] proposed a 3D CNN method consisting of several convolution and pooling layers to extract spatial-spectral feature information. Also, dropout layer and L2 regularization were used to overcome the overfitting problem caused by the limited training samples. Roy et al. [27] proposed a hybrid spectral CNN (HybridSN) for HSIC. The HybridSN consists of using 3D CNN and 2D CNN together. 3D CNN extracts common spatial-spectral features from a spectral band stack. 2D CNN on top of 3D CNN extracts spatial features. Ahmad [28] proposed a fast 3D CNN method using spatial-spectral feature information to achieve better classification performances in HSIC. Ge et al. [29] proposed a deep network architecture based on multibranch feature fusion with 3D CNN and 2D CNN with different kernel size for HSIC. They also used Mish activation function instead of ReLU. He et al. [30] proposed a multiscale 3D deep CNN (M3D-DCNN) that can learn 2D multiscale spatial feature and 1D spectral feature jointly for HSIC. Mu et al. [31] proposed a multiscale and multilevel spatial spectral feature fusion network for HSIC. In the proposed method, From 3D to 2D alternative residual block combines spectral feature information extracted from 3D CNN with spatial feature information extracted from 2D CNN. Zhong et al. [32] proposed spectral-spatial residual network (SSRN) for HSIC. SSRN takes the 3D hyperspectral cube as input data without using any dimension reduction method. Residual blocks connect each 3D convolutional layer through identity mapping, which facilitates the back propagation of gradients. Mohan et al. [33] proposed a hybrid CNN model consisting of a multi-scale spatio-spectral feature based 3D CNN and 2D CNN for HSIC. They used the hybrid dimension reduction method consisting of nonlinear kernel principal component analysis (KPCA) and linear gaussian random projection (GRP) to reduce spectral band redundancy and provide optimum band extraction. Mohan et al. [34] proposed a hybrid CNN architecture consisting of 3D CNN and 2D CNN for HSIC. The proposed method uses PCA as the preprocessing step. 3D CNN is used to extract spatial-spectral features simultaneously. Then 2D CNN is used to extract more spatial features. However, the redundancy in its spatial features is decreased by octave convolution instead of classic convolution.

Cao et al. [35] proposed a deep 3D-2D-SSHDR method for HSIC that continuously extracts spectral and spatial feature information by combining spectral residual blocks, spatial hybrid dilated convolutions residual blocks, and outermost skip connection. Hamida et al. [36] and Li et al. [21] designed a 3D CNN to simultaneously extract spatial-spectral feature information for HSI classification.

In recent years, there has been an intense development of CNN-based methods for HSIC in which spatial features are adapted by a 2D CNN method. Nevertheless, since only the spatial information in the HSI will be preserved with 2D CNN, good distinguishing feature maps cannot be extracted from spectral dimensions. That is, if 2D convolution processes are implemented to HSIs, important spectral information is lost. By using 3D CNN, this problem is overcome. Since spectral and spatial features are extracted simultaneously with 3D CNN, the structural features of 3D HSIs are fully utilized in this study. However, the use of 3D CNN alone is computationally complex. Therefore, in the proposed method, 3D CNN and 2D CNN are joined to eliminate the shortcomings of 3D CNN and 2D CNN methods. In this way, maximum classification accuracy is achieved by fully utilizing both spatial and spectral feature maps. Also, some of the traditional 2D convolution layers have been replaced by depthwise separable convolution (DSC) layers. The purpose here is to solve the parameter and operating costs problem and prevent overfitting. In this study, a hybrid method consisting of 3D CNN, 2D CNN and 2D DSC layers is proposed. First of all, PCA is applied to the 3D HSI as a preprocessing step. Then, 3D hypercube datas are obtained to be given to the input of the network by neighborhood extraction. Firstly, 3D CNN is applied to extract spatial-spectral features simultaneously. Then feature information is converted from 3D data to 2D feature map and 2D CNN is applied to extract spatial feature. Finally, more spatial features are extracted using 2D DSC layers. DSC can improve HSI's feature learning capability and decrease computational complexity.

The rest of this paper is organized as follows: The datasets used in the study with PCA, 3D/2D Convolutional neural network, Depthwise separable convolution which constitute the theoretical background of the proposed method in this study, are explained in Section 2. In addition, the proposed method is detailed in Section 2. Application results and discussions about results are given in Section 3. This study is summarized in Section 4.

II. MATERIALS AND METHODS

A. Principal Component Analysis (PCA)

PCA is the most commonly used method in dimension reduction. With PCA, a significant reduction in the number of spectral features is possible while preserving most of the spatial features included in the HSI dataset. Significant correlation between spectral bands in HSI forms the principal of PCA. Analysis with PCA defines the optimal linear combination of the original spectral bands that takes into account the change of pixel values and also attempts to remove the correlation between spectral bands in HSI. The mathematical basic of PCA is based on the eigenvalue decomposition of the covariance matrix of the HSI spectral bands. Pixels of HSI are adjusted as

a vector whose dimension is the same as the number of spectral bands. $X_i = [x_1, x_2, x_3, \dots, x_N]^T$, N is the number of HSI spectral bands. The mean of all pixel vectors is calculated as in Equation (1).

$$m = \frac{1}{M} \sum_{i=1}^M [x_1, x_2, x_3, \dots, x_N]_i^T \quad (1)$$

$M = p * q$ represents the number of pixel vectors for an HSI consisting of " q " columns and " p " rows. The covariance matrix is defined as in Equation (2).

$$C = \frac{1}{M} \sum_{i=1}^M (X_i - m)(X_i - m)^T \quad (2)$$

The eigen decomposition of the covariance matrix, that is, the eigen values together with their corresponding eigen vectors, is expressed as in Equation (3).

$$C = ADA^T \quad (3)$$

D is the diagonal matrix consisting of the eigenvalues $(\lambda_1, \lambda_2, \dots, \lambda_N)$ of C , and $A = a_1, a_2, a_3, \dots, a_N$, is the orthogonal matrix with corresponding eigenvectors (each of dimension N) as columns [37]. The linear transformation $y_i = A^T X_i, i = 1, 2, \dots, M$ is adapted to obtain changed pixel vectors, which are PCA-converted spectral bands of the original HSI. The first row K of the A^T matrix is chosen so that the rows are eigenvectors suitable to their eigenvalues arranged in a decreasing order. The selected K rows are multiplied by the pixel vector X_i to give PCA spectral bands consisting of most of the information included in the HSI spectral bands [38].

B. 3D/2D Convolutional Neural Network

Convolutional neural network (CNN), which is one of the deep learning-based methods in HSIC, has been commonly used recently. CNN is a deep artificial neural network consisting of convolution, pooling, flatten, fully connected and softmax layers. The convolution layer is the first layer of CNN used to extract features by applying convolution operation to the input data. The convolution operation is performed by the inner product of the input image matrix and the kernel or a filter matrix. In this way, the output (feature map) of the input image is formed by shifting the kernel matrix over the input image matrix. The pooling layer reduces the width and height of the output (feature map) obtained as a result of the convolution operation without losing its features. In this way, both the required processing power is reduced and the unnecessary features that are caught are ignored and more important features are focused on. There are two different pooling methods generally used in CNN. Maximum and average pooling. In the pooling layer, which has a kernel (filter) as in the convolution layer, the kernel moves over the input image matrix. But instead of the convolutional operation, it applies the determined pooling method. In other words, if you are applying max pooling, it takes the largest value in the area covered by the filter, and if you are applying average pooling, it takes the

average of the values in the filter. This reduces the spatial dimension and retains important features. The flatten layer prepares the data at the input of the last and most important layer, the fully connected layer. Generally, neural networks take input data from a one-dimensional array. The data in this neural network is the one-dimensional array of matrices from the convolutional and pooling layers. The fully connected layer receives the data from the flatten layer and performs the learning process through the neural network. Activation function used as the last layer for classification in deep learning methods is softmax function. As a result of the softmax function, the probabilistic distributions of the classes to be used in the classification process are generated. The probabilistic value generated for each class is in the range of 0 to 1. The total probability value of all classes is equal to 1 [33]. For x inputs, the output of a single neuron is calculated as in Equation (4).

$$t = f(w * x + bias) \quad (4)$$

In Equation (4), w represents the filter weight. $f(\cdot)$ denotes nonlinear activation function implemented to a weighted input sum. In the 2D CNN method, convolution is performed using the 2D kernel before passing the input HSI data through the activation function. The convolution performs by calculating the sum of the inner product between kernel and the input HSI. The kernel moves on on the HSI to cover the entire spatial dimension. This convolution operation provides to extract spatial features from the HSI. The 2D convolution output of each neuron is formulated as in Equation (5).

$$t_{mn} = f \left(\sum_l \sum_{i=0}^{h-1} \sum_{j=0}^{w-1} k_{ij} x_{(i+m)(j+n)} + bias_{mn} \right) \quad (5)$$

In Equation (5), t_{mn} is the feature extracted at (m, n) position. k is the 2D convolution kernel at dimension $h \times w$. In the case of a 2D image, this convolution operation is applied on all feature maps (l) in the receiver area and sums all values for non-linear activation. This process is repeated for all layers in the case of multi-dimensional data. When data are 3D they have spatial-spectral dimensions. 2D convolution fails for this input data. Because, in traditional 2D CNN, convolution processes are applied only to 2D feature maps that capture features in spatial dimension. 3D CNN is a changed version of 2D CNN methods that apply 3D convolution instead of 2D convolution in 2D CNN. When convolution processes are applied to 3D data, it is desirable to capture features from both spatial and spectral dimensions. For this purpose, 3D CNN is used, in which 3D convolution processes are implemented to 3D hypercubes to calculate co-spatial spectral features from 3D input data. 3D convolution enables the extraction of spatial-spectral features from 3D images. The feature extracted from the 3D CNN method is formulated as in Equation (6).

$$t_{mnd} = f \left(\sum_l \sum_{i=0}^{h-1} \sum_{j=0}^{w-1} \sum_{r=0}^{b-1} k_{ijr} X_{(i+m)(j+n)(r+d)} + bias_{mnd} \right) \quad (6)$$

In Equation (6), t_{mnd} is the feature extracted at (m,n,d) position. b is the dimension of the 3D kernel along the spectral dimension. The kernel (k) is 3D and the features are calculated by applying 3D convolution on the 3D input data. In traditional 2D CNNs, the convolution process to obtain 2D feature maps is applied only on spatial dimensions and covers all feature maps of the previous layer. However, for HSIC, it is desirable to obtain spectral features information as well as spatial features. 2D CNNs cannot extract spectral features. 3D CNN kernel can simultaneously extract spatial-spectral feature information from HSI despite increasing computational complexity. To benefit advantage of the automatic feature learning of both 3D CNN and 2D CNN, a hybrid CNN method consisting of 3D and 2D CNN is used under the HSIC.

C. Depthwise Separable Convolution

HSIs are different from 2D images and include a lot of information in the spatial dimension. Spatial features are determined by the excess information contained in the spatial dimension of HSIs. These spatial features can compensate for the shortcomings of spectral features to improve the model's ability to capture features. Spatial features of HSIs can be extracted with 2D CNN. However, since HSIs are 3D, they are not suitable for simultaneous extraction of spatial-spectral features with 2D CNN. Considering of the inadequate use of hyperspectral data information by 2D convolution, a DSC layer is added after the 2D convolution layer, which can increase spatial features and decrease parameters. With the addition of the DSC layer, a lot of spatial-spectral features are extracted to enable the method to losslessly distinguish spatial information of different spectral bands [39].

DSC consists of two separate operations, namely, depthwise convolution (DC) and pointwise convolution (PC) (1x1 convolution or also known as point-by-point convolution). As shown in Fig. 1, a separate convolution operation is performed on each channel of the input image in DC. With DC convolution operation, spatial features are extracted on each dimensions. In PC convolution operation, 1x1 standard convolution is applied to the depthwise feature map obtained as a result of DC convolution. PC convolution is used to assemble the feature map across channels [40].

Fig. 1a shows the standard convolution operation and Fig. 1b shows the DSC. In Figure 1, the size of the 3D input image is $W \times H \times C$. W is the width of the input image, H is the height of the input image, and C is the number of depth/channel. Suppose the size of the convolution kernel is $K \times K \times C$ and the number of convolution kernels is N . The input image is convoluted with each convolution kernel and a feature map (output) of size $W \times H \times N$ is obtained. The computational cost of the standard convolution operation is as in Equation (7).

$$cost_{standart_conv} = KxKxCxNxWxH \quad (7)$$

The DSC implements a single channel filtering operation for each input channel, added after the DC filtering, and then performs PC. In DCS, in order to apply DC to the input image of size $W \times H \times C$ first, it is divided into C groups with convolution kernel size $K \times K \times 1$. Each group then performs standart convolution, which is equivalent to extracting the spatial features of each input channel C , ie Depthwise feature. The output size obtained after this operation is $W \times H \times C$. In PC, on the other hand, $1 \times 1 \times C$ convolution kernels are used. In other words, $1 \times 1 \times C$ convolution operation with N number of convolution kernels is applied to the output feature map obtained at the end of DC. The size of the output feature map after DC and PC is $W \times H \times N$. Equations (8) and (9) give the computational costs of DC and PC, respectively.

$$cost_{DCconv} = KxKxWxHxC \quad (8)$$

$$cost_{PCconv} = CxWxHxN \quad (9)$$

The calculation required to compare the DSC with the standard convolution is given in Equation (10).

$$\frac{cost_{DCconv} + cost_{PCconv}}{cost_{standart_conv}} = \frac{1}{K^2} + \frac{1}{N} \quad (10)$$

Considering Equation (10), it can be seen that the trainable parameters and calculations of the DSC are $\frac{1}{K^2} + \frac{1}{N}$ times of the standard convolution. With the use of DSC, the number of trainable parameters and the computational cost are significantly reduced.

D. Information of HSI Datasets

Three HSI datasets including Indian Pines, Pavia of University and Salinas datasets were used to evaluate the performance of the proposed method and state-of-the-art techniques. Indian Pines (IP), Pavia of University (PU) and Salinas (SA) datasets are taken from the website http://www.ehu.es/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes.

The IP is data obtained by the Airborne Visible / Infrared Imaging Spectrometer (AVIRIS) sensor at the Indian Pines test site in North-western Indiana. Each spectral image in this dataset is 145×145 spatial dimension. The sensor achieved a total of 224 spectral bands in the 0.4-2.5 micrometers wavelength range. From these 224 bands, 24 bands that are in the water absorption region and not useful were removed. A total of 200 spectral bands are used in applications. The IP includes 16 classes and 10,249 samples. The PU is data obtained by Reflective Optics System Imaging Spectrometer (ROSIS) optical sensors over the Pavia University in northern Italy. PU is 340 pixels width and 610 pixels height, with a spatial resolution of 1.3 meters per pixel. The dataset has a 115 wavelength ranging from 0.43-0.86 micrometer. After removing 12 noise bands in applications, a total of 103 spectral bands are used. The PU includes 9 classes and 42,776 samples. The SA is data collected by the AVIRIS sensor over Salinas Valley region of California. SA has a spatial resolution of 3.7 meters per pixel. The dataset is a remote sensing image 217

pixels width and 512 pixels height and contains 224 spectral bands. From these 224 bands, 20 bands that are in the water absorption region and not useful were removed. A total of 204 spectral bands are used in applications. The SA includes a total

of 16 classes and 54,129 samples. The class and sample number information of the datasets are given in Table I, while the false-color images are shown in Fig. 2.

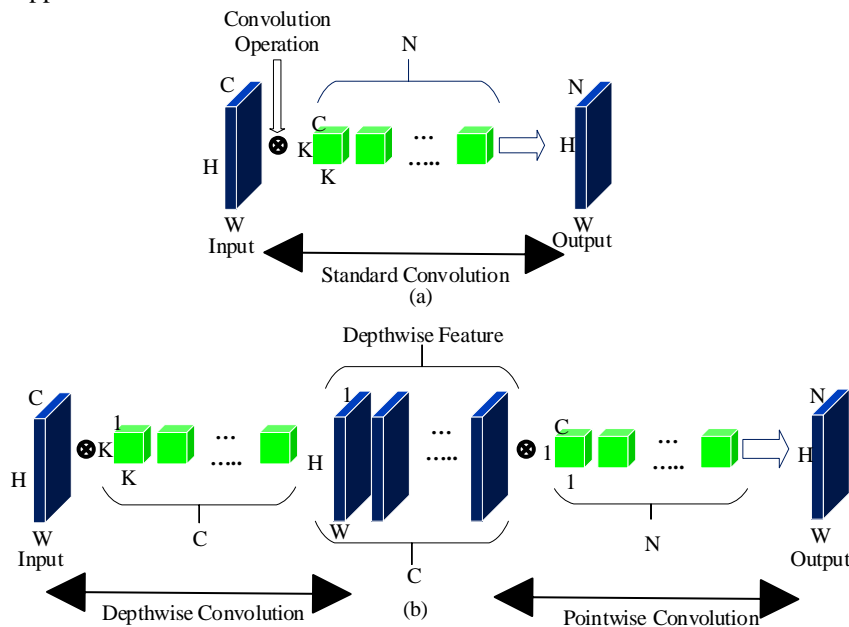


Fig. 1. (a) Standart convolution operation. (b) Depthwise separable convolution operation

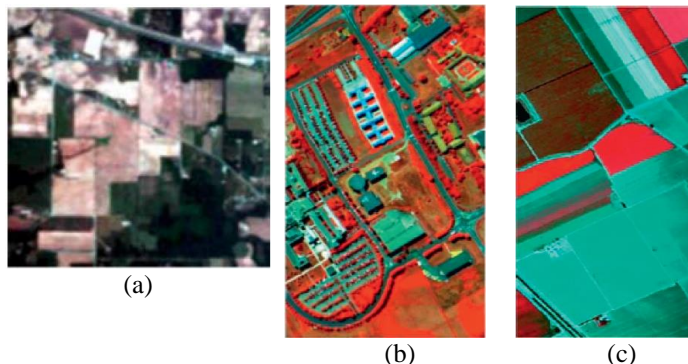


Fig. 2. False-color images of (a) Indian Pines, (b) University of Pavia, (c) Salinas datasets

TABLE I
INFORMATION REGARDING SAMPLES OF EACH CLASS IN IP, PU, AND SA DATASETS

No	Indian Pines (IP)		Salinas (SA)		Pavia University (PU)	
	Classes	Samples	Classes	Samples	Classes	Samples
1	Alfalfa	46	Broccoli_green_weeds_1	2009	Asphalt	6631
2	Corn-notill	1428	Broccoli_green_weeds_2	3726	Meadows	18,649
3	Corn-mintill	830	Fallow	1976	Gravel	2099
4	Corn	237	Fallow_rough_plow	1394	Trees	3064
5	Grass-pasture	483	Fallow_smooth	2678	Painted metal sheets	1345
6	Grass-trees	730	Stubble	3959	Bare soil	5029
7	Grass-pasture-mowed	28	Celery	3579	Bitumen	1330
8	Hay-windrowed	478	Grapes_untrained	11,271	Self-blocking Bricks	3682
9	Oats	20	Soil_vinyard_develop	6203	Shadows	947
10	Soybean-notill	972	Corn_senesced_green_weeds	3278		
11	Soybean-mintill	2455	Lettuce_roumaine_4wk	1068		
12	Soybean-clean	593	Lettuce_roumaine_5wk	1927		
13	Wheat	205	Lettuce_roumaine_6wk	916		
14	Woods	1265	Lettuce_roumaine_7wk	1070		
15	Buildings-grass-trees-drives	386	Vinyard_untrained	7268		
16	Stone-steel-towers	93	Vinyard_vertical_trellis	1807		
	Total Number	10,249		54,129		42,776

E. Our Proposed Method

HSI X is specified as a 3D data cube with two spatial and one spectral dimensions of size $M \times N \times D$. X is the original input. While D indicates the number of spectral bands, M and N are the spatial width and height of the HSI. The ground truth of the input image Y is converted using one-hot encoding and expressed as $Y = (y_1, y_2, \dots, y_C)$. C indicates the number of classes in the HSI. HSI pixels exhibit high inter-class similarity, high intra-class variability, overlapping and nested regions, which require extensive effort for any classification method. To overcome these problems, it is necessary to remove the spectral band redundancy. As a preprocessing step to remove spectral band redundancy, conventional principal component analysis (PCA) is first applied to the original HSI (X) data along the spectral bands. With PCA, the number of spectral bands is reduced from D to B while maintaining the same spatial dimensions (i.e., width M and height N). That is, using PCA, the spatial feature information, which is very important for recognizing any class, is preserved, while only the number of spectral bands is reduced. Thus, it is reduced to the desired number of bands. After the PCA, the modified input image X is represent as $M \times N \times B$. M is the width, N is the height, and B is the number of newly obtained spectral bands. In the second stage, to use HSIC techniques, the HSI cube is split into small overlapping 3D spatial patches from which real labels based on the central pixel are generated. 3D neighboring patches $S \times S \times B$ is created from the HSI cube (X) centered in the spatial position (a, b) and covering the $S \times S$ window size or spatial dimension and all spectral bands B . The total number of 3D patches (n) produced from the HSI cube (X) is found by $(M - S + 1) \times (N - S + 1)$. Thus, these patches at location (a, b) , covers the width from $a - (S - 1)/2$ to $a + (S - 1)/2$, height from $b - (S - 1)/2$ to $b + (S - 1)/2$, and all spectral bands (B) of HSI cube (X) [27][28]. The generated 3D HSI data cubes is given to the input of the Hybrid 3D/2D CNN architecture for feature extraction and classification.

The proposed Hybrid 3D/2D CNN consists of 3 3D CNNs, 2 2D CNNs, 1 Separable Conv layers (DC + PC layers), flattening, 2 Fully connected (FC), dropout and softmax layers. To preserve the maximum spatial-spectral information, the pooling layer in CNN is not to be used. The 3D convolution process is applied three times to simultaneously increase the number of spatial-spectral feature maps and can protect the spectral information of the input HSI data in the output volume. Dimensions of the applied 3D convolution kernels and the filters used are as follows. In the first convolutional layer, 32 filters with the dimension of $3 \times 3 \times 7$ (ie, two spatial and one spectral dimensions), 64 filters with the dimension of $3 \times 3 \times 3$ in the 2nd convolution layer, 64 filters with the dimension of $1 \times 1 \times 1$ in the 3rd convolution layer are used. 2D convolution is applied twice before the flattened layer, bearing in mind that it strongly discriminates spatial information in different spectral bands without significant loss of spectral information, which is crucial for HSI data. The dimensions of the 2D convolution kernels and the filters used are as follows. 128 filters with the dimension of 3×3 in the 1st convolution layer and 128 filters with the dimension of 1×1 in the 2nd convolutional layer are used. Similarly, 2D DSC layers containing depthwise and

pointwise convolution layers before the flattening layer are applied. The dimensions of the 2D DSC layer kernels and the filters used are as follows. 128 filters with the dimension of 3×3 in the DSC layer are used. Since $f(\cdot)$ ReLU is the most frequently used activation function in CNN, it is used in all convolution layers. ReLU is faster than other activation functions when using gradient descent techniques for training. ReLU is formulated as in Equation (11).

$$ReLU(t) = \max(0, t) \quad (11)$$

The features extracted after the 3D and 2D convolution layers are flattened and given as input to the FC layers for classification. In proposed method, two FC layers with 256 and 128 neurons are used. To prevent overfitting, a dropout layer with a dropout rate of 0,4% is implemented after each FC layer. The output of the FC layer is given to the simple softmax classifier to produce the classification result. Softmax is another activation function often used in the last layer of a deep learning method for classification. This activation function produces the probability distribution of all samples and their sum equals one. More details on the proposed method are shown in Table II. The total number of trainable weight parameters in the proposed method for PU dataset is 1,465,481.

TABLE II
SUMMARY OF THE PROPOSED METHOD FOR PU

Layer(Type)	Output Shape	Parameters
Input (InputLayer)	(11, 11, 15, 1)	0
Conv3d	(9, 9, 9, 32)	2048
Conv3d_1	(7, 7, 7, 64)	55360
Conv3d_2	(7, 7, 7, 64)	4160
Reshape	(7, 7, 448)	0
Conv2d	(5, 5, 128)	516224
Separable_conv2d	(5, 5, 128)	17664
Conv2d_1	(5, 5, 128)	16512
Flatten	3200	0
Dense	256	819456
Dropout	256	0
Dense_1	128	32896
Dropout_1	128	0
Dense_2	9	1161
Total Trainable Parameters		1,465,481

III. APPLICATION RESULTS AND DISCUSSIONS

All applications are performed using Colaboratory (Colab) Notebook. Colab is an online platform offered by Google that provides free access to the Graphical Processing Unit (GPU) and Tensor Processing Units (TPU) as hardware accelerators. Google Colab needs a good internet speed while running on any hardware accelerator. It is preferred to use the TPU as a hardware accelerator while running our applications. The TPU offers the ability to run code on a Python 3 desktop computer with 35 GB of RAM and 107.77 GB of storage for data computing. In all applications, categorical crossentropy with 0.001 learning rate and Adam optimizer with $1e-06$ decay function used for optimization is used. The training process repeats for 256 batch sizes and 100 epochs. ReLU is used as activation function in all layers except the Softmax layer. The proportion of training samples and the size of the input are factors that affect the accuracies of the HSIC. All three datasets use 11×11 neighborhoods (window size) for convolution. Training sample rates of IP, PU and SA datasets are taken as

20%, 10% and 10%, respectively. On the other hand, in our applications, 30, 15 and 15 principal components are used for IP, PU and SA datasets, respectively. In order to make a better comparison with other deep learning-based methods, the PC values were chosen in this way. That is, dimension of the 3D patches obtained as a result of neighborhood extraction was set to $11 \times 11 \times 30$ for IP, $11 \times 11 \times 15$ for PU and SA datasets, respectively. Overall accuracy (OA), Average accuracy (AA) and Kappa statistics (K) have used to evaluate the classification results of each method. OA is computed by the ratio between correctly classified samples in the test samples and the total number of test samples (Eq. (12)). AA is the average value of the accuracy of each class (Eq. (13)). K is a statistical measurement metric that provides mutual information regarding a strong agreement between the basic truth map and the classification map (Eq. (14)). The advantage of using the

Kappa value is to consider the effect of uncertainties on classification accuracy when predicting classification accuracy.

$$OA = \sum_{i=1}^K M_{ii} / N \tag{12}$$

$$AA = \sum_{i=1}^K \left(M_{ii} / \sum_{j=1}^K M_{ij} \right) / K \tag{13}$$

$$Kappa = \left(N \sum_{i=1}^K M_{ii} - \sum_{i=1}^K (M_{i+} \times M_{+i}) \right) / \left(N^2 - \sum_{i=1}^K (M_{i+} \times M_{+i}) \right) \tag{14}$$

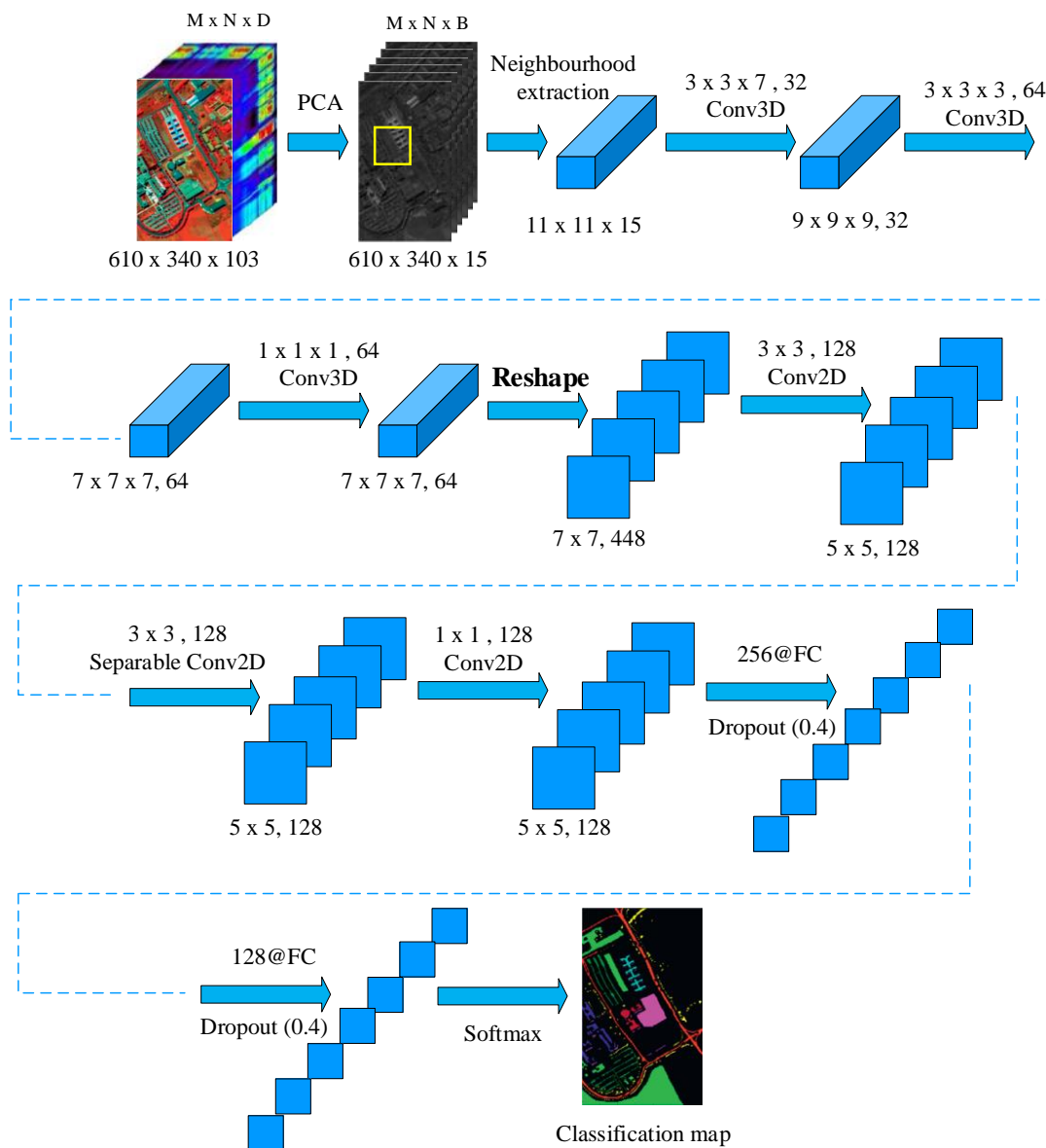


Fig. 3. Proposed hybrid 3D/2D CNN method for HSIC

In Equations (12), (13) and (14), K is the number of classes used for HSIC. M represents the confusion matrix. The confusion matrix is created by comparing the predicted classification map with ground truth that is actual value. M_{ii} shows the main diagonal values of the confusion matrix and is the number of correctly classified samples belonging to class i . N is the total number of test samples. M_{i+} indicates the total number of samples in row i , while M_{+i} indicates the total number of samples in column i .

The proposed method (PM) has been compared with HSIC methods based on deep learning from the literature, such as SVM-RBF [41], 2D CNN [42], 3D CNN [21], SSRN [32], and HybridSN[27].

In the first application, IP dataset is used. Classification accuracy of each class according to class-based and OA, AA and K evaluation metrics is shown in Table III. Ground truth and the classification maps obtained as a result of the applications performed with the IP dataset are given in Fig. 4. When Table III is examined, it is seen that the proposed method obtained the best classification result with 99.32%, 99.46% and

99.22%, respectively, according to the OA, AA and K evaluation metrics. The closest results to the proposed method were obtained with SSRN with 99.19% OA, 98.93% AA, and 99.07% K values, HybridSN with 99.26% OA, 98.98% AA, and 99.09% K. Among the other classification methods, the lowest OA, AA and K values were obtained in SVM-RBF with 82.83%, 80.84% and 82.23%, respectively. Considering the AA value in SVM-RBF, it is seen that the number of correct classifications for each class is lower than other methods. With the proposed method in class-based classification, the best classification accuracies of 100% are obtained in all classes except for the 2nd, 3rd, 5th, 10th, 11th, 12th and 16th classes. The best classification accuracy for class 2 was obtained in the SSRN method with a difference of 0.57% from the proposed method. The best classification accuracy for class 3 was obtained in the SSRN method with a difference of 1.79% from the proposed method. Similarly, the best classification accuracy was obtained in SSRN with 99.24% for class 5, HybridSN with 98.76% for class 10, PM with 99.75% for class 11, HybridSN with 99.12% for class 12, and PM with 98.65% for class 16.

TABLE III
NUMBER OF TRAINING AND TEST SAMPLES OF IP DATASET AND CLASSIFICATION ACCURACIES (%)

No. of classes	Train/Test	SVM-RBF	2D CNN	3D CNN	SSRN	HybridSN	PM
1	9/37	61.5	85.88	94.63	97.82	97.98	100
2	286/1142	78.68	91.31	93.9	99.17	98.37	98.6
3	166/664	73.41	91.07	94.85	99.53	99.48	97.74
4	47/190	71.58	80.38	93.48	97.79	97.38	100
5	97/386	80.38	91.89	93.56	99.24	99.23	98.96
6	146/584	92.27	99.01	94.2	99.51	99.14	100
7	6/22	79.52	82.59	89.73	98.7	99	100
8	96/382	87.38	100	96.01	99.85	100	100
9	4/16	85.87	66.55	95	98.5	99.01	100
10	194/778	77.58	86.38	94.55	98.74	98.76	98.71
11	491/1964	83.75	90.47	93.87	99.3	99.64	99.75
12	119/474	83.21	82.89	91.52	98.43	99.12	98.95
13	41/164	84.64	99.06	93.89	100	100	100
14	253/1012	98.01	97.86	91.77	99.31	100	100
15	77/309	94.3	90.52	95.03	99.2	99.35	100
16	19/74	61.43	98.94	93.57	97.82	97.26	98.65
OA (%)		82.83	90.89	94.07	99.19	99.26	99.32
AA (%)		80.84	89.68	93.72	98.93	98.98	99.46
Kappa x 100		82.23	88.56	93.87	99.07	99.09	99.22

The second application is performed with the PU dataset. The classification accuracies obtained as a result of the applications performed with the PU dataset are given in Table IV. The classification maps obtained as a result of the classification are shown in Fig. 5. When Table IV is examined, it is seen that the proposed method obtained the best classification result with 99.83%, 99.70% and 99.78%, respectively, according to the OA, AA and K evaluation metrics. The proposed method obtained 0.11%, 0.1%, 0.14% better results in terms of OA, AA and K values, respectively, than HybridSN, which is one of the most advanced methods. In addition, the proposed method obtained 0.21%, 0.21% and 0.28% better results in terms of OA, AA and K values, respectively, compared to SSRN, which is another of the most advanced methods. The lowest classification results are obtained in the SVM-RBF method

with 82.67% OA, 80.84% AA and 81.21% K values. Classification accuracies of 96.89% OA, 95.79% AA, 2D CNN with 96.56% K and 3D CNN with 99.07% OA, 98.75% AA, 98.87% K were obtained, respectively. When class-based classification accuracies are compared, the proposed method has the best classification performance with 100% in the 5th, 6th and 9th classes. In addition, it is seen that the best classification results are obtained with 99.95% in class 1, 99.98% in class 2, 99.53% in class 4 and 99.46% in class 8. The best classification results are obtained in the HybridSN method, with 99.01% in class 3 and 99.69% in class 7. Considering all classification accuracies, it is seen that the classification map obtained by the proposed method is quite close to the ground truth.

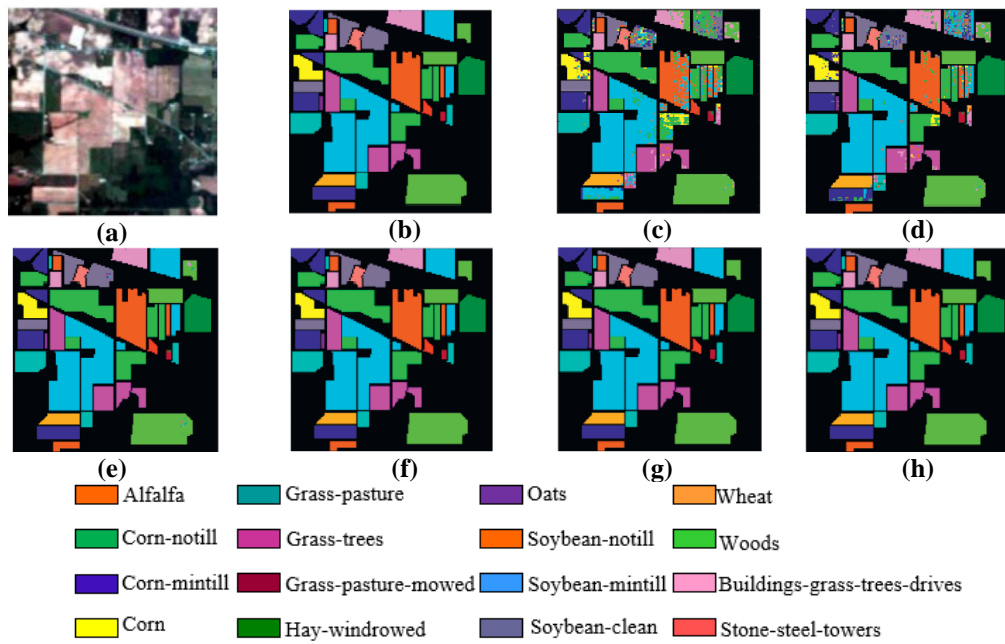


Fig. 4. Classification maps obtained as a result of predicted for IP. (a) False-color image, (b) ground truth, (c) SVM-RBF, (d) 2D CNN, (e) 3D CNN, (f) SSRN, (g) HybridSN, and (h) Proposed method

TABLE IV
NUMBER OF TRAINING AND TEST SAMPLES OF PU DATASET AND CLASSIFICATION ACCURACIES (%)

No. of classes	Train/Test	SVM-RBF	2D CNN	3D CNN	SSRN	HybridSN	PM
1	663/5968	93.68	97.37	97.4	99.75	99.76	99.95
2	1865/16784	97.02	99.26	94.73	99.79	99.78	99.98
3	210/1889	82.41	80.73	95.05	98.29	99.01	98.94
4	306/2758	96.51	95.54	98.04	99.52	99.53	99.53
5	135/1211	98.38	99.75	99.01	99.82	99.87	100
6	503/4526	90.01	93.14	98.62	99.77	99.77	100
7	133/1197	85.92	91.65	97.02	99.65	99.69	99.42
8	368/3314	88.08	92.39	98.23	99.05	99.21	99.46
9	95/852	99.85	99.09	99.29	99.78	99.80	100
OA (%)		82.67	96.89	99.07	99.62	99.72	99.83
AA (%)		80.84	95.79	98.75	99.49	99.60	99.70
Kappa x 100		81.21	96.56	98.87	99.50	99.64	99.78

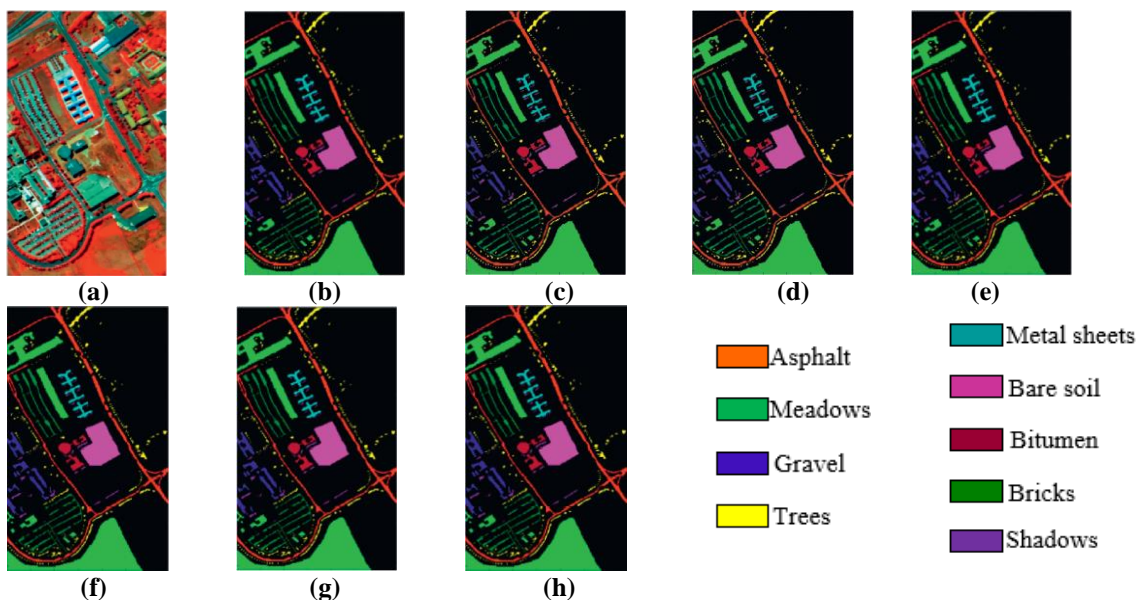


Fig. 5. Classification maps obtained as a result of predicted for PU. (a) False-color image, (b) ground truth, (c) SVM-RBF, (d) 2D CNN, (e) 3D CNN, (f) SSRN, (g) HybridSN, and (h) Proposed method

The third application is performed with the SA dataset. The classification accuracies obtained as a result of the applications performed with the SA dataset are given in Table V. The classification maps obtained as a result of the classification are shown in Fig. 6. When Table V is examined, the best classification accuracies were obtained with 99.90% OA, 99.93% AA and 99.89 K values in the proposed method. The closest results to the proposed method were obtained in the HybridSN method with 99.80% and SSRN method with 99.64% OA, 98.76% AA, 99.60% K values in all evaluation metrics. The proposed method achieved better classification accuracy of 7.23% OA, 8.32% AA and 7.68% K, respectively, from SVM-RBF, 4.56% OA, 5.18% AA and 4.96% K from 2D CNN, and 5.88% OA, 6.44% AA and 6.32% K from 3D CNN. When the class-based classification results were examined, the

proposed method achieved the best classification accuracy with 100% in the 1st, 2nd, 3rd, 6th, 7th, 9th, 11th, 12th, 13th, 16th classes. The best classification results were obtained in the HybridSN method with 100% for the 4th, 5th, 8th, 10th and 14th classes. The proposed method with 99.63% for class 15 gives the best results. Considering Table V and Figure 6, the classification maps obtained by the proposed method and HybridSN methods are very similar to the ground truth.

TABLE V
NUMBER OF TRAINING AND TEST SAMPLES OF SA DATASET AND CLASSIFICATION ACCURACIES (%)

No. of classes	Train/Test	SVM-RBF	2D CNN	3D CNN	SSRN	HybridSN	PM
1	201/1808	92.56	95.51	95.19	99.78	100	100
2	373/3353	94.78	95.87	96.12	99.78	100	100
3	198/1778	94.22	95.64	95.89	99.78	100	100
4	139/1255	97.36	99.99	97.18	99.99	100	99.76
5	268/2410	93.38	94.51	94.70	99.78	100	99.92
6	396/3563	95.36	96.71	95.63	99.78	100	100
7	358/3221	95.79	95.51	95.36	99.78	100	100
8	1127/10144	80.21	86.93	86.33	99.78	100	99.85
9	620/5583	97.67	97.42	97.19	99.78	100	100
10	328/2950	88.99	90.93	90.94	92.69	100	99.90
11	107/961	90.20	94.24	90.20	90.68	98.24	100
12	193/1734	96.42	99.91	96.24	99.95	99.61	100
13	92/824	95.53	95.53	95.31	99.85	99.59	100
14	107/963	91.26	92.97	94.18	99.88	100	99.90
15	727/6541	72.67	90.93	81.39	99.12	99.32	99.63
16	181/1626	89.37	93.43	94.09	99.78	100	100
OA (%)		92.67	95.34	94.02	99.64	99.80	99.90
AA (%)		91.61	94.75	93.49	98.76	99.80	99.93
Kappa x 100		92.21	94.93	93.57	99.60	99.80	99.89

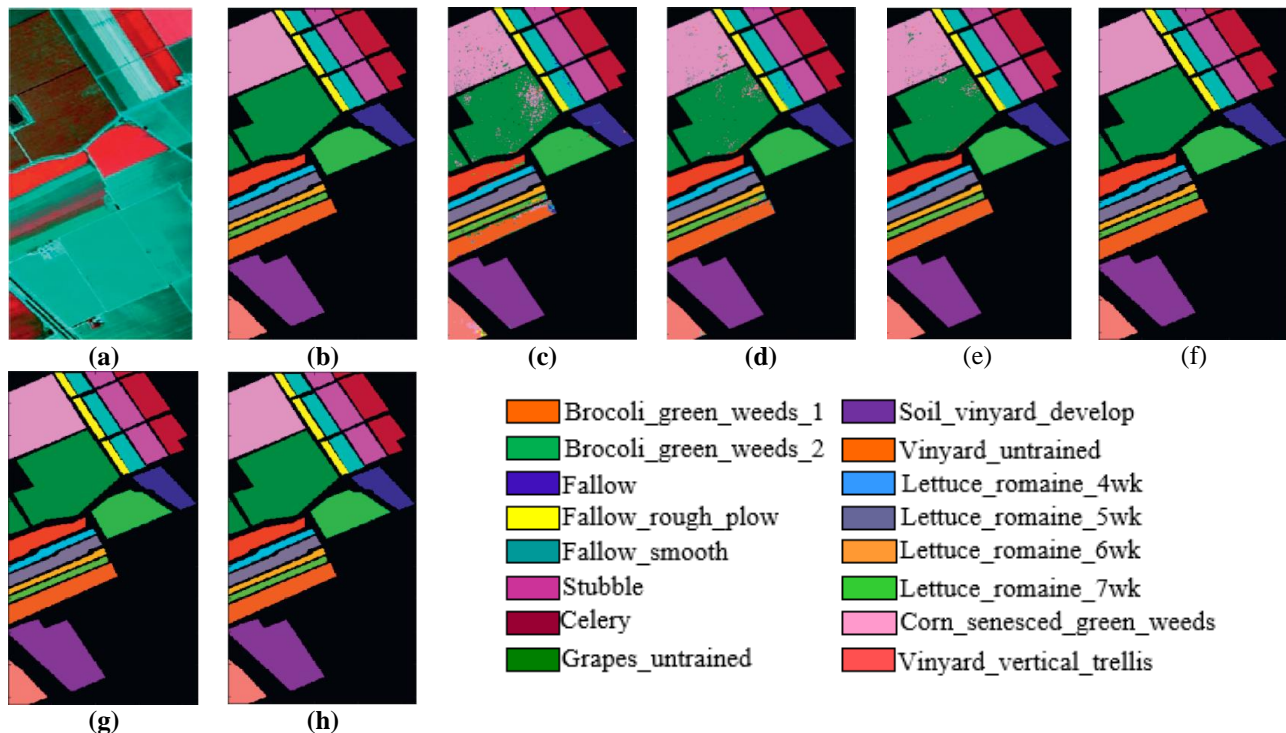


Fig. 6. Classification maps obtained as a result of predicted for SA. (a) False-color image, (b) ground truth, (c) SVM-RBF, (d) 2D CNN, (e) 3D CNN, (f) SSRN, (g) HybridSN, and (h) Proposed method

IV. CONCLUSION

In this study, a hybrid depthwise separable CNN method for HSIC is proposed. The method combines 3D/2D CNN and spatially depthwise separable convolution layers. While spatial-spectral features are extracted with 3D CNN, spatial features are continuously extracted using 2D CNN. With depthwise separable convolution, the number of trainable parameters and computation time are reduced and more spatial features are extracted. Considering the applications performed with IP, PU and SA benchmark datasets, with the proposed method, 99.32% OA, 99.46% AA, 99.22% K in IP, 99.83% OA, 99.70% AA,

99.78% K in PU, 99.90% OA, 99.93% AA, 99.89% K values in SA were obtained. In line with all these values, the proposed method was compared with SVM-RBF, 2D CNN, 3D CNN, SSRN, HybridSN methods. When Table III-IV-V is examined, it is seen that better classification performance is achieved with the proposed method. In future studies, Dilated convolution, which is another convolution method, will be examined first to solve the HSIC problem. Later, applications will be developed on how it affects HSIC performance when used with depthwise separable convolution.

REFERENCES

- [1] H. Gao, Y. Yang, C. Li, L. Gao, and B. Zhang, "Multiscale Residual Network with Mixed Depthwise Convolution for Hyperspectral Image Classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 4, pp. 3396–3408, 2021, doi: 10.1109/TGRS.2020.3008286.
- [2] H. Firat and D. Hanbay, "4CF-Net: Hiperpektral uzaktan algılama görüntülerinin spektral uzamsal sınıflandırılması için yeni 3B evrişimli sinir ağı," *Gazi Üniversitesi Mühendislik-Mimarlık Fakültesi Dergisi*, vol. 1, pp. 439–453, 2021, doi: 10.17341/gazimmfd.901291.
- [3] H. Firat, M. Uçan, and D. Hanbay, "Classification of Hyperspectral Remote Sensing Images Using Hybrid 3D-2D CNN Architecture," *Journal of Computer Science*, vol. IDAP-2021, no. Special, pp. 132–140, 2021.
- [4] Y. Wang, W. Yu, and Z. Fang, "Multiple Kernel-based SVM classification of hyperspectral images by combining spectral, spatial, and semantic information," *Remote Sensing*, vol. 12, no. 1, 2020, doi: 10.3390/RS12010120.
- [5] M. Ahmad et al., "Spatial prior fuzziness pool-based interactive classification of hyperspectral images," *Remote Sensing*, vol. 11, no. 9, pp. 1–19, 2019, doi: 10.3390/rs11091136.
- [6] A. Alcolea, M. E. Paoletti, J. M. Haut, J. Resano, and A. Plaza, "Inference in supervised spectral classifiers for on-board hyperspectral imaging: An overview," *Remote Sensing*, vol. 12, no. 3, pp. 1–29, 2020, doi: 10.3390/rs12030534.
- [7] J. S. Ham, Y. Chen, M. M. Crawford, and J. Ghosh, "Investigation of the random forest framework for classification of hyperspectral data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 3, pp. 492–501, 2005, doi: 10.1109/TGRS.2004.842481.
- [8] S. Ghaderizadeh, D. Abbasi-Moghadam, A. Sharifi, N. Zhao, and A. Tariq, "Hyperspectral Image Classification Using a Hybrid 3D-2D Convolutional Neural Networks," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 7570–7588, 2021, doi: 10.1109/JSTARS.2021.3099118.
- [9] M. F. Özdemir and D. Hanbay, "Comparison of Optimization Algorithms for Multi-Object Tracking FairMOT Algorithm," *Journal of Computer Science*, vol. IDAP-2021, no. Special, pp. 147–153, 2021.
- [10] İ. Sel and D. Hanbay, "Creating a Parallel Corpora for Turkish-English Academic Translations," *Journal of Computer Science*, vol. IDAP-2021, no. Special, pp. 335–340, 2021.
- [11] G. MENDİ and C. BUDAK, "Automatic Cell Nucleus Segmentation Using Superpixels and Clustering Methods in Histopathological Images," *Balkan Journal of Electrical and Computer Engineering*, vol. 9, no. 3, pp. 304–309, 2021, doi: 10.17694/bajece.864266.
- [12] C. Budak, V. Mençik, and V. Gider, "Determining similarities of COVID-19 – lung cancer drugs and affinity binding mode analysis by graph neural network-based GEFA method," *Journal of Biomolecular Structure and Dynamics*, vol. 0, no. 0, pp. 1–13, 2021, doi: 10.1080/07391102.2021.2010601.
- [13] H. Üzen, H. Firat, A. Karçl, and D. Hanbay, "Automatic Thresholding Method Developed With Entropy For Fabric Defect Detection," in *2019 International Artificial Intelligence and Data Processing Symposium (IDAP)*, 2019, pp. 14–17.
- [14] H. Uzen, M. Turkoglu, and D. Hanbay, "Texture defect classification with multiple pooling and filter ensemble based on deep neural network," *Expert Systems with Applications*, vol. 175, no. March, p. 114838, 2021, doi: 10.1016/j.eswa.2021.114838.
- [15] C. Zhao, X. Wan, G. Zhao, B. Cui, W. Liu, and B. Qi, "Spectral-Spatial Classification of Hyperspectral Imagery Based on Stacked Sparse Autoencoder and Random Forest," *European Journal of Remote Sensing*, vol. 50, no. 1, pp. 47–63, 2017, doi: 10.1080/22797254.2017.1274566.
- [16] H. Data et al., "Deep Learning-Based Classification of Hyperspectral Data," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 6, pp. 2094–2107, 2014, doi: 10.1109/JSTARS.2014.2329330.
- [17] A. Mughees and L. Tao, "Efficient deep auto-encoder learning for the classification of hyperspectral images," *Proc. - 2016 Int. Conf. Virtual Real. Vis. ICVRV 2016*, no. September, pp. 44–51, 2017, doi: 10.1109/ICVRV.2016.16.
- [18] P. Zhong, Z. Gong, S. Li, and C. B. Schonlieb, "Learning to Diversify Deep Belief Networks for Hyperspectral Image Classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 6, pp. 3516–3530, 2017, doi: 10.1109/TGRS.2017.2675902.
- [19] Y. Chen, X. Zhao, and X. Jia, "Spectral-Spatial Classification of Hyperspectral Data Based on Deep Belief Network," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 6, pp. 2381–2392, 2015, doi: 10.1109/JSTARS.2015.2388577.
- [20] J. Li, B. Xi, Y. Li, Q. Du, and K. Wang, "Hyperspectral classification based on texture feature enhancement and deep belief networks," *Remote Sensing*, vol. 10, no. 3, 2018, doi: 10.3390/rs10030396.
- [21] Y. Li, H. Zhang, and Q. Shen, "Spectral-spatial classification of hyperspectral imagery with 3D convolutional neural network," *Remote Sensing*, vol. 9, no. 1, 2017, doi: 10.3390/rs9010067.
- [22] C. BUDAK, V. MENÇİK, and M. E. ASKER, "Effect on model performance of regularization methods," *DÜMF Mühendislik Dergisi*, vol. 5, pp. 757–765, 2021, doi: 10.24012/dumf.1051352.
- [23] H. S. Nogay, T. C. Akinci, and M. Yilmaz, "Detection of invisible cracks in ceramic materials using by pre-trained deep convolutional neural network," *Neural Computings and Applications*, vol. 0123456789, 2021, doi: 10.1007/s00521-021-06652-w.
- [24] C. Zhang et al., "Joint Deep Learning for land cover and land use classification," *Remote Sensing of Environment*, vol. 221, no. May 2018, pp. 173–187, 2019, doi: 10.1016/j.rse.2018.11.014.
- [25] H. Firat, M. E. Asker, and D. Hanbay, "Classification of hyperspectral remote sensing images using different dimension reduction methods with 3D/2D CNN," *Remote Sensing Applications: Society and Environment*, p. 100694, 2022, doi: 10.1016/j.rsase.2022.100694.
- [26] P. Chen, Y.; Jiang, H.; Li, C.; Jia, X.; Ghamisi, "Deep feature extraction and classification of hyperspectral images based on convolutional neural networks," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 10, pp. 6232–6251, 2016, doi: 10.1049/iet-ipt.2019.1282.
- [27] S. K. Roy, G. Krishna, S. R. Dubey, and B. B. Chaudhuri, "HybridSN: Exploring 3D-2D CNN Feature Hierarchy for Hyperspectral Image Classification," *arXiv*, vol. 17, no. 2, pp. 277–281, 2019.
- [28] M. Ahmad, A. M. Khan, M. Mazzara, S. Distefano, M. Ali, and M. S. Sarfraz, "A Fast and Compact 3-D CNN for Hyperspectral Image Classification," *IEEE Geoscience and Remote Sensing Letters*, no. April, pp. 1–5, 2020, doi: 10.1109/LGRS.2020.3043710.
- [29] Z. Ge, G. Cao, X. Li, and P. Fu, "Hyperspectral Image Classification

Method Based on 2D-3D CNN and Multibranch Feature Fusion,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 5776–5788, 2020, doi: 10.1109/JSTARS.2020.3024841.

- [30] H. C. Mingyi He, Bo Li, “Multi-scale 3D deep convolutional neural network for hyperspectral image classification,” 2017 *IEEE International Conference on Image Processing (ICIP)*, pp. 3904–3908, 2017.
- [31] C. Mu, Z. Guo, and Y. Liu, “A multi-scale and multi-level spectral-spatial feature fusion network for hyperspectral image classification,” *Remote Sensing*, vol. 12, no. 1, 2020, doi: 10.3390/RS12010125.
- [32] Z. Zhong, J. Li, Z. Luo, and M. Chapman, “Spectral-Spatial Residual Network for Hyperspectral Image Classification: A 3-D Deep Learning Framework,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 2, pp. 847–858, 2018, doi: 10.1109/TGRS.2017.2755542.
- [33] A. Mohan and M. Venkatesan, “HybridCNN based hyperspectral image classification using multiscale spatio-spectral features,” *Infrared Physics and Technology*, vol. 108, no. March, 2020, doi: 10.1016/j.infrared.2020.103326.
- [34] A. Mohan and V. Meenakshi Sundaram, “V3O2: hybrid deep learning model for hyperspectral image classification using vanilla-3D and octave-2D convolution,” *Journal of Real-Time Image Processing*, no. 0123456789, 2020, doi: 10.1007/s11554-020-00966-z.
- [35] F. Cao and W. Guo, “Deep hybrid dilated residual networks for hyperspectral image classification,” *Neurocomputing*, vol. 384, pp. 170–181, 2020, doi: 10.1016/j.neucom.2019.11.092.
- [36] A. Ben Hamida, A. Benoit, P. Lambert, and C. Ben Amar, “3-D deep learning approach for remote sensing image classification,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 8, pp. 4420–4434, 2018, doi: 10.1109/TGRS.2018.2818945.
- [37] M. Ahmad, S. Shabbir, R. A. Raza, M. Mazzara, S. Distefano, and A. M. Khan, “Hyperspectral Image Classification: Artifacts of Dimension Reduction on Hybrid CNN,” no. January, pp. 1–9, 2021.
- [38] L. Jiang, B. Zhu, and Y. Tao, “Hyperspectral Image Classification Methods,” *Hyperspectral Imaging for Food Quality Analysis and Control*, pp. 79–98, 2010, doi: 10.1016/B978-0-12-374753-2.10003-6.
- [39] A. Wang, C. Liu, D. Xue, H. Wu, Y. Zhang, and M. Liu, “Depthwise separable relation network for small sample hyperspectral image classification,” *Symmetry (Basel)*, vol. 13, no. 9, 2021, doi: 10.3390/sym13091673.
- [40] L. Dang, P. Pang, and J. Lee, “Depth-wise separable convolution neural network with residual connection for hyperspectral image classification,” *Remote Sensing*, vol. 12, no. 20, pp. 1–20, 2020, doi: 10.3390/rs12203408.
- [41] B. C. Kuo, H. H. Ho, C. H. Li, C. C. Hung, and J. S. Taur, “A kernel-based feature selection method for SVM with RBF kernel for hyperspectral image classification,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 1, pp. 317–326, 2014, doi: 10.1109/JSTARS.2013.2262926.
- [42] Q. Wang, J. Gao, and Y. Yuan, “A Joint Convolutional Neural Networks and Context Transfer for Street Scenes Labeling,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 5, pp. 1457–1470, 2018, doi: 10.1109/TITS.2017.2726546.

BIOGRAPHIES



HÜSEYİN FIRAT received the B.S. degree in computer engineering from Cukurova University, Adana, Turkey, in 2014. He received the M.S. degree in computer engineering from Inonu University in 2018. He is currently working toward the Ph.D. degree in the computer engineering. He also works as a lecturer at Dicle University in Turkey. His current interests include remote sensing, deep learning, and hyperspectral image classification.



MEHMET EMİN ASKER was born in Diyarbakır, Turkey, in 1976. He received the B.S. degree in electrical electronics engineering, from Firat University, Elazığ, Turkey in 1993, the M.S. degree and the Ph.D. degree in electrical machines, power electronics from Firat University, Elazığ, Turkey, in 2009 and 2016, respectively. He is an Assistant Professor with Dicle University, Department of electrical power and energy. Where he teaches courses on power system, power electronics, circuit theory and electrical machines since 2007. His research interests include electrical machines, power electronics and power systems.



DAVUT HANBAY received the B.S., M.S. and Ph.D. degrees in electric–electronic engineering from Firat University, Elazığ, Turkey, in 1999, 2003, and 2007, respectively. He is the author or coauthor of more than 30 papers in journals and conferences. His current interests include classification, optimization and pattern recognition.

Estimation of Survival According to Body Mass Index (BMI), Hypertension, Diabetes and Heart Disease with Optimizable Decision Trees

N. Hakime Nogay and H. Selcuk Nogay


Abstract— Non-communicable chronic diseases such as cardiovascular diseases and diabetes and the risk factors of these diseases are becoming an increasing health and development problem in the world. Non-communicable chronic diseases are among the most important causes of death according to the World Health Organization (WHO). The prediction of death or survival is very important in terms of contributing to scientific studies for the earlier diagnosis of non-communicable chronic diseases. Today's developing world, where technology and artificial intelligence can be used in every field, enables the prediction of survival in chronic diseases to be realized with many machine learning methods. In order to know which artificial intelligence or machine learning method is the most effective, it will be very useful to make applications with the methods used and even with the subclasses of the same method and to compare the classification results obtained from the applications with each other. In this study, survival in chronic diseases was estimated by using decision tree methods in four different structures designed by training with body mass index taken from individuals with chronic diseases and other hospital records. The highest accuracy rate was obtained with the optimizable decision trees (ODT) method, which is the simplest model among these models, which allows the most optimal selection of hyperparameters.

Index Terms—Survival, Coarse, Fine, Medium, Optimizable Decision Tree, Body Mass Index,


I. INTRODUCTION

Among the machine learning methods, the decision tree method is one of the most preferred methods by many researchers for problems that can be solved by binary classification [1]. Deaths from chronic non-communicable diseases (such as cardiovascular diseases, diabetes, cancer, etc.) are a growing problem of global health in middle- and high-income countries.

NALAN HAKIME NOGAY, is with Department of Nutrition and Diabetic, Erciyes University, Kayseri, Turkey.(e-mail: nalannogay@erciyes.edu.tr).

 <https://orcid.org/0000-0002-9435-5755>

HIDIR SELCUK NOGAY, is with Department of Electrical, Kayseri University, Kayseri, Turkey (e-mail: nogay@kayseri.edu.tr).

 <https://orcid.org/0000-0001-9105-508X>

Manuscript received December 16, 2021; accepted January 30, 2022.
DOI: [10.17694/bajece.1037645](https://doi.org/10.17694/bajece.1037645)

According to WHO, six out of 10 causes of death in 2018 were chronic noncommunicable diseases. Again, according to WHO statistics, cardiovascular diseases alone caused the death of 17.65 million people in 2015 [2]. Studies have shown that an appropriate cardiovascular health profile is associated with a lower risk of developing other chronic diseases [3]. In a study investigating the risks of premature death from noncommunicable chronic diseases, high systolic blood pressure, high body mass index and risks related to dietary intake emerged as the main risk factors among both women and men [4]. Detection of risk factors for deaths from chronic diseases may contribute to the development of new health policies to reduce the burden of these diseases [5]. Considering the scientific studies examined, it is seen that there is a need for more machine learning algorithm-based studies for the prediction of death or survival in chronic diseases. In this study, the survival of individuals with chronic diseases was predicted by decision tree methods, which is one of the most popular machine learning algorithms.

The rest of the study is organized as follows; In the second section, technical information about the decision tree method used in the study is given. In the third part of the study, explanations about the preferred method and data set are given in order to carry out the study. In this section, the stages of the study are presented. In the results section, the results and graphics obtained from the study are given. In the last section, the interpretation of the results is given.

II. THEORETICAL FRAMEWORK

Among the machine learning methods used for classification, the decision tree method continues to maintain its popularity thanks to its comprehensible and simple rules, easy interpretability and intelligibility [6]. In order to classify or categorize a data with the decision tree method, it is necessary to train the decision tree model and then subject it to the classification process. During the training of the model, a training data is used to create and train the model. A decision tree is obtained by using model classification rules trained with training data. In the second stage, which is the classification stage, test or validation data is used. Test or validation data tests the accuracy or reliability of the decision tree. If the obtained accuracy rate has reached the desired level, the rules are used to

classify the next data row. Training data plays a key role in building the tree [7, 8]. A decision tree model has nodes, branches, and leaves, and each attribute is represented by a node. In decision trees, the top part of the tree is called the root, as opposed to a normal tree view. The branches are between the root and leaves [8, 9]. In order to create a decision tree, conclusions are drawn from the clues obtained from the data reserved for training. With these clues, decision rules are created to classify the data at the root node of the decision tree. At the root node, which is the first node of the tree, questions are started to be asked based on the attributes in the data in order to classify the data and create the tree structure. In this way nodes, branches and leaves are created. Test data from the root of the tree is applied to the tree so that the created decision tree can produce predictions with each data set and test the results. This new dataset, which is used for testing starting from the root, is sent to the child nodes or branches according to the results from the testing process. This process continues like an

iteration until it reaches a certain last leaf on the tree. There is a separate path, or a separate decision rule, from the root to each leaf of the tree. Figure 1 shows a decision tree structure consisting of two-dimensional attribute values belonging to two classes. In the figure, the "column_9" attribute values; column_9 < values a, b, c, and d represent the threshold values for branching, and 1 and 0 represent the class labels. Variables in the tree structure can be one or more [10-12].

In decision trees, each variable is divided into at least two sub-variables. A threshold value of "c" can be set to maximize differences or reduce similarities [13-15]. In the decision tree design process, criteria and rules are determined by using some techniques for the formation of branches and nodes. Some of these techniques are; information gain, information gain ratio, Gini index, Towing rule, and Ki-Square probabilistic table statistics. CHAID, QUEST, ID3, C4.5, C5.0, and CART algorithms are some of the algorithms created using these techniques.

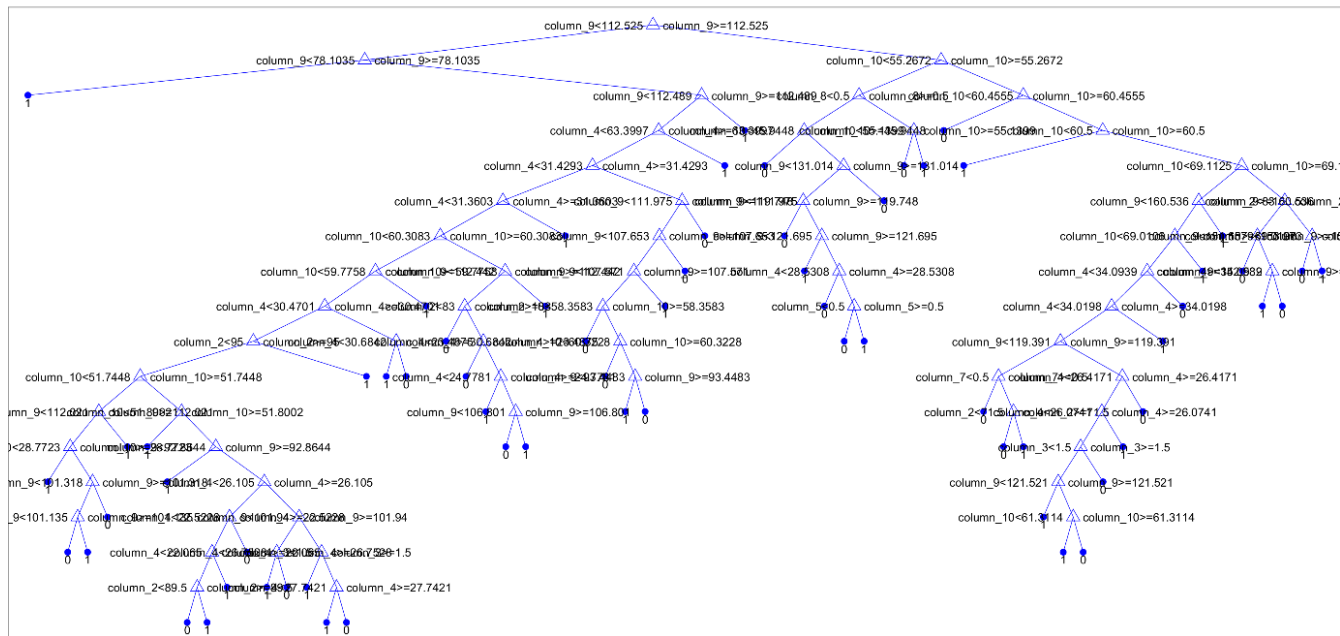


Fig. 1. Fine DT

In the information gain method, entropy rules are used to represent disorder in a system. ID3 and C4.5 algorithms can be given as examples of algorithms using the information gain approach [15]. In this study, the Gini index technique was used while creating the decision tree branches and nodes. In the Gini Index, hypothetically, all variables are continuous. The possibility that each variable can be divided into as many possible categories is considered [15]. The Gini Index is calculated as follows for the T data set with n categories and N samples.

$$gini(T) = 1 - \sum_{j=1}^n p_j^2 \tag{1}$$

Where p_j is the relative frequency of class j in T. If the T data set is divided into two classes as T1 and T2 and the data

numbers of these classes are N1 and N2, the Gini Index is calculated as follows.

$$gini_a(T) = \frac{N_1}{N} gini(T_1) + \frac{N_2}{N} gini(T_2) \tag{2}$$

Where 'a' is the number of separations, and the variable with the lowest Gini value is selected.

When a very complex structure emerges in the decision trees, the parts of the decision tree that do not affect the classification accuracy can be removed. This process which is simplifying the decision tree is called pruning [16-20]. In Figure 2, pruned branches are shown for the Fine Decision Tree (FDT) model.

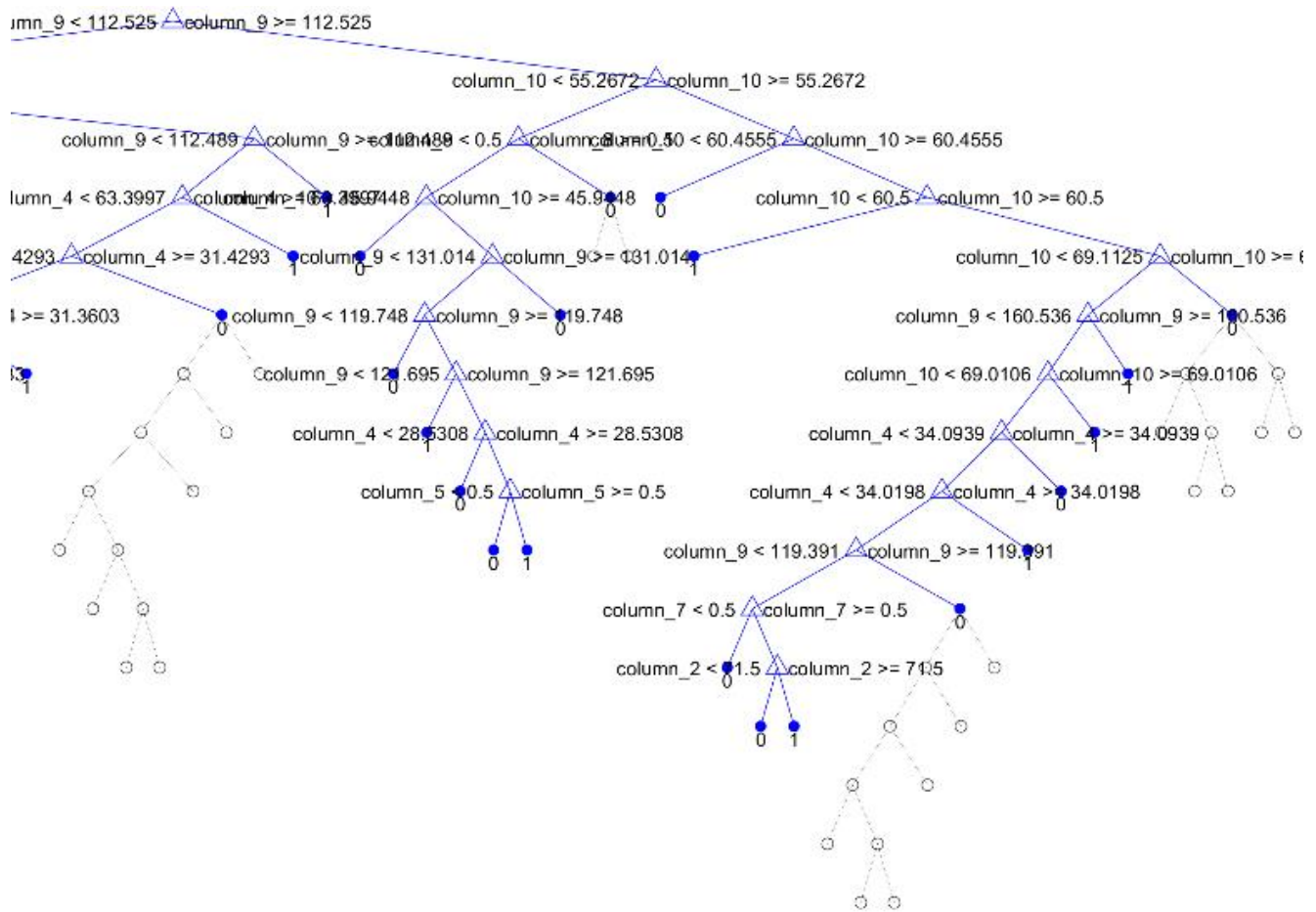


Figure 2. Right part of the fine DT model

III. METHODOLOGY

This study was carried out in the "Classification Learner Toolbox" environment under the "Machine Learning and Deep Learning" tab in the MATLAB R2021a package program. Figure 3 shows the Medium Decision Tree (MDT) model used in the study. The internal structure of the Optimizable Decision Tree (ODT) model designed in the study is shown in Figure 4. In the decision tree model, the "Optimizable" option was selected, which automatically selects the most optimal selection of hyperparameters and gives us minimum error curves at the end of this process. Therefore, the decision tree model proposed in the study is the ODT model. The Gini Index technique was preferred in determining the rules. Training and test data are needed to create the decision tree model. The data set used in the study was obtained from the data analysis study performed by Jingmin et al. to make a more accurate prediction of mortality among heart patients admitted to intensive care units [21]. In the created data set, body mass index (BMI), hypertensive, atrial fibrillation, diabetes, hyperlipidemia,

systolic blood pressure, diastolic blood pressure, and survival status of 1177 participants were taken to be used for this study. It was recognized that BMI data of some patients were not recorded. So these unsaved rows were removed from the data set, the most recently used data set in the study consisted of 962 rows and 10 columns. The data set used in the study can be accessed from the <https://doi.org/10.5061/dryad.0p2ngf1zd> web address. Table 1 presents the summary of the data set used in the study. In addition, the definitions of the variables used for the training and test data in the study are shown in Table 1. 'Respond' in the dataset is survival. Survival is represented numerically as 0 and death as 1.

In addition to the ODT method in the study, in order to show the effectiveness of the proposed ODT method, estimation was made with FDT, MDT and CDT in the MATLAB environment. In the decision tree models (Figure 1), the numbers at the end of each branch represent the survival status. In order to increase the reliability of the study and test its accuracy, the 5-fold cross-validation method was used.

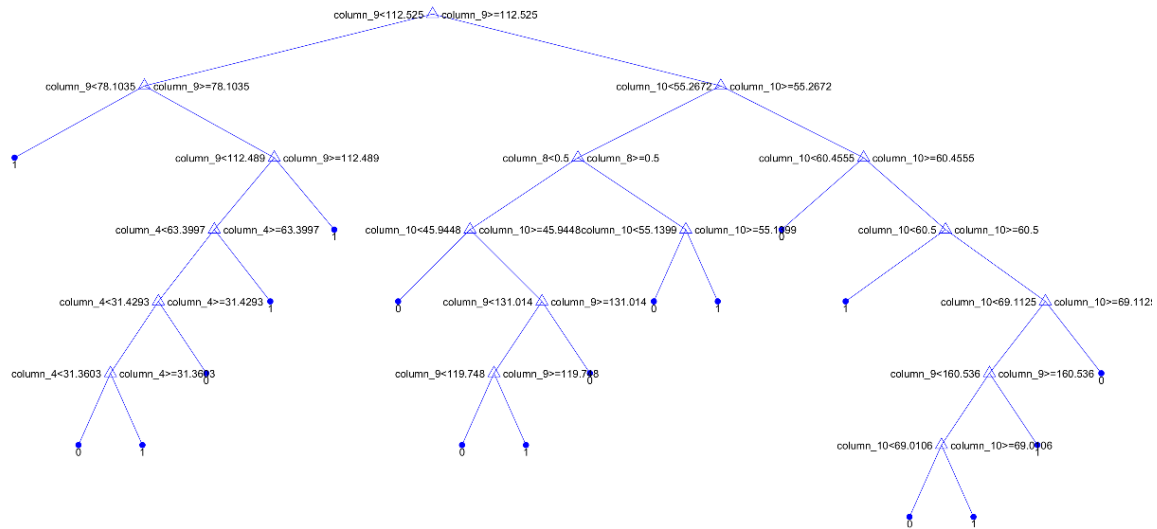


Figure 3. Medium DT model

TABLE I
SUMMARY OF THE DATA SET

Data	Variable Label	Variable Definition	Min	Max
Predictors	Column_2	Age	19	99
	Column_3	Gender (1:F, 2:M)	1	2
	Column_4	BMI	133.46	104.9
	Column_5	Hypertensive	0	1
	Column_6	Atrial fibrillation	0	1
	Column_7	Diabetes	0	1
	Column_8	Hyperlipemia	0	1
	Column_9	Systolic blood pressure	75	203
	Column_10	Diastolic blood pressure	247.36	107
	Respond	Column_1	Outcome	0

IV. RESULTS

The results obtained from the decision tree classifiers designed and applied in the study are presented in Table 2. According to the results in Table 2, the highest accuracy rate was obtained with the ODT method. Estimates of survival and death among individuals participating in the study can be examined through the confusion matrices in Figure 5 obtained for each of the four models.

TABLE II
THE RESULTS OF APPLICATIONS

	FDT	MDT	CDT	ODT
Accuracy (%)	81.1	86.2	87.8	88.1
Prediction speed (obs/sec)	20000	67000	80000	68000
Training time (sec)	34.154	0.4704	0.3466	32.441
Maximum number of splits	100	20	4	1

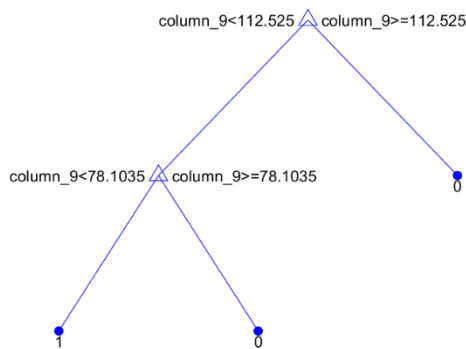


Figure 4. Structure of a decision tree consisting of two-dimensional attribute values

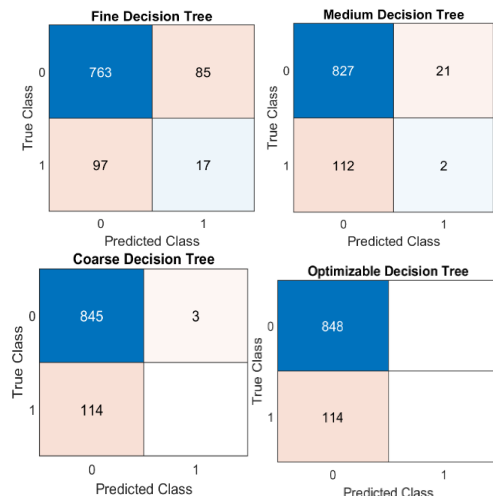


Fig. 5. Confusion matrixes of the DT models

According to the binary classification results obtained from the decision tree models, which method produces more successful results can be understood from the ROC curves shown in Figure 7. Among the four different decision tree methods used, the minimum classification error curve in Figure 8 was obtained in the ODT method that we applied and proposed in our study.

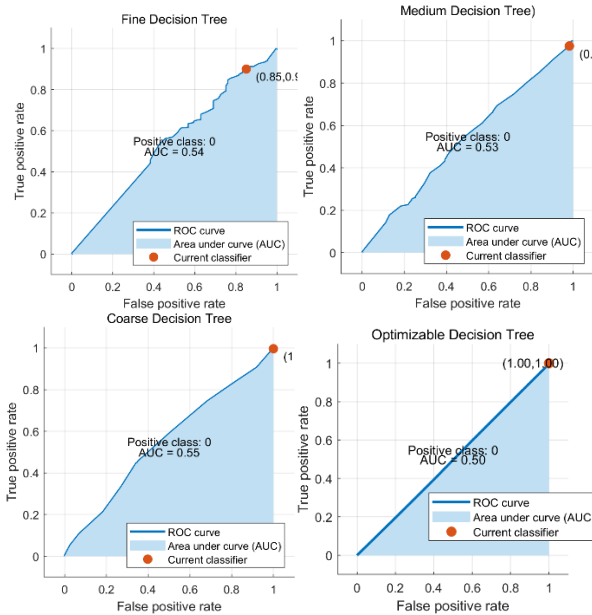


Figure 7. ROC curves of the DT models

The minimum classification error curves in Figure 8 give the difference between the observed actual minimum classification error and the estimated minimum classification error. In addition, the most optimal value point of the hyperparameters obtained as a result of the optimization is marked on this curve.

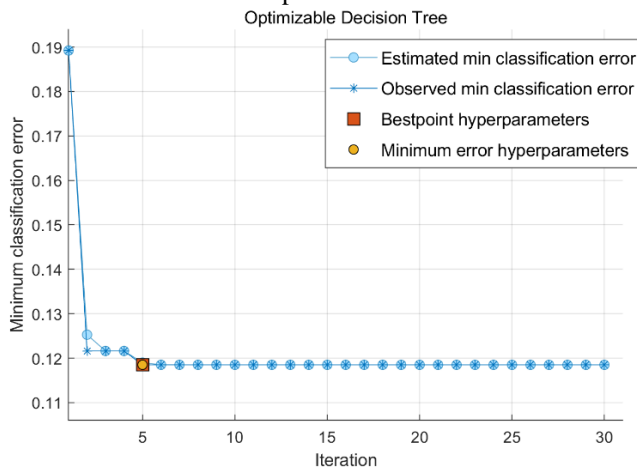


Figure 6. Minimum classification error curve of the DT model

V. CONCLUSION AND DISCUSSION

When Table 2 is examined, it is clear that the ODT model has the highest accuracy rate. Among these four decision tree techniques, the simplest ones are CDT and MDT. The training times of these simple ones are also very short compared to other

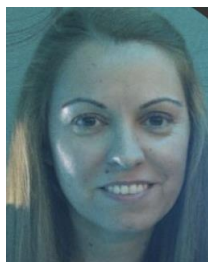
models. On the other hand, the most successful model in terms of prediction speed is FDT. The number of splits in decision trees is directly related to the number of nodes used in the tree. The low number of splits also gives us an idea about the complexity of the designed decision tree. In the ODT tree with the lowest number of splits of one, no branches are actually visible. However, we created a symbolic decision tree by creating branches for two nodes and for three nodes. The ODT model has only one split. However, despite the simplest structure, the highest accuracy rate was obtained. The CDT has only 4 split counts and still has the second highest accuracy rate of 87%. When the confusion matrices in Figure 5 are examined, it is seen that FDT produced the lowest survival estimate with 763, and ODT produced the highest survival estimate with 848. Looking at the ROC curves in Figure 8, it is seen that the AUC (Area Under the Curve) values are very close to each other. However, it is understood that the ODT method with 0.5 is more successful. As a result, survival of patients with heart disease can be predicted with high accuracy using decision tree models with 9 variables, including BMI, using some of the hospital data. If one of these decision tree models is ODT, it has been seen that the highest accuracy rate can be achieved with the most optimal decision tree hyperparameters. In the future, a more generalizable model can be designed with more datasets and deep learning methods, taking into account the frequency of food consumption, in which more variables can be added.

REFERENCES

- [1] B. Gupta, A. Rawat, A. Jain, A. Arora, N. Dhama, "Analysis of Various Decision Tree Algorithms for Classification in Data Mining", International Journal of Computer Applications (0975 – 8887) Volume 163 – No 8, April 2017.
- [2] S. J. Szydlowski, M. Luliak. Prevention of Disease-related Mortality from Chronic Non-communicable Diseases. CSWHI 2020; 11.2. 28 – 33; DOI: 10.22359/cswghi_11_2_06
- [3] Z. Zhang, S. Jackson, R. Merritt, C. Gillespie, Q. Yang, "Association between cardiovascular health metrics and depression among U.S. adults: National Health and Nutrition Examination Survey", Ann. Epidemiol. (2007-2014) 2019.
- [4] D. C. Malta, B. B. Dunca, M. I. Schmidt, R. Teixeira, A.L.P. Ribeiro, M. S. Felisbino-Mendes et al. "Trends in mortality due to non-communicable diseases in the Brazilian adult population: national and subnational estimates and projections for 2030". Popul Health Metr. 2020.18(Suppl 1).16.
- [5] P. T. Istilli, L. H. Arroyo, R. A. Dias Lima et al. Premature mortality from chronic non-communicable diseases according to social vulnerability. Mundo da Saúde 2021,45: 187-194.
- [6] C. F. Chien, L. F. Chen, "Data Mining to Improve Personnel Selection and Enhance Human Capital: A Case Study in High-Technology Industry," Expert Systems with Applications, vol. 34, 2008, pp. 280-290
- [7] S. Tsang, B. Kao, K. Y. Yip, Wai-Shing Ho, and S. D. Lee, "Decision Trees for Uncertain Data", IEEE Transactions on Knowledge and Data Engineering, vol. 23. 1, January 2011.
- [8] L. Rokach, O. Maimon, "Data Mining with Decision Trees Theory and Applications", 2nd edition, volume 81, World Scientific Publishing Co. Pte. Ltd. April 2014.
- [9] J. R. Quinlan. "C4.5: Programs for Machine Learning", Morgan Kaufmann, San Mateo, CA, 302, 1993.
- [10] G. Dougherty, "Pattern Recognition and Classification, Springer New York Heidelberg Dordrecht London", first edition, DOI 10.1007/978-1-4614-5323-9
- [11] WY. Loh and Yu-Shan Shih, "Split Selection Methods for Classification Trees, Statistica Sinica", vol. 7. 4 (October 1997), pp. 815-840

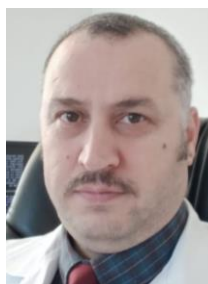
- [12] M. A. Friedl, C. E. Brodley, "Decision tree classification of land cover from remotely sensed data", *Remote Sensing of Environment*, 61, 1997, 399-409
- [13] S. R. Safavian, D. Landgrebe, "A survey of decision tree classifier methodology", *IEEE Transactions on Systems Man and Cybernetics*, 21, 1991, 660-674
- [14] PN. Tan, M. Steinbach, V. Kumar, "Introduction to Data Mining" (First Edition) (March 25, 2006) Copyright 2006, Pearson Addison-Wesley.
- [15] O. Maimon, L. Rokach, "Data Mining and Knowledge Discovery Handbook", Springer; 2nd ed. 2010.
- [16] N. Suneetha, Ch. V. M. Hari, V. Sunil Kumar, "Modified Gini Index Classification: A Case Study of Heart Disease Dataset", *International Journal on Computer Science and Engineering* Vol. 02, No. 06, 2010, 1959-1965
- [17] L. Breiman, J.H. Friedman, R.A. Olshen and C. J. Stone. 1984, "Classification and Regression Trees" Monterey, CA: Wadsworth, 358
- [18] L. E. Raileanu, and K. Stoffel, Theoretical Comparison between the Gini Index and Information Gain Criteria, *Annals of Mathematics and Artificial Intelligence* 41.1, 77-93. May 2004
- [19] K. Teknomo, "Decision Tree Tutorial", *Revoledu.com* Online edition, Last Update: October 2012
- [20] J. Mingers, "An empirical comparison of pruning methods for decision tree induction", *Machine Learning*, 4, 1989, 227-243
- [21] Z. Jingmin, et al. "Prediction model of in-hospital mortality in intensive care unit patients with heart failure: machine learning-based, retrospective analysis of the MIMIC-III database", *Dryad, Dataset*, 2021, <https://doi.org/10.5061/dryad.0p2ngf1zd>.

BIOGRAPHIES



NALAN HAKIME NOGAY Istanbul, Turkey, in 1977. He received the B.S., M.S. and Ph.D. degrees in nutrition and dietetics from Hacettepe University, Ankara, Turkey, in 1999, 2002 and 2009 respectively.

From 2009 to 2015, she was an Assistant Professor with the Nutrition and Dietetic Department. Since 2015, she has been an Associate Professor with the Nutrition and Dietetic Department, Ankara Hacettepe University. She has more than 40 papers. Her research areas are nutrition in the autistic and disabled people, obesity and community nutrition.



HIDIR SELCUK NOGAY Isparta, Turkey, in 1975. He received the B.S. , M.S. and Ph.D. degrees in electrical education from the Kocaeli University, Marmara University and Marmara University respectively in 1999, 2002 and 2008.

From 1999 to 2008, he was a Research Assistant with the Deep Learning and Electrical Machine Laboratory. From 2009 to 2011, he was an Assistant Professor with the Electrical Engineering. From 2011 to 2015, he was an Associate Professor with the Electrical Engineering. Since 2016, he has been a Professor with the Electrical Engineering Department, Kayseri University. His research interests include deep learning applications in medicine, electrical power systems, electrical machinery and renewable energy.

A Cost-Based Approach to Elicit Ancillary Service Tariffs from Unified Capacity and Energy Tariffs

Hikmet Sezer, Ali Nezhil Güven, Osman Bülent Tör, Mahmut Erkut Cebeci, and Saeed Teimourzadeh


Abstract—Establishing separated markets for capacity, energy, and ancillary services (AS) is one of the key steps to liberalize electricity sector and realize electricity markets. This study aims at eliciting AS tariffs from unified capacity and energy tariffs through a cost-based approach. In the proposed framework, maintaining annual revenue requirements of the power plants is considered as the main objective while separating the AS tariff from unified tariffs. The conducted study considers two types of unified tariffs, i.e., capacity and energy tariffs, and offers proper AS eliciting scheme for each type of unified tariffs. The main challenges addressed here are: 1) incentivizing/penalizing AS provision more/less than scheduled amount which is tackled through an incentive factor, and 2) preventing overall tariff increment (customer tariff) after eliciting AS tariff which is addressed by applying fixing constraint to the overall tariff at before and after AS tariff eliciting conditions. Last but not the least, AS trade between adjacent utilities and calculating the benefits from this trade are addressed. The effectiveness of the proposed methodology is assured through numerical investigations.

Index Terms—Ancillary Service, Balancing Market, Frequency Regulation, Capacity Tariff, Energy Tariff.


I. INTRODUCTION

ANCILLARY SERVICES and balancing facilities are countermeasures to ensure nodal power balances when


HIKMET SEZER is with Çalık Enerji, Istanbul, Turkey, (e-mail: hikmet.sezer@yesilirmakedas.com).

 <https://orcid.org/0000-0002-3628-7619>


ALI NEZHIL GÜVEN is with Middle East Technical University, Ankara, Turkey, (e-mail: guven@metu.edu.tr).

 <https://orcid.org/0000-0001-6041-7199>


OSMAN BÜLENT TÖR, is with EPRA Electric Energy Co., Ankara, Turkey (e-mail: osman.tor@epra.com.tr).

 <https://orcid.org/0000-0002-7634-2475>

MAHMUT ERKUT CEBECI is with EPRA Electric Energy Co., Ankara, Turkey (e-mail: erkut.cebeci@epra.com.tr).

 <https://orcid.org/0000-0003-2519-0309>

SAEED TEIMOURZADEH is with EPRA Electric Energy Co., Ankara, Turkey (e-mail: saeed@epra.com.tr).

 <https://orcid.org/0000-0001-8295-8404>

Manuscript received April 1, 2020; accepted Jan 30, 2022.
DOI: [10.17694/bajece.711158](https://doi.org/10.17694/bajece.711158)

the system operation parameters are violated from pre-scheduled condition [1]-[3]. Day-ahead scheduling and market clearance identify the committed generation units, determine associated operating points and allocate required reserves for the next operating day. Meanwhile, power system operator must ensure that sufficient resources are available, given any variation in system operation conditions from the day-ahead schedules, and power plants are dispatched on a moment-to-moment basis maintaining system balance and security [4]-[6].

To maintain the generation/consumption balance, the power system operator needs the ability of calling on identified generation assets or demand-side assets. AS is the service offered or facilitated by the power system operators for power quality procurement and system balancing purposes which allows multiple users to utilize the transmission network while preserving system security. The power system operators keep on hand to meet unexpected imbalances between pre-scheduled and actual load and generation. Balancing energy is the actual deployment of this capacity realized through set-aside reserve capacity which might be procured from market participants or neighbouring systems [7]. In other words, balancing services means either or both balancing capacity and balancing energy.

The language used to describe ancillary service, balancing facilities and imbalance notions is not always consistent across localities. One of the important aspects of the European Union (EU) Internal Energy Market (IEM) is to harmonize network operating practices, including how to define ancillary service and balancing facilities and imbalance notions [8]. “Balancing” is a process, while AS are diverse products to be defined and supplied for balancing. Balancing and AS are often discussed together because they are core functions for secure operation of power system. Some AS are part of the balancing function itself and designed to ensure continuous balancing of the power system. Other AS, say voltage (or reactive power) regulation, black start, are necessary for grid operation and security, but not necessarily for the balancing process [9]. This study addresses balancing aspects of AS which focuses on active power (MW) regulation capability (ramp up/down) of power plants.

This study addresses separation of AS tariffs from uniform capacity and energy tariffs for vertically integrated utilities (VIU) which are starting electricity market liberations and implementing uniform tariffs for power plants. Georgia,

Armenia and Black Sea region countries are among the countries which are at transition from VIU to liberalized electricity markets. Unified tariffs are implemented in many utilities round-the-globe, particularly in developing countries, to cover annual revenue requirements (ARR) of the power plants [10]. Eliciting AS tariffs from unified tariffs is among the most important duties of energy market regulatory authorities in these VIU as a critical step towards liberalizing electricity market opening. By the separated AS tariffs in place, 1) the true cost levels of providing AS can be identified; 2) Obeying to/violating from the pre-scheduled AS can be remunerated/penalized; 3) AS trade between the adjacent utilities can be initiated. However, the discussed benefits are in challenge with concerns such as how to incentivise AS providers; and how to prevent overall tariff increment (customer tariff) after eliciting AS tariff (see Fig. 1).

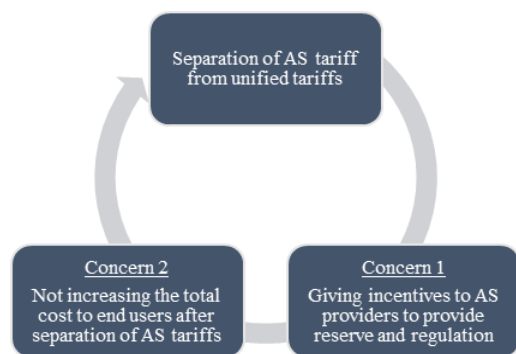


Fig.1. Concerns for separating AS tariffs from unified tariffs

Unified capacity tariffs (UCT) already include remuneration for providing AS independent from amount of AS provision. In other words, under unified tariff mechanisms, there is no motivation for the power plants to provide AS unless a penalty mechanism is introduced for the cases that the power plant refuse to provide AS. Under such a circumstance, the obvious solution is to separate the AS tariffs from the uniform tariffs and remunerate provision of AS through a separate AS tariffs mechanism. This is also required to calculate the benefits of utilities attained from coupling their ancillary service and balancing facilities services with other utilities through interconnection lines. Quantification of coupled AS provision, such as exchanging of spinning reserves between two countries, cannot be made properly unless AS tariff is separately calculated. Here, relevant studies in the literature are generally based on cost-based or market-based assumptions [11]. By separating AS tariff, those power plants who have enjoyed the remuneration of AS provision through unified tariffs will definitely be unwilling to adopt separated AS tariffs. The main reason is that such power plants will be remunerated - if and only if they provide AS satisfactorily. Therefore, an incentive mechanism should be considered in case of separated AS tariff methodology [12]. Nevertheless, such incentive mechanism should not result a significant increase in total payments to power plants as illustrated in Figure 1. Otherwise, it will be difficult for the energy market regulatory authorities to explain the tariff increment to the

end-users.

In a competitive AS market environment, price of AS provision is determined by the competition among the AS providers (referred as “market-based” approach) [13], [14]. Although it is aimed to procure AS through such market-based approaches, “cost-based” approaches are inevitable particularly at the initial phase of market opening [15]. In the cost-based approaches, price of a commodity is determined commonly based on either “marginal-based” or “pay-as-bid” mechanisms [16]. In “pay-as-bid” mechanisms, cost-based approaches should consider a profit component in determining ARR of the power plants [17]. Profit component in pay-as-bid mechanisms is more critical than that of marginal-based mechanisms given that the market clearing at marginal price already provides a room for profit of power plants unless they are marginal.

Unified capacity and energy tariffs of power plants are generally determined annually using cost-based approaches [18]. Each year, power plants provide associated costs (fixed and variable) to energy market regulatory authorities. Afterwards, the ARRs for power plants are settled after some negotiations between the power plants and energy market regulatory authorities. ARR is encompasses fixed costs (i.e., costs which are independent of power generation amount), variable costs (costs associated with power generation amount), depreciation costs, and profit component [19]. It is worth mentioning that cost-based approaches are also being utilized in developed markets when a market power/abuse is observed [18]. In such a case, energy market regulatory authorities identify the price based on cost-based approaches. Hence, separated AS tariffing is not an interim solution until market opening indeed and can be used in case of liberalized electricity market.

This paper offers a cost-based approach to elicit the AS tariffs from the unified tariffs. In the proposed framework, maintaining annual revenue requirements of the power plants are considered as the main objective while separating the AS tariff from unified tariffs. The conducted study considers two types of unified tariffs, i.e., capacity and energy tariffs, and offers proper AS eliciting scheme for each type of unified tariffs. The main challenges addressed here are: 1) incentivizing/penalizing AS provision more/less than scheduled amount which is tackled through an incentive factor, and 2) preventing overall tariff increment (customer tariff) after eliciting AS tariff which is addressed by applying fixing constraint to the overall tariff at before and after AS tariff eliciting conditions [20]. Last but not the least, AS trade between adjacent utilities and calculating the benefits from this trade are addressed. The proposed approach is tested through numerical analysis for generic power plants. Fixed and variable cost parameters of the generic power plants are given in [19], [21].

II. PROPOSED METHODOLOGY

In this section, first, the proposed methodologies for eliciting AS tariff from capacity and energy tariffs are presented. Next, the proposed scheme for AS trade between

adjacent utilities and calculating the benefits of this trade is presented. The following assumptions are made in separating AS tariffs from unified tariffs:

- AS tariff is constant for two different type of spinning reserves: 1) Primary reserves, (frequency containment reserve FCR); and 2) Secondary reserves (frequency restoration reserve; FRR) [22]. In other words, there is a single AS tariff both for FCR and FRR. However, provision of FCR and FRR are remunerated separately, based on the allocated reserve amount as illustrated in Fig. 2.
- AS providers are assumed to allocate spinning reserves based on power system operator requirements which are responsible from identifying total amount of spinning reserve requirements in term of FCR and FRR, separately.
- AS tariff is symmetrical in terms of upward and downward activation of FCR and FRR. For instance, if a power plant is allocated for providing 10% FCR, associated upper level for dispatch can be 90% of its capacity.

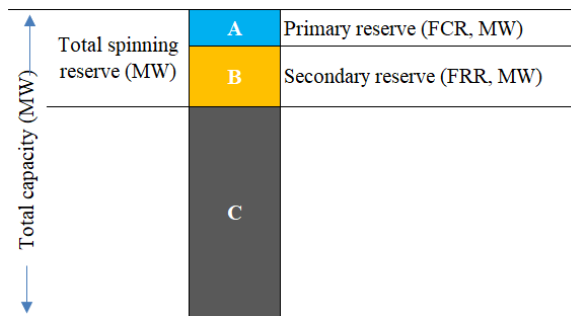


Fig.2. Spinning reserve concept in the proposed AS tariff mechanism

- Settling time is assumed to be one hour. That is, ancillary service tariff is in \$/MW/h.
- No separate capacity tariff for tertiary reserves (i.e., replacement reserve, RR) and RRs are procured from the balancing market.
- It is assumed that upward and downward activation of FCR cancel each other gradually, as illustrated in Fig. 3. Therefore, there is no additional remuneration for the energy generated during the activation of FCR in the upward direction.
- In contrast to FCR, the amount of energy activated to provide FRR in both upward and downward directions, is subjected to balancing market. That is, the net energy balance of the AS providers should be remunerated in the balancing market, as illustrated in Fig. 3.

Figure 3 presents chronological activation of FCR, FRR, and RR after losing a generation unit in a typical power grid. Subsequent to this event, the frequency immediately starts decreasing (alteration #1 in Fig. 3). Here, the generation/demand balance is tried to be recovered through activating FCR in upward direction (#2). The frequency is restored to the pre-disturbance level by activating the FRR (#3) and RR (#6) in upward direction, respectively. This process ensures restoration of the FCR and FRR (downward activation) to their initial values as well. Although deployment of unified energy tariffs (\$/MWh) is a common practice, some

countries implement capacity tariffs (\$/MW/h) in addition to energy tariffs. Capacity tariff mechanisms are generally implemented to mitigate security and reliability concerns [18]. In economics, variable costs and fixed costs are the two main costs a company has when producing goods and services. A variable cost varies with the amount produced, while a fixed cost remains fix independent from the output. It is common to determine capacity tariffs with respect to the fixed-cost driven ARR of the power plants.

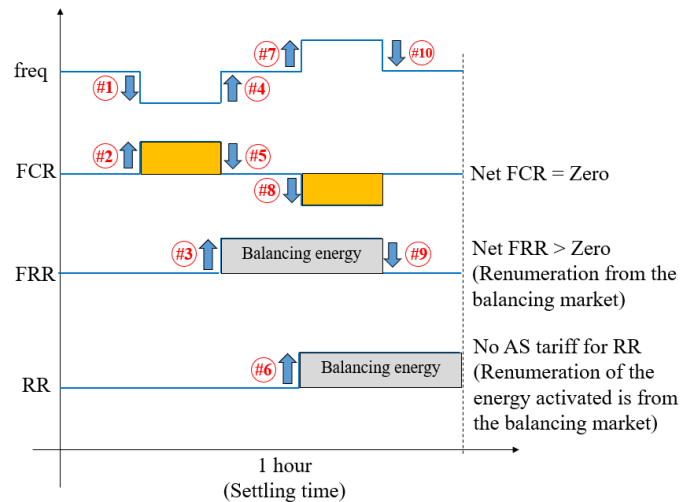


Fig. 3 Sequential activation of FCR, FRR, and RR after loss of generation.

Given the above-discussed assumptions, the proposed methodologies for eliciting AS tariff from unified energy and capacity tariff are addressed in the following:

A. Eliciting AS Tariff from Unified Capacity Tariff (UCT)

Eliciting AS tariff from UCT (\$/MW/h) is investigated in this section. The proposed methodology is based on fixed-costs driven ARR of the power plants. Figure 4 depicts the outline of the proposed approach.

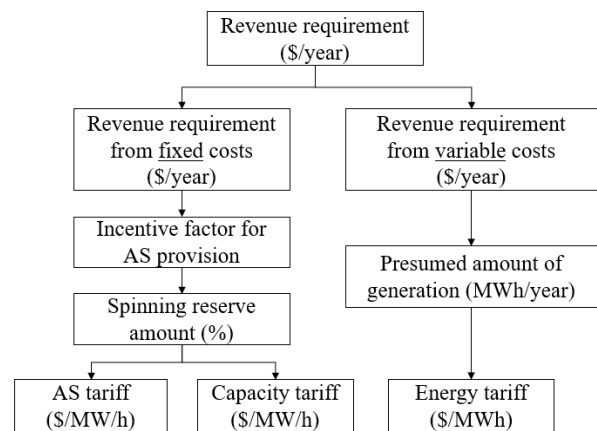


Fig.4. Spinning reserve concept in the proposed AS tariff mechanism

Separating AS tariff from the existing UCT should be made in a way that the power plants, which are supposed to provide AS reserve (either FCR, FRR, or both), should be remunerated if and only if they allocate and activate the AS reserve as planned by the transmission system operator (TSO). In this

respect, the following planning process by the TSO:

- TSO should determine total amount of AS reserve (MW) required for the grid in terms of FCR and FRR. To do so the common practice is to determine them on a monthly basis at the initial phase of the market opening.

- Total AS reserve should be allocated to power plants based on their availability and technical capability to provide AS according to the grid code of the country. This requires a prequalification and certification process [23].

Unified cost-based capacity tariffs are determined by the regulatory authority based on ARR of the power plants. Fixed-costs include the costs associated with the provision of AS services. The power plants are remunerated for their available capacity through their UCTs, as described by (1):

$$FC_i = \sum_{t=1}^{12} (UCT_i \times AVC_i^t) \quad (1)$$

where,

- i Generator index
- t Month index
- FC_i Annual fixed cost of the generator i (\$/year)
- UCT_i Unified capacity tariff of generator i (\$/MW)
- AVC_i^t Average available capacity of generator i at month t (A+B+C in Fig. 2) (MW)

AS tariff of each power plant is separated from the UCT in a way that, if it provides AS reserves as allocated; then, the summation of the payments from AS tariff and recalculated capacity tariff should be equal to payment from the current UCT. Mathematically speaking:

$$FC_i = \sum_{t=1}^{12} (CT_i \times AVCC_i^t + AST_i \times AVC_i^t) \quad (2)$$

$$AVC_i^t = AVCC_i^t + ASC_i^t \quad (3)$$

where ,

- CT_i Capacity tariff of generator i for $AVCC_i^t$ after separation of AS tariff (\$/MW)
- $AVCC_i^t$ Average available capacity excluding AS reserve (C in Fig. 2) (MW)
- AST_i Ancillary service tariff of generator i (\$/MW)
- ASC_i^t Average available AS capacity of generator i at month t (A + B in Fig. 2) (MW)

Constant FC_i satisfies the second concern indicated above (i.e., increase of overall tariffs after separation of AS tariffs). Incentive factor, namely K , is for the second concern; to give incentives to those generators which are asked to provide more AS than planned, as illustrated in Fig. 5. The incentive factor K represents the slope in Fig. 5 which can be calculated as:

$$AST_i = K \times CT_i \quad (4)$$

If the slope is equal to one, total payment to generators is same as the case of UCT given by $CT_i = AST_i$. Here, the higher slope results in a greater gap between the AST_i and CT_i . The difference between AST_i and CT_i makes sense in terms of total payment to generators if they are asked to provide more AS than that of planned. In such a case, total payment to generators will increase. However if the power plant provides

less than planned, associated income will reduce which is calculated by (5). The incentive factor should be negotiated and settled between the power plants and ISO. Recalculation of the capacity tariff after separation of AS tariff is presented by (6).

$$FC_i = CT_i \times \sum_{t=1}^{12} (AVCC_i^t + K \times ASC_i^t) \quad (5)$$

$$CT_i = \frac{FC_i}{\sum_{t=1}^{12} (AVCC_i^t + K \times ASC_i^t)} \quad (6)$$

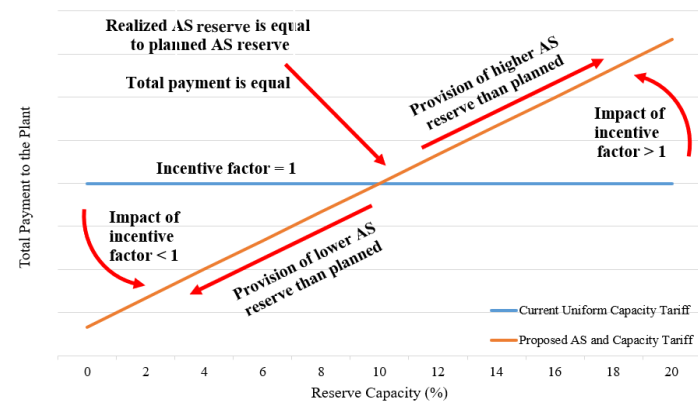


Fig. 5 Impact of incentive factor K for 10% allocated capacity

B. Eliciting AS Tariff from Unified Energy Tariff (UET)

Eliciting AS tariff from UET (\$/MWh) is investigated in this section. The proposed methodology for separation of AS tariff from UCT (Section II.A) is modified as illustrated in Fig. 6. The methodology again starts by classifying fixed- and variable-costs of power plants and calculating fixed-cost based ARR of the power plants. However, since there is no capacity tariff, AST covers only fixed-costs which correspond to spinning reserves provision. The remaining fixed-costs should be converted to energy tariff as indicated in Fig. 6.

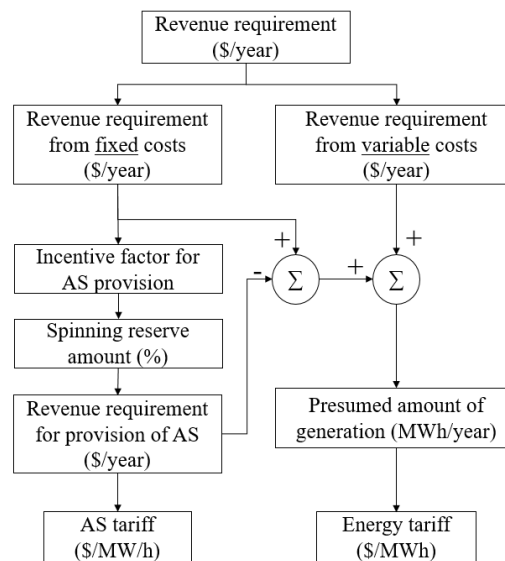


Fig.6. Flowchart of proposed methodology for eliciting AS tariff in case of UET

Equations (7)-(9) express the calculation process of UET , AST , and re-calculation of energy tariff after separation AST , respectively:

$$UET_i = \frac{FC_i + VC_i}{E_i} \quad (7)$$

$$AST_i = \frac{K \times FC_i}{E_i \times (1 - ASR_i) + ASR_i \times K} \quad (8)$$

$$ET_i = \frac{FC_i + VC_i - AST_i \times E_i \times ASR}{E_i} \quad (9)$$

where,

VC_i	Annual variable cost of the generator i (\$/year)
E_i	Total amount of energy presumed to be produced by generator i (MWh/year)
ET_i	Energy tariff of power plant i (\$/MWh) after separation of AST

To summarize the proposed approaches, Fig. 4 depicts the outline of the proposed approach for eliciting ancillary service from the tariff when unified capacity mechanism is applied. In the contrary, Fig. 6 represents eliciting ancillary service from the tariff when unified energy mechanism is applied. In brief, Fig 4 deals with capacity, however, Fig. 6 deals with energy.

III. NUMERICAL INVESTIGATIONS AND DISCUSSIONS

This section testifies effectiveness of the proposed method through numerical studies.

A. Eliciting AS Tariff from UCT

1) Illustrative Example 1 - Calculation of UCT

This example deals with calculation of UCT , CT , and AST for a generic thermal power plant for different values of incentive factor, K . Essentially, annual fixed costs of the power plant include those costs associated with the provision of AS services plus some profit to ensure that the power plants are satisfied with the tariffs. Table I and II represent the main parameters and fixed and variable costs of a generic thermal power plant. Note that profit ratio should be determined by the regulator. Here, profit ratio is assumed to be 10% and UCT is calculated as 3.52 \$/MW/h through (10):

$$UCT = 7,392,000.00(\$) / 300(MW) / 7,000 (h) \quad (10)$$

$$= 3.52 \$ / MW / h$$

TABLE I
MAIN PARAMETERS OF A GENERIC THERMAL POWER PLANT

Type of power plant	Thermal
Plant capacity (MW)	300
Availability hours per year (8760 - maintenance hours)	7,000
Full load hours per year	2,500
Estimated energy (MWh/year)	750,000 (300*2,500)
Annual operation hours (capacity factor)	3,333

2) Illustrative Example 1 - Calculation of AS Tariff

After calculating UCT , AS tariff should be calculated. Given that the generic power plant is committed to provide 10%

spinning reserve at real-time operation. CT and ACT are calculated using (1)-(6) for different incentive factors, K , as presented in Table III. For the calculations presented in Table III, it is assumed that the power plant allocates and activates 10% AS at the satisfactory level. Therefore, although total payment to the power plant from AS tariff increases along with incentive factor, total payment remains fix (i.e., capacity payment plus AS payment). Such a behavior ensures the second concern in Fig. 1. Note that, capacity payment is significantly higher than that of AS in this example. The main reason is that CT is subjected to 90% of the power plant capacity during available hours (7,000 h); whereas, ACT is subjected to only 10% capacity of the power plant for operation hours (3,333 h), as presented in Table I.

TABLE II
FIXED AND VARIABLE COSTS OF A GENERIC POWER PLANT

Cost item	Fixed Cost (k\$/year)	Variable Cost (k\$/year)
Salary fund	1,500.00	-
Depreciation / amortization of production facilities	1,750.00	-
Expense on basic and supporting materials	200.00	-
Cost of operational maintenance	800.00	-
Fuel cost (natural gas, oil, coal, etc.)	-	25,000.00
Chemicals	-	10.00
Technical water consumption costs	-	15.00
Electricity utility and internal consumption costs	200.00	-
Utility costs (except electricity)	10.00	-
Communication expenses	10.00	-
Office cost	10.00	-
Business trips	5.00	-
Advisory costs	500.00	-
Security costs	50.00	-
Rental costs	5.00	-
Insurance expense	400.00	-
Market operator service fee	70.00	-
Commission regulation fee	60.00	-
property tax	200.00	-
Property tax on land	50.00	-
Other operating costs	100.00	-
Interest rate cost	650.00	-
Other non-operational costs	150.00	-
Total cost per year (k\$/year)	6,720.00	25,025.00
Profit (10%)	672.00	2502.50
Total cost for tariff calculation (k\$/year)	7,392.00	27,527.50

TABLE III
CT AND ACT CALCULATIONS FOR DIFFERENT INCENTIVE FACTORS

Incentive factor (K)	1	2	4
AS Reserve (Planned / Provided)	10% / 10%	10% / 10%	10% / 10%
CT (\$/MW/h)	3.52	3.36	3.08
ACT (\$/MW/h)	3.52	6.72	12.32
Capacity Payment (k\$/year)	7,040	6,720	6,160
AS Payment (k\$/year)	352	672	1,232
Total Payment (AS+Capacity) (k\$/year)	7,392	7,392	7,392

3) Illustrative Example 1 – Violating from scheduled amount of AS at real-time operation

Assume that the power plant is planned to provide 10% AS reserve; however, lower amount is realized in real-time operation. Consequently, associated revenue from AS tariff will considerably reduce. On the contrary, if asked to provide more than planned amount, associated revenue will increase with respect to the incentive factor, as presented in Table IV for the case of $K=2$.

TABLE IV
SCHEDULED VS PROVIDED ($K=2$)

Incentive factor (K)	2	2	2
AS Reserve	Planned: 10% Provided: 10%	Planned: 10% Provided: 15%	Planned: 10% Provided: 5%
Capacity Tariff (\$/MW/h)	3.36	3.36	3.36
AS Capacity Tariff (\$/MW/h)	6.72	6.72	6.72
Capacity Payment (k\$/year)	6,720	6,552	6,888
AS Payment (k\$/year)	672	1,008	336
Total Payment (AS+Capacity) (k\$/year)	7,392	7,560	7,224

4) Illustrative Example 2 – More than one power plant contribute to AS

In a case that a power plant is planned for AS provision; however, couldn't realize associated commitment in real-time operation, TSO should attain the required reserve from other power plants. In this illustrative example, such a condition is investigated through two different cases. In Case 1, five generic power plants (see Fig. 7) are considered to provide satisfactory level of spinning reserves which are subjected to provide 5% spinning reserve (Case 1). The scheduled and provided AS for these power plants are illustrated in Fig. 8. In Case 2, it is assumed that generic power plant B refuses to provide AS at a satisfactory level, and hence, generic power plant F is providing 5% more spinning reserve given that both plants have the same capacity as illustrated in Fig. 8 (Case 2). *UCT*, *CT*, and *AST* of the power plants are presented in Fig. 7. Incentive factor, K , is assumed to be 2.

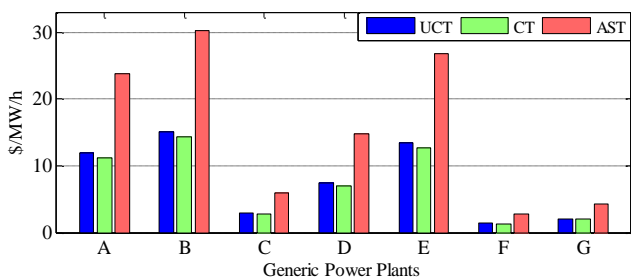


Fig.7. *UCT*, *CT*, and *AST* of the generic power plants ($K=2$)

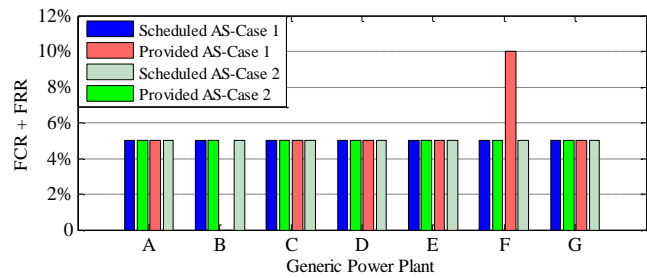


Fig.8. Scheduled versus provided AS reserve by the generic power plants (Cases 1 and 2)

Note that generic power plant F is selected to provide additional AS owing to low *AST* in comparison to the other power plants of this example. It is worth mentioning that tariffs of the power plants are disregarded while scheduling and allocating spinning reserve requirement among the power plants. Such an indicator could be considered; however, it might be out of scope of conducted study.

Annual payments to power plants for Case 1 and 2 are represented in Fig. 9 and Fig. 10, respectively. Total payment to power plant B reduces in Case 2 comparing to Case 1 as power plant B is excluded from *AST* payment. In Case 2, total payment to the plant F is more than that of Case 1 since the incentive factor greater than 1, i.e. $K = 2$.

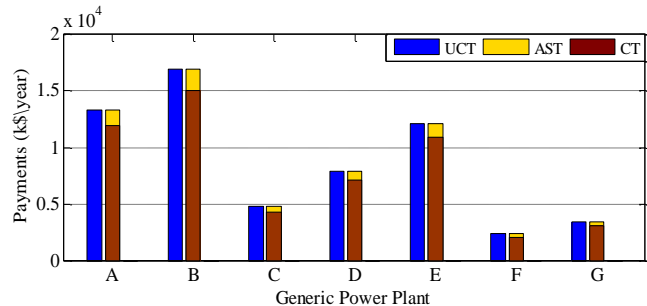


Fig.9. Annual revenue from *UCT*, *CT*, and *AST* in Case 1

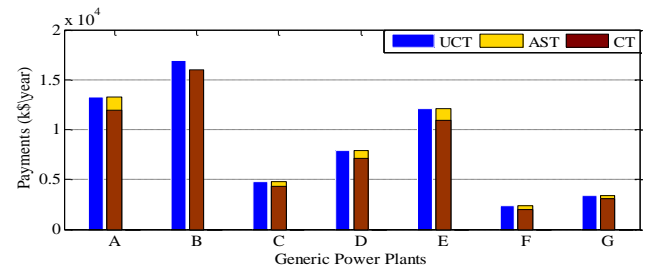


Fig.10. Annual revenue from *UCT*, *CT*, and *AST* in Case 2

B. Eliciting AS Tariff from UET

In this example, the same condition as that of illustrative examples presented in Section III.A are considered; unless, the power plants are assumed to be subjected to *UET*. That is ARR of the power plant is remunerated through *UET*. In this numerical example, $FC= 27,528.00$ k\$/year; $VC= 7,392.00$ k\$/year; $E= 750,000$ MWh/year. *UET* of the generic power plant is 4.66 \$-cent/kWh which can be calculated as:

$$UET = \frac{27528 + 7392}{750000} \times 100 = 4.66\$ - \text{cent/kWh} \quad (11)$$

AST of the power plant is 0.99 \$-cent/kW/h (12):

$$AST = \frac{K \times FC_i}{E_i \times (1 - ASR_i) + ASR_i \times K} = 0.99 \text{ \$-cent/kW/h} \quad (12)$$

where, *ASR* is ancillary service in per units. Finally, energy tariff will be 4.56 \$-cent/kWh (13). As can be seen, after separation of *AST*, *ET* is lower than *UET* (i.e., 4.56 < 4.66).

$$ET_i = \frac{(FC_i + VC_i) - AST_i \times E_i \times ASR_i}{E_i} = 4.56 \text{ \$-cent/kW/h} \quad (13)$$

C. Calculation of Benefits of Countries from AS Coupling

In this section, numerical analyses are performed to calculate benefits of two countries from coupling their AS and balancing mechanisms, based on the proposed cost-based AS tariffs eliciting approach. The following direct benefits are addressed:

- **Common Dimensioning:** It is an opportunity for TSOs to jointly dimensioning the total volume of balancing reserve, as illustrated in Fig. 11.

- **Exchange of Reserve:** It is the opportunity for TSOs to procure part of the balancing energy in another area (with no impact on the total amount of balancing reserve in the system) as illustrated in Fig. 11. The scenarios investigated here are summarized in Table V. In Fig. 11, three different conditions are depicted. First, each region holds spinning reserves based on their own need. In the common dimensioning, both side are holding 150 MW spinning reserve as a common reserve. Finally, for exchanging reserve case, the 20 MW reserve is exchanged between the regions.

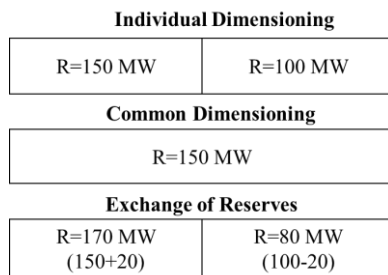


Fig.11. Direct benefits from coupling of AS (R: amount of allocated spinning reserve)

TABLE V
SCENARIOS INVESTIGATED IN SECTION III.C

Scenario	Condition	AS and Balancing Coupling
1	No interconnection between countries A and B	-
2	Interconnection between country A and B	Common dimensioning
3	Interconnection between country A and B	Exchange of AS reserves

1) Key Assumptions

Given two interconnected countries, five power plants each with 100 MW capacity are assumed in both countries. Each power plant provides 20% spinning reserve, totally 100 MW, to compensate for loss of any power plant in the countries.

Available hours of the power plants are assumed to be 7,000 h. The following annual parameters are assigned differently among the power plants to differentiate their cost-based tariffs: i) In-service hours, ii) Total fixed costs. More details of the assumptions are described in the Sections III.C.2, III.C.3, and III.C.4 on a country basis.

Parameters of the power plants and associated *CT* and *AST* are provided in Tables VI and VII for Country A, and Tables VIII and IX for Country B, respectively. Incentive factor, *K*, is assumed to be 1.5 in all power plants. Additional cost of power plants due to provision of AS is assumed to be 10% of corresponding fixed costs.

TABLE VI
PARAMETERS OF THE POWER PLANTS – COUNTRY A

Parameter	G.A.1	G.A.2	G.A.3	G.A.4	G.A.5
Plant Capacity (MW)	100	100	100	100	100
Estimated Availability Hours (h)	7,000	7,000	7,000	7,000	7,000
In service hours along the year (h)	3,300	6,000	5,000	3,000	5,000
FC (k\$/year)	2	2,500	3	2,500	3,500
Additional costs related to AS (%)	10%	10%	10%	10%	10%
Profit (10%)	10%	10%	10%	10%	10%
Spinning reserve (%)	20%	20%	20%	20%	20%
ARR from CT and AST (k\$/year)	2,420	3,025	3,630	3,025	4,235

TABLE VII
TARIFF CALCULATIONS – COUNTRY A

Parameter	G.A.1	G.A.2	G.A.3	G.A.4	G.A.5
UCT (\$/MW/h)	3.46	4.32	5.19	4.32	6.05
<i>K</i>	1.5	1.5	1.5	1.5	1.5
CT (\$/MW/h)	3.30	3.98	4.84	4.14	5.65
AST (\$/MW/h)	4.95	5.97	7.26	6.22	8.47

TABLE VIII
PARAMETERS OF THE POWER PLANTS – COUNTRY B

Parameter	G.B.1	G.B.2	G.B.3	G.B.4	G.B.5
Plant Capacity (MW)	100	100	100	100	100
Availability hours (h)	7,000	7,000	7,000	7,000	7,000
In service hours (h)	3,300	6,000	5,000	3,000	5,000
FC (\$/year)	1	1,250	1,500	1,250	1,750
Additional costs related to AS (%)	10%	10%	10%	10%	10%
Profit (10%)	10%	10%	10%	10%	10%
Spinning reserve (%)	20%	20%	20%	20%	20%
ARR from CT and AST (k\$/year)	1,210	1,513	1,815	1,513	2,118

TABLE IX
TARIFF CALCULATIONS – COUNTRY B

Parameter	G.A.1	G.A.2	G.A.3	G.A.4	G.A.5
UCT (\$/MW/h)	1.73	2.16	2.59	2.16	3.03
<i>K</i>	1.5	1.5	1.5	1.5	1.5
CT (\$/MW/h)	1.65	1.99	2.42	2.07	2.82
AST (\$/MW/h)	2.48	2.99	3.63	3.11	4.24

CT and *AST* of power plants in Country B are less than those of Country A as illustrated in Fig. 12. Difference is mainly due to fixed cost parameters of the power plants (see Table VI and Table VIII).

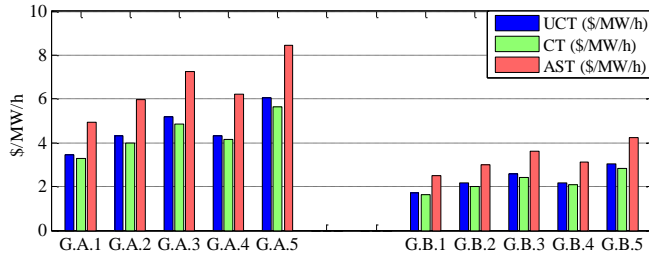


Fig.12. CT and AST of power plants – Country A vs Country B

2) Scenario 1 – No interconnection between Country A and Country B

In case of no interconnection between the countries, the payments to power plants located at Countries A and B from CT and AST are calculated as presented in Table X and Table XI, respectively.

TABLE X
CALCULATIONS OF PAYMENTS FOR CT AND AST – COUNTRY A

Parameter	G.A.1	G.A.2	G.A.3	G.A.4	G.A.5
Capacity Payment from CT (k\$/year)	2,093	2,308	2,904	2,652	3,388
Total Payment from CT (k\$/year)	13,346				
AS Payment from AST (k\$/year)	326	716	726	372	847
Total Payment from AST (k\$/year)	2,989				
Total Payment for CT and AST (k\$/year)	2,420	3,025	3,630	3,025	4,235
Total Payment for CT and AST (k\$/year)	16,335				

TABLE XI
CALCULATIONS OF PAYMENTS FOR CT AND AST – COUNTRY B

Parameter	G.A.1	G.A.2	G.A.3	G.A.4	G.A.5
Capacity Payment from CT (k\$/year)	1,047	1,154	1,452	1,326	1,694
Total Payment from CT (k\$/year)	6,673				
AS Payment from AST (k\$/year)	163	358	363	186	423
Total Payment from AST (k\$/year)	1,495				
Total payment for CT and AST (k\$/year)	1,210	1,512	1,815	1,512	2,117
Total payment for CT and AST (k\$/year)	8,168				

3) Scenario 2 – Common Dimensioning

In this scenario, it is assumed that two countries are interconnected and they implement common dimensioning for their AS and balancing requirements. After common dimensioning, the amount of AS allocation in both countries reduces to 50 MW representing total spinning reserve of 100 MW. Each country is assumed to reduce spinning reserve amounts from those power plants which have the highest AST. This reduction will occur in G.A.5, G.A.3, and G.A.2 in Country A, and G.B.5, G.B.3, and G.B.4 in Country B given their high relatively high AST, as illustrated in Table XII and Table XIII, respectively. The tables also present recalculated

values of ARR to be recovered from CT and AST.

TABLE XII
RECALCULATION OF PARAMETERS - COUNTRY A

Parameter	G.A.3 (1)	G.A.3 (2)	G.A.4 (1)	G.A.4 (2)	G.A.5 (1)	G.A.5 (2)
Plant capacity (mw)	100	100	100	100	100	100
Estimated Availability Hours per year (h)	7,000	7,000	7,000	7,000	7,000	7,000
In service hours along the year (h)	5,000	5,000	3,000	3,000	5,000	5,000
Total cost for CT calculation (k\$/year)	3,000	3,000	2,500	2,500	3,500	3,500
Costs related to AS provision (%)	10%	0%	10%	10%	10%	0%
Costs related to AS provision (k\$/year)	300	-	250	250	350	-
Profit (10%)	10%	10%	10%	10%	10%	10%
Spinning reserve (%)	20%	0%	20%	10%	20%	0%
Total spinning reserve amount (MW)	20	-	20	10	20	-
ARR from CT and AST per year (k\$/year)	3,630	3,300	3,025	3,025	4,235	3,850

TABLE XIII
RECALCULATION OF PARAMETERS - COUNTRY A

Parameter	G.A.3 (1)	G.A.3 (2)	G.A.4 (1)	G.A.4 (2)	G.A.5 (1)	G.A.5 (2)
Plant capacity (mw)	100	100	100	100	100	100
Estimated Availability Hours per year (h)	7,000	7,000	7,000	7,000	7,000	7,000
In service hours along the year (h)	5,000	5,000	3,000	3,000	5,000	5,000
Total cost for CT calculation (k\$/year)	1,500	1,500	1,250	1,250	1,750	1,750
Costs related to AS provision (%)	10%	-	10%	10%	10%	-
Costs related to AS provision (k\$/year)	150	-	125	125	175	-
Profit (10%)	10%	10%	10%	10%	10%	10%
Spinning reserve (%)	20%	0%	20%	10%	20%	0%
Total spinning reserve amount (MW)	20	-	20	10	20	-
ARR from CT and AST per year (k\$/year)	1,815	1,650	1,513	1,513	2,118	1,925

Since G.A.3, G.A.5, G.B.3, and G.B.5 are not providing AS, their total annual fixed cost reduces since AS provision payment is excluded. This reduction will essentially reduce their CTs, as illustrated in Fig. 13. Note that, although these power plants will not provide AS, their capacity connected with AS in Scenario 1 should be still be remunerated through the CT in Case 2. That is, total capacity subjected to CT in Case 2 increases if compared to Case 1 by the amount of spinning reserve allocated in Case 1, as illustrated in Fig. 14. Nevertheless, total payments of country A and B for CT and AST reduces as presented in Fig. 15. This result shows that total savings of Countries A and B from common dimensioning is 1,073 k\$/year for this particular generic example. Note that ARRs of G.A.4 and G.B.4 from CT and

AST remained fixed given the assumption that costs related to AS provision do not change along with reduction of AS amount from 20% to 10%.

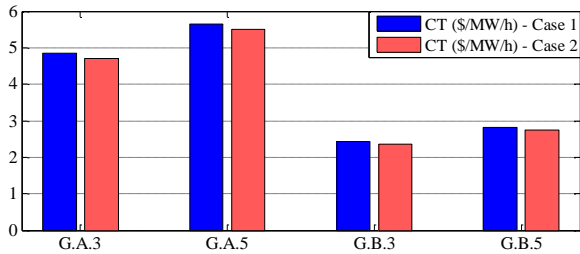


Fig. 13. CT of G.A.3, G.A.5, G.B.3, and G.B.5 (Case 1 vs Case 2)

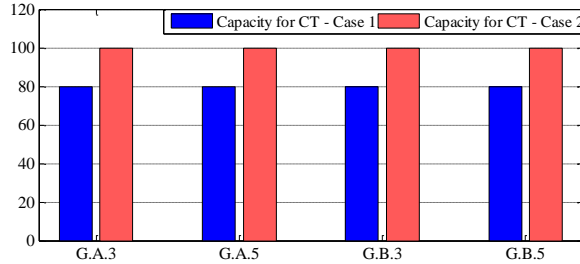


Fig. 14. Capacity of G.A.3, G.A.5, G.B.3, and G.B.5 subjected to CT (Case 1 vs Case 2)

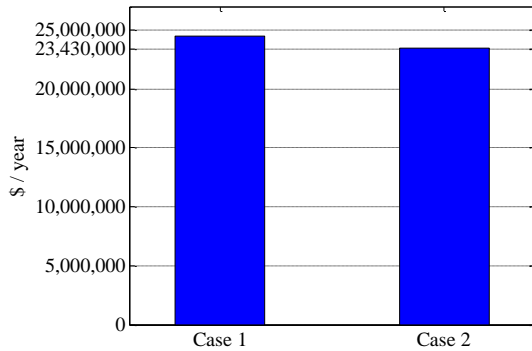


Fig. 15. Total payment for CT and AST (Country A + Country B)

4) Scenario 3 – Exchange of AS

In this scenario, it is assumed that Countries A and B are exchanging 20 MW spinning reserve. The exchange is between G.A.1 and G.B.1. That is, G.A.1 is not providing any spinning reserve while G.B.1 is providing additional 20 MW to compensate for. This exchange still ensures 100 MW total spinning reserve in total (i.e., Country A plus Country B). Recalculated values of ARRs, tariffs, and payments are provided in Table XIV-Table XVI for Country A, and Table XVII-Table XIX for Country B, respectively.

Benefits of the countries from AS and coupling mechanisms are presented in Fig. 16. One can argue that the benefit is not observable and relatively small. However in the numerical analysis, normalized fixed costs (\$/MW) of the power plants are assumed almost similar in order to be realistic (fixed cost range is taken from the literature [19], [21]). In addition, profit component of the power plants (10%) and incentive factor ($K=1.5$) are assumed to be same. Such parameters can result in higher benefits than calculated in the generic numerical examples. Nevertheless, main focus is given in the study to

propose an AS tariff separation methodology which satisfies the main concerns mentioned above.

TABLE XIV
RECALCULATION OF PARAMETERS - COUNTRY A

Parameter	G.A.1	G.A.2	G.A.3	G.A.4	G.A.5
Plant Capacity (MW)	100	100	100	100	100
Estimated Availability Hours per year (h)	7,000	7,000	7,000	7,000	7,000
In service hours along the year (h)	3,300	6,000	5,000	3,000	5,000
Total cost for CT calculation (\$/year)	2,000	2,500	3,000	2,500	3,500
Costs related to AS provision (%)	-	10%	-	10%	-
Costs related to AS provision (\$/year)	-	250	-	250	-
Profit (10%)	10%	10%	10%	10%	10%
Spinning reserve (%)	-	20%	-	10%	-
Total spinning reserve amount (MW)	-	20	-	10	-
ARR from CT and AST per year (\$/year)	2,200	3,025	3,300	3,025	3,850

TABLE XV
RECALCULATION OF CT AND AST - COUNTRY A

Parameter	G.A.1	G.A.2	G.A.3	G.A.4	G.A.5
CT (\$/MW/h)	3.14	3.98	4.71	4.23	5.50
AST (\$/MW/h)	-	5.97	-	6.35	-

TABLE XVI
RECALCULATION OF PAYMENT - COUNTRY A

Parameter	G.A.1	G.A.2	G.A.3	G.A.4	G.A.5
Capacity Payment from CT (k\$/year)	2,200	2,309	3,300	2,835	3,850
Total Payment from CT (k\$/year)	14,493				
AS Payment from AST (k\$/year)	-	716	-	190	-
Total Payment from AST (k\$/year)	907				
Total Payment for CT and AST (k\$/year)	2,200	3,025	3,300	3,025	3,850
Total Payment for CT and AST (k\$/year)	15,400				

TABLE XVII
RECALCULATION OF PARAMETERS - COUNTRY B

Parameter	G.A.1	G.A.2	G.A.3	G.A.4	G.A.5
Plant Capacity (MW)	100	100	100	100	100
Estimated Availability Hours per year (h)	7,000	7,000	7,000	7,000	7,000
In service hours along the year (h)	3,300	6,000	5,000	3,000	5,000
Total cost for CT calculation (\$/year)	1,000	1,250	1,500	1,250	1,750
Costs related to AS provision (%)	10%	10%	-	10%	-
Costs related to AS provision (\$/year)	10%	10%	10%	10%	10%
Profit (10%)	20%	20%	-	10%	-
Spinning reserve (%)	1,210	1,513	1,650	1,513	1,925
Total spinning reserve amount (MW)	100	100	100	100	100
ARR from CT and AST per year (\$/year)	7,000	7,000	7,000	7,000	7,000

TABLE XVIII
RECALCULATION OF CT AND AST - COUNTRY B

Parameter	G.A.1	G.A.2	G.A.3	G.A.4	G.A.5
CT (\$/MW/h)	1.65	1.99	2.36	2.12	2.75
AST (\$/MW/h)	2.48	2.99	-	3.17	-

TABLE XIX
RECALCULATION OF PAYMENT - COUNTRY B

Parameter	G.A.1	G.A.2	G.A.3	G.A.4	G.A.5
Capacity Payment from CT (k\$/year)	938	1,154	1,650	1,417	1,925
Total Payment from CT (k\$/year)	7,084				
AS Payment from AST (k\$/year)	229	358	-	95	-
Total Payment from AST (k\$/year)	682,840.28				
Total Payment for CT and AST (k\$/year)	1,167	1,513	1,650	1,513	1,925
Total Payment for CT and AST (k\$/year)	7,767				

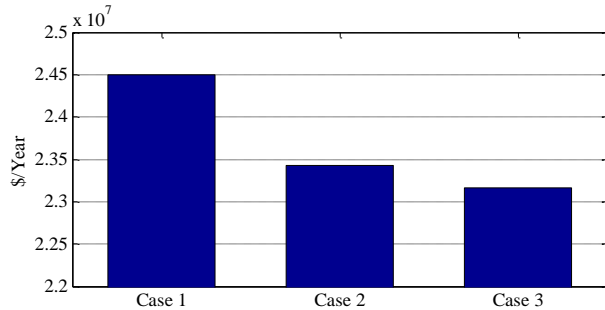


Fig.16. Total payment for CT and AST (Country A + Country B)

IV. DISCUSSIONS AND CONCLUSIONS

This study dealt with eliciting AS tariffs from unified capacity and energy tariffs through a cost-based approach. In the proposed framework, maintaining annual revenue requirements of the power plants are considered as the main objective while separating the AS tariff from unified tariffs. The conducted study addressed the following challenges: 1) incentivizing/penalizing AS provision more/less than scheduled amount which is tackled through an incentive factor, and 2) preventing overall tariff increment (customer tariff) after eliciting AS tariff which is addressed by applying fixing constraint to the overall tariff at before and after AS tariff eliciting conditions. Main concern of power plants in providing AS is the additional costs connected with providing active power regulation and loss of opportunity due to allocation of some reserves of AS provision. On the other hand, main concern of transmission system operator and electricity market regulatory authorities is remunerating the power plants which do not provide AS, through unified tariffs. Numerical case studies demonstrated that the proposed approach satisfies such concerns. In addition, electricity market regulatory authorities of the countries, which implement unified capacity and/or energy tariffs, can consider the proposed tariff separation approach for determining true cost level of their country in procuring AS. Therefore, such

countries do not need to wait until market opening to start benefitting from AS coupling with neighbour countries. The proposed approach enables them to quantify their benefits.

ACKNOWLEDGMENT

This paper presents the scientific results of the transnational project “Multi-layer aggregator solutions to facilitate optimum demand response and grid flexibility” (SMART-MLA Project No: 89029), co-financed by TÜBİTAK (Turkey), SWEA (Sweden), EUDP (Denmark), UEFISCDI (Romania), and RCN (Norway) under ERA-Net Smart Energy Systems, SG+ 2017 Program. The study is supported by “TÜBİTAK TEYDEB 1509 - Uluslararası Sanayi Ar-Ge Projeleri” program with the project number of 9180003.

REFERENCES

- [1] Y. G. Rebours, D. S. Kirschen, M. Trotignon, and S. Rossignol, "A survey of frequency and voltage control ancillary services—Part I: Technical features," *IEEE Trans. Power Syst.*, vol. 22, no. 1, pp. 350-357, 2007.
- [2] N. A. El-Taweel, H. Khani, and H. E. Z. Farag, "Optimal Sizing and Scheduling of LOHC-Based Generation and Storage Plants for Concurrent Services to Transportation Sector and Ancillary Services Market," *IEEE Transactions on Sustainable Energy*, vol. 11, pp. 1381-1393, 2020.
- [3] D. Jay and S. S. K., "Game Theoretical Approach to Novel Reactive Power Ancillary Service Market Mechanism," *IEEE Transactions on Power Systems*, pp. 1-1, 2020.
- [4] A. Zobian and M. D. Ilic, "Unbundling of transmission and ancillary services. I. Technical issues," *IEEE Trans. Power Syst.*, vol. 12, no. 2, pp. 539-548, 1997.
- [5] M. Vanouni and N. Lu, "A Reward Allocation Mechanism for Thermostatically Controlled Loads Participating in Intra-Hour Ancillary Services," *IEEE Transactions on Smart Grid*, vol. 9, pp. 4209-4219, 2018.
- [6] G. D. Zotti, S. A. Pourmousavi, H. Madsen, and N. K. Poulsen, "Ancillary Services 4.0: A Top-to-Bottom Control-Based Approach for Solving Ancillary Services Problems in Smart Grids," *IEEE Access*, vol. 6, pp. 11694-11706, 2018.
- [7] Y. G. Rebours, D. S. Kirschen, M. Trotignon, and S. Rossignol, "A survey of frequency and voltage control ancillary services—Part II: Economic features," *IEEE Trans. Power Syst.*, vol. 22, no. 1, pp. 358-366, 2007.
- [8] L. Meeus, K. Purchala, and R. Belmans, "Development of the internal electricity market in Europe," *The Electricity Journal*, vol. 18, no. 6, pp. 25-35, 2005.
- [9] K. Bhattacharya and J. Zhong, "Reactive power as an ancillary service," *IEEE Trans. Power Syst.*, vol. 16, 2, pp. 294-300, 2001.
- [10] R. Raineri, S. Rios, and D. Schiele, "Technical and economic aspects of ancillary services markets in the electric power industry: an international comparison," *Energy policy*, vol. 34, no. 13, pp. 1540-1555, 2006.
- [11] O. B. Tör, M.E. Cebeci, M. Stojkovic, D. Orlic, C. Godfrey, G. Wray, *et al.*, "Sub-Regional Direct Benefits from Cross-Border Provision of Balancing Energy and Reserves at Black Sea Region Based on Market-Based and Cost-Based Approaches," presented at the CIGRE RSEEC October 2018.
- [12] E. Ela, A. Tuohy, M. Milligan, B. Kirby, and D. Brooks, "Alternative approaches for incentivizing the frequency responsive reserve ancillary service," *The Electricity Journal*, vol. 25, no. 4, pp. 88-102, 2012.

- [13] H. Singh and A. Papalexopoulos, "Competitive procurement of ancillary services by an independent system operator," *IEEE Trans. Power Syst.*, vol. 14, no. 2, pp. 498-504, 1999.
- [14] M. B. Zammit, D. J. Hill, and R. J. Kaye, "Designing ancillary services markets for power system security," *IEEE Trans. Power Syst.*, vol. 15, no. 2, pp. 675-680, 2000.
- [15] I. De la Fuente, D. Soler, G. Relafío, O. Largo, T. Gómez, R. Martínez, *et al.*, "F. Remuneration Framework and Cost Allocation Methodology for AGC Ancillary Services Markets," in *Power System Computation Conference PSCC*, 1999.
- [16] A. Zobian and M. D. Ilic, "Unbundling of transmission and ancillary services. II. Cost-based pricing framework," *IEEE Trans. Power Syst.*, vol. 12, no. 2, pp. 549-558, 1997.
- [17] S. F. Tierney, T. Schatzki, and R. Mukerji, "Uniform-pricing versus pay-as-bid in wholesale electricity markets: does it make a difference?," *New York ISO*, 2008.
- [18] J. E. Bowring, "The evolution of PJM's capacity market," in *Competitive Electricity Markets*, ed: Elsevier, 2008, pp. 363-386.
- [19] U. EIA, "Capital cost estimates for utility scale electricity generating plants," *US Department of Energy, Energy Information Administration. Parts on PHP and batteries* http://www.eia.gov/analysis/studies/powerplants/capitalcost/pdf/capcost_assumption.pdf, 2016.
- [20] M. Scherer, M. Zima, and G. Andersson, "An integrated pan-European ancillary services market for frequency control," *Energy policy*, vol. 62, pp. 292-300, 2013.
- [21] R. Tidball, J. Bluestein, N. Rodriguez, and S. Knoke, "Cost and performance assumptions for modeling electricity generation technologies," National Renewable Energy Lab.(NREL), Golden, CO (United States)2010.
- [22] E. N. o. T. S. Operators, "Network Code on Load-Frequency Control and Reserves 6," 2013.
- [23] B. Kirby, "Ancillary services: Technical and commercial insights," *Retrieved October*, vol. 4, p. 2012, 2007.



Mahmut Erkut Cebeci received his B.S. and M.S. degrees from Middle East Technical University (METU), Ankara, Turkey, in 2005 and 2008, respectively. He is working as managing partner and director at EPRA Electric Energy Co., Ankara, Turkey.



Saeed Teimourzadeh received the Ph.D. degree in power electrical engineering from School of Electrical and Computer Engineering, University of Tehran, Tehran, Iran, in 2018. From 2016 to 2017, he was a Research Associate with the Electrical and Computer Engineering Department at the Illinois Institute of Technology (IIT), Chicago, IL, USA. He is currently R&D Director with EPRA Electric Energy Co., Ankara, Turkey. Dr. Teimourzadeh served the Journal of Modern Power Systems and Clean Energy (MPCE) as the editor and received the 2018 excellence in reviewing award from MPCE Journal. His research interests include microgrid protection, control and stability, and smart grid initiatives.

BIOGRAPHIES



Hikmet Sezer received his BS degree from Middle East Technical University (METU), Turkey, in 1991. He is working as Chairman of the Board at Yeşilirmak Elektrik Dağıtım A.Ş., İstanbul, Turkey.



Ali Nezh Güven received the B.S. degree from Middle East Technical University (METU), in 1979 and the M.S. and Ph.D. degrees from the Ohio State University, USA in 1981 and 1984, respectively. He is currently a professor in the Department of Electrical and Electronics Engineering at METU, Ankara, Turkey. His research interests are the analysis, design, and operation of power systems, and distribution automation.



Osman Bülent Tör received his B.S., M.S. and Ph.D. degrees from Middle East Technical University (METU), Turkey, in 1998, 2001 and 2008, respectively. He is working as managing partner and director at EPRA Electric Energy Co., Ankara, Turkey.

The Effect of Road Slope on Parameters of Disabled Vehicle Driven by PMDC Motor

Fatih Alpaslan Kazan * and Ramazan Akkaya

Abstract— In this study, the effect of road slope on the electrical and mechanical parameters of a disabled vehicle and regenerative braking was investigated on a 4 wheel disabled vehicle driven by a permanent magnet direct current (PMDC) motor. For this purpose, the test system, which was designed by the authors in order to test such vehicles in a laboratory environment, was developed in a way that allows the vehicle to be tested in external environments. This test system consists of a data acquisition card (DAC), related sensors, and an interface prepared in C# program. Using the DAC, currents, voltages, speed, road slope, and temperatures were measured instantaneously during the test process and then transferred to the computer instantly. Using these data, the motor's speed, the motor's torque, the motor's shaft power, the yields of the motor and the driver, the vehicle's instantaneous speed, and the total covered distance were calculated during the experiment in real-time by the interface. At the same time, these data were graphed and saved. Tests were carried out on roads with a positive and negative slope. Finally, the data obtained from these tests were analyzed and the alterations in the electrical and mechanical parameters of the disabled vehicle depending on the road slope were clearly demonstrated by numerical data.

Index Terms— Data acquisition card, disabled vehicle, PMDC motor, regenerative braking, road slope, measurement.

I. INTRODUCTION

A VEHICLE THAT some or all of the mechanical energy required for its movement is provided from the electric motor(s) is called an electric vehicle. While determining the powers of the motors to be mounted on electric vehicles, the climbing angle of the vehicle and some forces affecting the vehicle are taken into account. These forces that oppose the

movement of the vehicle; aerodynamic (air) resistance, rolling resistance, acceleration resistance and gradient resistance.


The effect of aerodynamic resistance occurs when the vehicle's speed exceeds about 50 km/h [1]. The rolling resistance consists of shape changes in roads and tires during rolling of the wheels and is expressed by a coefficient. The main reason for this resistance is that the tire is constantly deformed. It is also possible to reduce this with the appropriate wheel selection and tire pressure. Acceleration resistance occurs when the vehicle moves from the first speed to a second speed. This resistance is calculated by multiplying the vehicle's mass and the vehicle's instant acceleration. Therefore, this resistance value is zero for constant speed applications. Gradient resistance occurs due to the component of the weight parallel to the road during the movement of the vehicle on the inclined road. This resistance occurs on inclined roads [2].

Another point that is taken into consideration when determining the powers of the motors to be mounted on the vehicles is the vehicle's climbing ability. The maximum angle of inclination that a vehicle can climb at any constant speed is defined as the "ability to climb" of that vehicle.

Similar points need also to be taken into account when determining the power of the motor(s) to be installed in the battery-operated disabled vehicle, which are the biggest helpers of elderly and disabled people in carrying out their daily activities. In the study carried out by Akpunar [3], the power of a motor designed to carry 120 kg load on the battery-operated disabled vehicle was calculated by considering the above-mentioned points. As a result of the calculations, it has been found that the motor to be mounted to the vehicle should produce 16.25 Nm, 10 Nm and 3.75 Nm of torque on + 20°, 0° and -20° roads, respectively. This means that every 1° change in road slope causes a change by 3.125% in torque. The literature was searched to find similar studies examining the change of the vehicle's parameters in relation to the road slope in the battery-operated disabled vehicles.


In the literature searches, it has been seen that very different studies have been carried out regarding the battery-operated disabled vehicles. Some of these are: controlling the disabled vehicle with head movement [4]-[9], voice [10]-[17], eye [18]-[24] and mouth movement [25]-[27]; using a fuzzy logic controller instead of conventional controller [28], [29]; examination of the vibration of the disabled vehicle under different road conditions [30]; measuring the mechanical and electrical impedances of the disabled vehicle with a sensor

FATİH ALPASLAN KAZAN, is with Department of Aviation Electrical and Electronic of Selçuk University, Konya, Turkey, (e-mail: akazan@selcuk.edu.tr).

 <https://orcid.org/0000-0002-5461-0117>

*Corresponding Author.

RAMAZAN AKKAYA, is with Department of Electrical and Electronic Engineering of Konya Technical University, Konya, Turkey, (e-mail: rakkaya@ktun.edu.tr).

 <https://orcid.org/0000-0002-6314-1500>

Manuscript received May 9, 2021; accepted Jan 28, 2022.

DOI: [10.17694/bajece.935156](https://doi.org/10.17694/bajece.935156)

[31]; use of efficient motors to increase the vehicle's range [32]- [34]; detection of road slope [35]-[39]; examination of the effect of road gradient on the energy consumption of electric cars [40], [41]; investigation on the effect of different user weights on the energy consumption of the disabled vehicle driven by permanent magnet synchronous motor (PMSM) [42]; design of dc-dc buck converter for electric vehicles [43], [44]; maximizing of regenerative braking energy recovery of electric vehicles [45]. However, an experimental study that examines the change of vehicle parameters depending on the road slope and revealing this change clearly was not found in a disabled vehicle driven by permanent magnet direct current (PMDC) motor.

In the study previously conducted by the authors and presented in [42], the disabled vehicle was driven by PMSM (650 W), and the experiments were carried out on 2 constant slope routes with average slopes of 0.21° and 1.44° , respectively. In the study where the effect of different user weights on the energy consumption of the vehicle was investigated, only the battery parameters were taken into consideration. The change of the parameters of the motor depending on the road slope or the regenerative braking effect was not investigated. In this new study, the changes of vehicle parameters, such as vehicle speed, battery current, battery voltage, the power drawn from the battery, motor's shaft power, and system efficiency, depending on the continuously changed road slope of the disabled vehicle driven by PMDC

motor were demonstrated by experimental studies. In addition, it was also made an experimental study about regenerative braking becomes continuous at how many degrees.

II. DESIGNED MEASURING SYSTEM AND COMPONENTS

A system that enables such disabled vehicles to be tested in the laboratory was previously introduced by the authors [46]. In this test system, battery's and motor's currents (I_b and I_m), motor's and battery's voltages (V_m and V_b), temperatures of motor and ambient (T_m and T_a), and vehicle speed (v) were measured instantly and then sent to the computer.

These data sent to the computer were both monitored and recorded in real-time in the interface prepared in C#. In addition, using these data, the unmeasured quantities such as the speed of the motor (N_m), the power drawn from the battery (P_b), the motor's shaft power (P_s), the torque produced by the motor (T), and the motor and driver efficiencies (Eff_m and Eff_d) were also calculated instantly and plotted. Details of all these calculations were given in detail in [46]. The principle diagram of the new system with PMDC motor is given in Fig. 1. The rated voltage, the shaft power, and the speed of the PMDC motor in the disabled vehicle are 24 V, 750 W, and 4000 rpm respectively. The motor is controlled by a commercial driver with the code D51445 produced by PG Drives for use in disabled vehicles. The driver uses the PWM method.

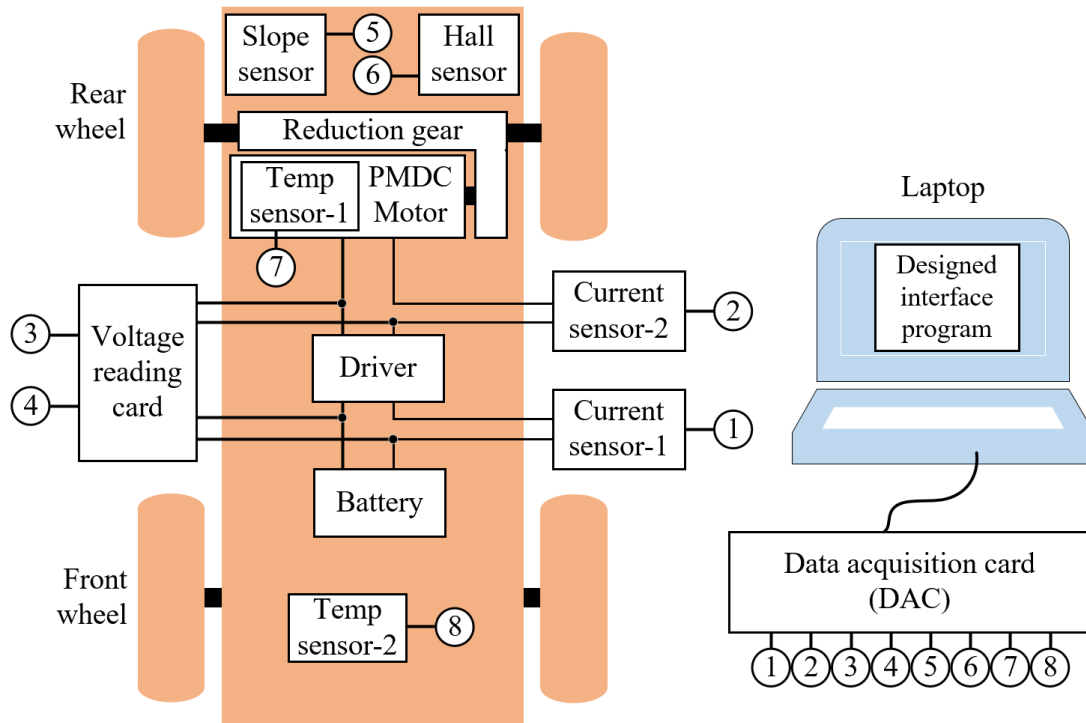


Fig. 1. The principle diagram of the designed test system.

A. Voltage Measurement

LV25P type voltage sensors manufactured by the LEM have been used to measure battery and motor terminal voltages. As

is known, this sensor is a sensor that changes the output current according to the magnitude of the voltage at its input. The measurement is done by measuring the voltage that the

current drops on a resistor connected to the output. The circuit diagram and final form of voltage reading card designed for measuring battery and motor terminal voltages are shown in Fig. 2.

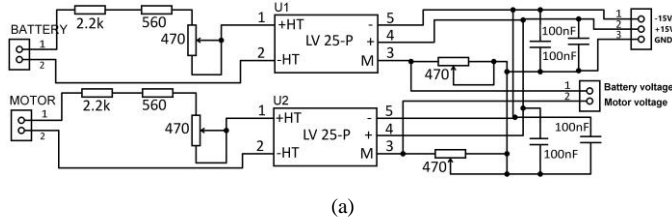


Fig. 2. The voltage reading card. (a) The open-circuit diagram. (b) Final form of the card.

B. Current Measurement

ACS712 series 30A current sensors (Fig. 3) have been used to measure battery and motor currents. It will also be possible to observe the regenerative braking effect as they allow bidirectional current measurement. As it is known, since these sensors give analogue output and can be supplied as a complete module, it is sufficient to connect their outputs directly to an analogue channel of a microcontroller.

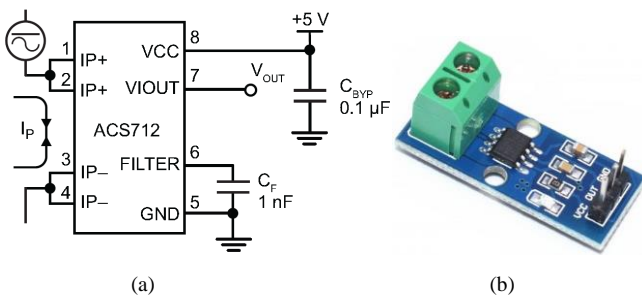


Fig. 3. ACS712-30A current sensor. (a) The open-circuit diagram [47] (b) Final form of the card.

C. Measurement of Road Slope

SCA61T-FAHH1G coded incline sensor manufactured by the company named Murata has been used to measure the slope of the road on which the vehicle was tested. The measuring limit of this sensor is $\pm 30^\circ$. This sensor, which has a sensitivity of 0.0025° , can give analogue output. The sensor produces an oscillating output when exposed to accelerated motion or when used on a vibrating surface. The circuit

diagram of the card designed for measuring the road slope is given in Fig. 4.

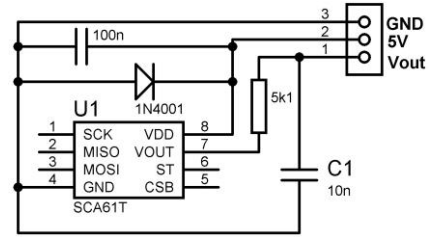


Fig. 4. The circuit diagram of the card designed for measuring the road slope

D. Measurement of Vehicle Speed

Vehicle speed plays a key role in calculating many parameters of the vehicle. Therefore, it is a quantity that must be measured. In the vehicle's outdoor tests, Hall sensor and magnets have been used to measure the speed. In this context, 24 equally spaced magnets have been placed on the rear right wheel of the vehicle to be tested. The A1101 series Hall sensor manufactured by the company named Allegro is positioned so that the distance between them is below 1 cm when it aligns with the magnets. The placement of the magnets and the Hall sensor are seen in Fig. 5.

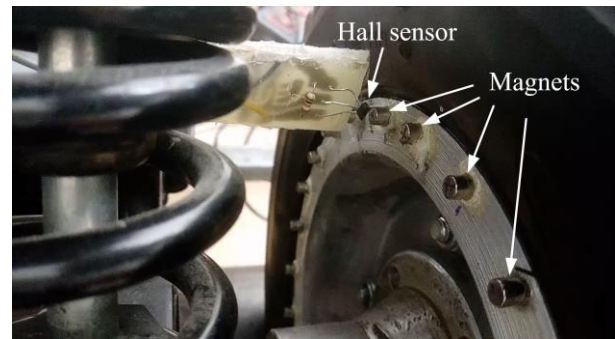
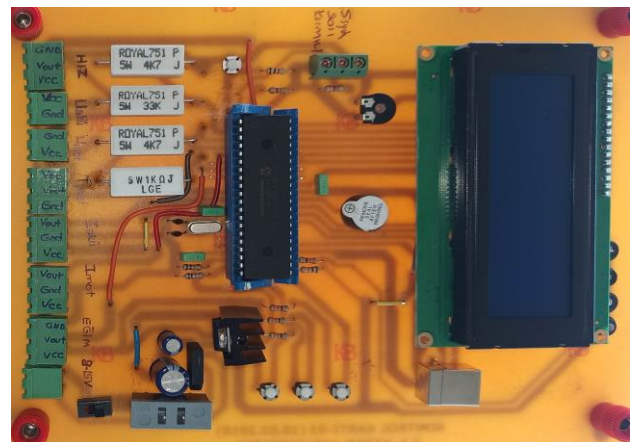


Fig. 5. Placement of the magnets and Hall sensor.

E. Data Acquisition Card

A data acquisition card (DAC) was designed to read the parameters and then to transfer them to the computer in real-time via USB. Low-priced PIC18F4550 was preferred as the microcontroller on the DAC and operated at 48 MHz by using PLL (Phase Locked Loop) feature. The designed DAC is shown in Fig. 6.



(a)

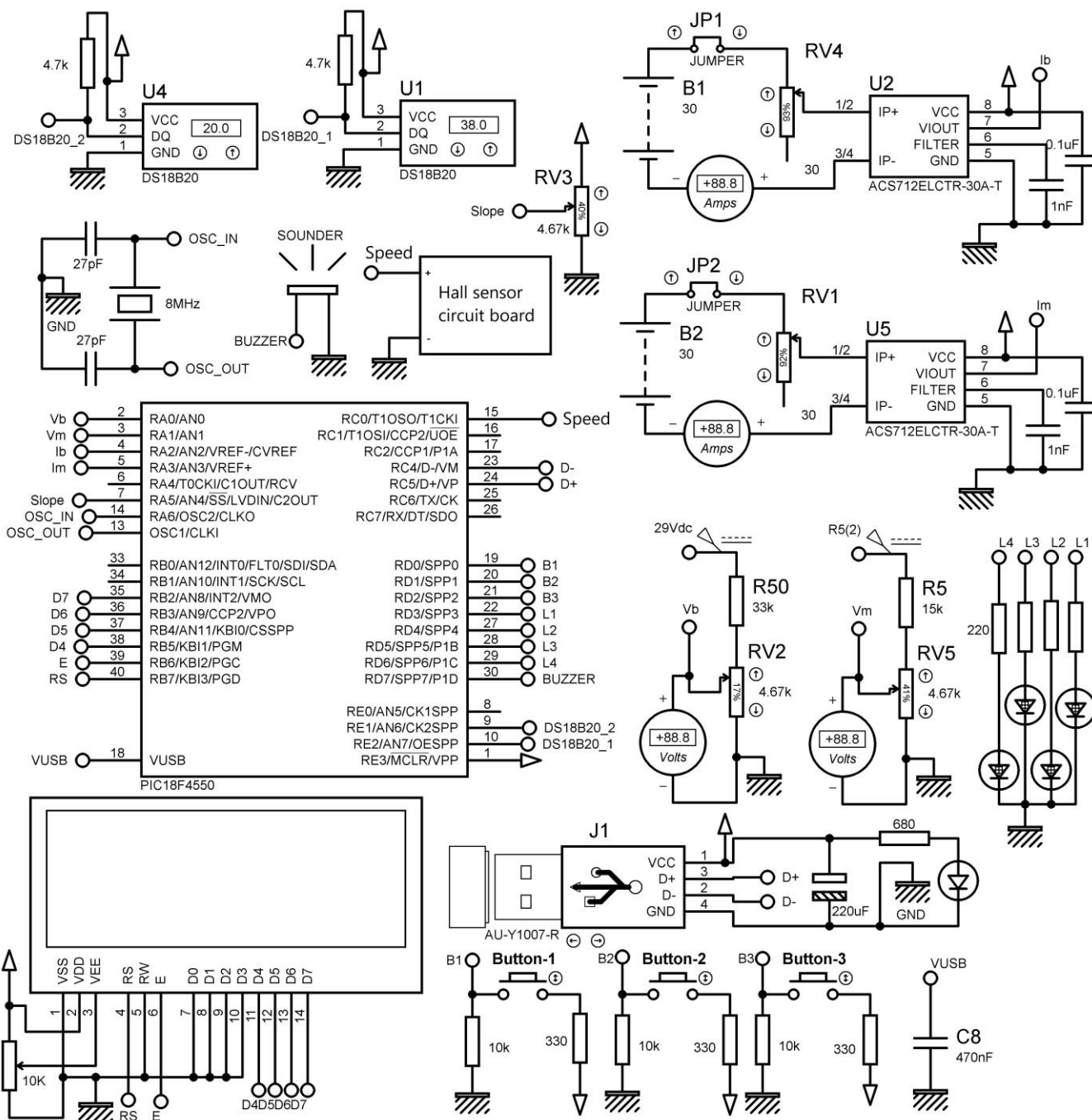


Fig. 6. The DAC. (a) Final form of the DAC. (b) The open-circuit diagram for simulation.

F. Interface Program

An interface program was designed using Visual C # for viewing, processing, and storing the data sent by the DAC on the computer. The screenshot taken at the moment of the simulation of the DAC from the designed interface program is given in Fig. 7.

The interface program consists of 4 main sections. These are parameters section to be entered, measured and calculated values section, table section, and graphic section. How other quantities obtained using the measured data are calculated, is

explained in detail in the study previously introduced by the authors in [46]. Therefore, it will not be explained here again. However, since the road slope was also measured in this study, minor revisions were made to the interface. The number of graphics drawn in the graphics section has also been increased from 10 to 13.

Another innovation to the interface is the constant updating of the torque constant value entered into the interface, taking into account the test data from the PMDC motor manufacturer installed in the vehicle. The update process was done using an equation that was obtained by the curve fitting method and

embedded in the interface program. Thus, the efficiency of the motor was obtained with an average error of 0.98% and without using any torque sensor. This situation is also seen in

the efficiency graphs, which are presented in Fig. 8 and obtained when the motor is loaded at 7 different levels.

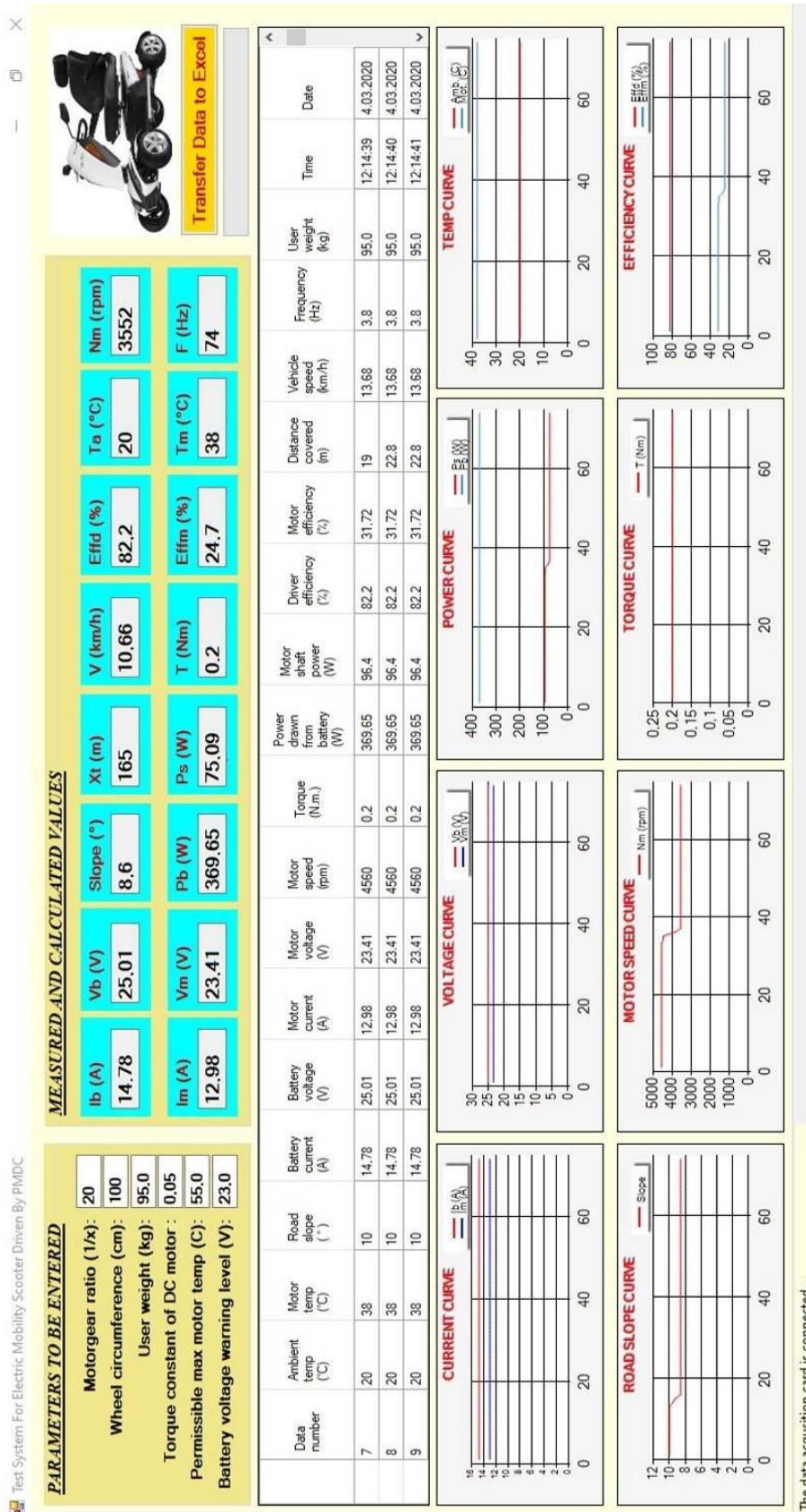


Fig. 7. Designed interface program.

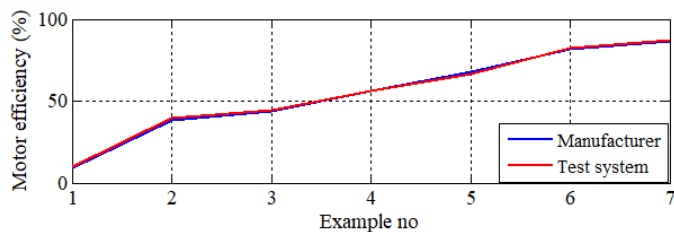


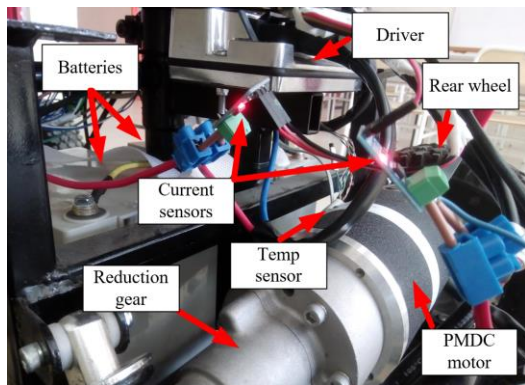
Fig. 8. Efficiency graphs of the PMDC motor in the disabled vehicle

G. The Layout of System Components on the Disabled Vehicle

The layouts of the system components on the vehicle is given in Fig. 9. The position of the designed DAC and voltage reading card can be seen in Fig. 9.a. In Fig. 9.b shows the rear of the disabled vehicle and the equipment located there.



(a)



(b)

Fig. 9. The layouts of the system components on the vehicle. (a) The position of the designed DAC and voltage reading card. (b) The rear of the disabled vehicle and the equipment located there.



Fig. 10. Route-1.

III. EXPERIMENTAL STUDIES

In order to investigate the effect of the slope on the disabled vehicle driven with a PMDC motor, tests were carried out on different roads with different slopes at the Alaeddin Keykubat Campus of Selçuk University. First of all, tests which made on roads with positive slopes will be presented.

A. Experimental Studies on Positive Slope Roads

The first test was carried out on Route-1, shown in Fig. 10. During the test, the direction of the vehicle's movement is from point A to point B.

Changes in road slope, vehicle speed, current and power drawn from the battery, and battery terminal voltage throughout the test at Route-1 are given in Fig. 11.

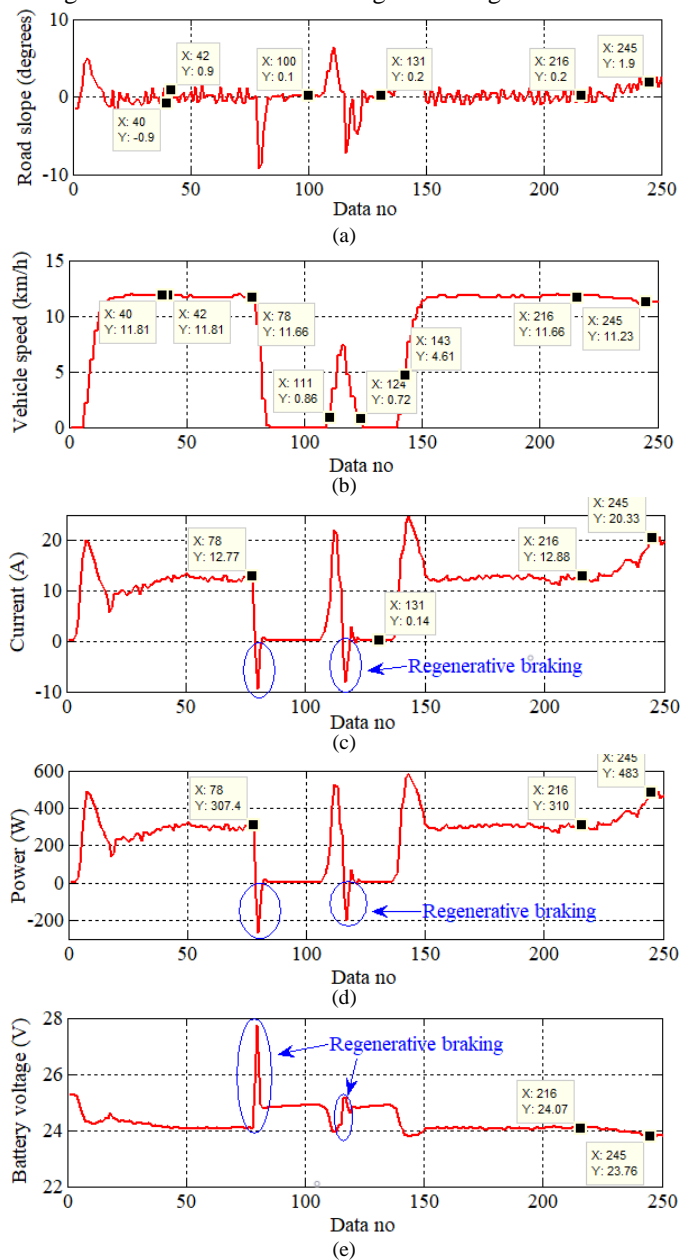


Fig. 11. Graphs related to the test on Route-1. (a) Road slope, (b) Vehicle speed, (c) Battery current, (d) Power drawn from the battery, (e) Battery terminal voltage. (Time difference between two consecutive data 0.4748 s).

The incline sensor produces oscillating results due to the accelerated and vibrating movement of the vehicle. When the slope and vehicle speed graphs in Fig. 11 are examined together: Although the tilt sensor produces a result well above and below normal during acceleration and deceleration, it produces an oscillating output that results only from the vibration of the vehicle but is relatively low compared to that of accelerated motion in areas where the speed is constant.

For example, when looking at the 40th and 42nd data in the speed graph, it is seen that the speed is constant (11.81 km/h). When looking at the same numbered data in the slope graph, it is seen that the slope is $\pm 0.9^\circ$, so there is a difference of 1.8° from peak to peak. This indicates that if the vehicle does not accelerate, there is only an error caused by the vehicle's vibration and the error value is approximately $\pm 0.9^\circ$ compared to the normal value. However, if the vibration of the sensor can be prevented, then a non-oscillating (absolute) result is obtained. This situation can be easily seen if the 100th and 131st data regions, where the vehicle is not moving, are examined. When the 100th and 131st data in the slope graph are taken into consideration, it is seen that the slopes are 0.1° and 0.2° , respectively.

In the slope graph in Fig. 11.a, it is seen that the road slope, which has an average of 0.2° , started to increase after 216th data and reached an average of 1.9° at the end of the experiment. The effect of this 1.7° increase in slope on vehicle speed, current drawn from the battery, average of instantaneous power drawn from the battery (P_b) and battery voltage is given in Table 1.

TABLE 1. SUMMARY INFORMATION OF ROUTE-1.

α ($^\circ$)	V_b (V)	I_b (A)	v (km/h)	P_b (W)
0.2	24.07	12.88	11.66	310
1.9	23.76	20.33	11.23	483

As can be find out from Table 1, the rising of road slope from 0.2° to 1.9° caused a decrease in battery voltage and vehicle speed by 1.29% and 3.69%, respectively. However, the battery current and the average of instantaneous power drawn from the battery rose 57.8% and 55.8%, respectively.

Testing on a road with an increasing slope will allow seeing the effect of the slope on parameters more clearly. For this purpose, the test was performed on Route-2, which shown in Fig. 12. Here, too, the direction of the vehicle's movement is from point A to point B.



Fig. 12. Route-2

Changes in vehicle speed, road slope, battery current, battery voltage, the power drawn from the battery, and shaft power throughout the test at Route-1 are given in Fig. 13. In this test, only the behavior of the vehicle in the steady-state will be examined. The transient regimes of the vehicle, such as

acceleration and deceleration, will not be taken into account. For this reason, the section up to 24th data, which is the acceleration zone of the vehicle, will not be taken into consideration in the graphics. Likewise, the region where the vehicle slows down and accelerates due to the speed bump placed on the road will be excluded from the analysis.

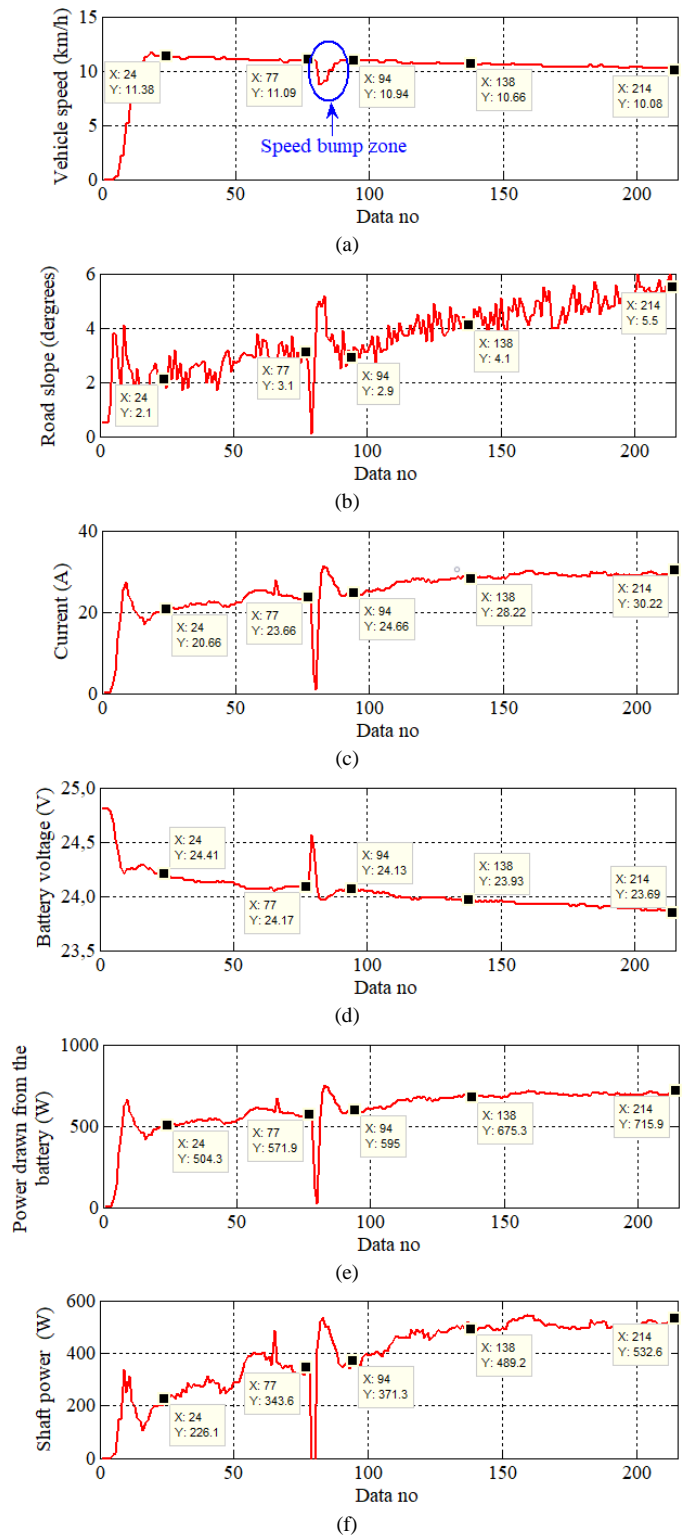


Fig. 13. Graphs related to the test on Route-2. (a) Vehicle speed, (b) Road slope, (c) Battery current, (d) Battery voltage, (e) Power drawn from the battery, (f) Shaft power (Time difference between two consecutive data 0.4711 s).

To facilitate understanding and interpretation of the data in the graphics, the points marked on the graphics and their related values have been collected in Table 2. The system efficiency (Eff_s), which is the ratio of the shaft power to the power drawn from the battery, has been also added to the last column of Table 2.

TABLE 2.
POINTS AND THEIR VALUES MARKED WITH THE HELP OF DATA CURSOR IN THE GRAPHS IN FIG. 13.

Data no	α (°)	I_b (A)	V_b (V)	v (km/h)	P_b (W)	P_s (W)	Eff_s (%)
27	2.1	20.66	24.41	11.38	504.3	226.1	44.83
77	3.1	23.66	24.17	11.09	571.9	343.6	60.08
94	2.9	24.66	24.13	10.94	595.0	371.3	62.40
138	4.1	28.22	23.93	10.66	675.3	489.2	72.44
214	5.5	30.22	23.69	10.08	715.9	532.6	74.40

As may remember, in the previous test, the battery current, which was 12.88 A at a road slope of 0.2°, increased to 20.33 A when the road slope was rose to 1.9°. In this test, too, the measurement of the battery current as 20.66 A at a road slope of 2.1 ° indicates that the system operates stably.

B. Experimental Studies on Negative Slope Roads

In order to research the effect of negative slope on the vehicle parameters, Route-3 given in Fig. 14 was chosen. In Route-3, which has a negative slope along approximately 900 m, the direction of the vehicle's movement is from point A to point B.

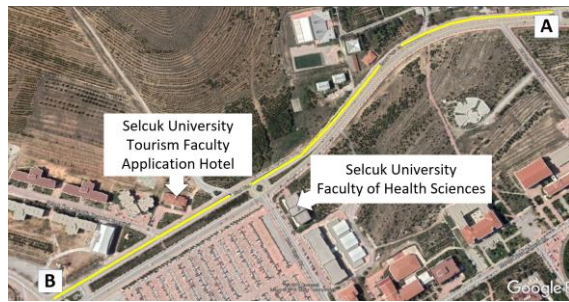


Fig. 14. Route-3

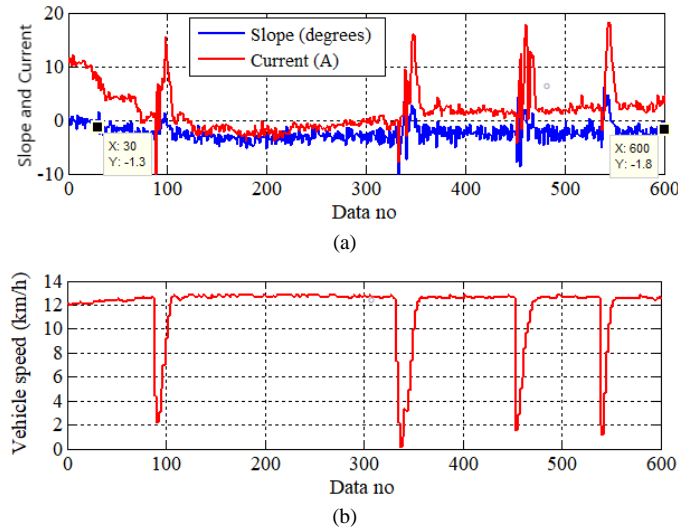


Fig. 15. Graphics of the test on the road with a negative slope (a) Slope and current graphic, (b) Vehicle speed graphic (Time difference between two consecutive data 0.4859 s).

In the test carried out in Route-3, changes in road slope, current drawn from the battery, and vehicle speed throughout the test are given in Fig. 15. The current and slope are given in the same graph (in Fig. 15.a) in order to see the change of current depending on the road slope more clearly.

When the slope graph is examined, it is seen that the slope starts to get negative value continuously starting from the 30th data. The negative slope continues regularly up to 600th data, except for the transient regime resulting from lowering and re-raising the vehicle speed due to speed bump on the road. The decrease and increase in vehicle speed which is the source of this transient regime will be seen when the speed graph in Fig. 15.b is examined.

Depending on the value of the slope, in order to see more clearly the current value that regenerative braking starts, the slope-current graphic has been enlarged between 120th-300th data and 350th-600th data and given in Fig. 16.

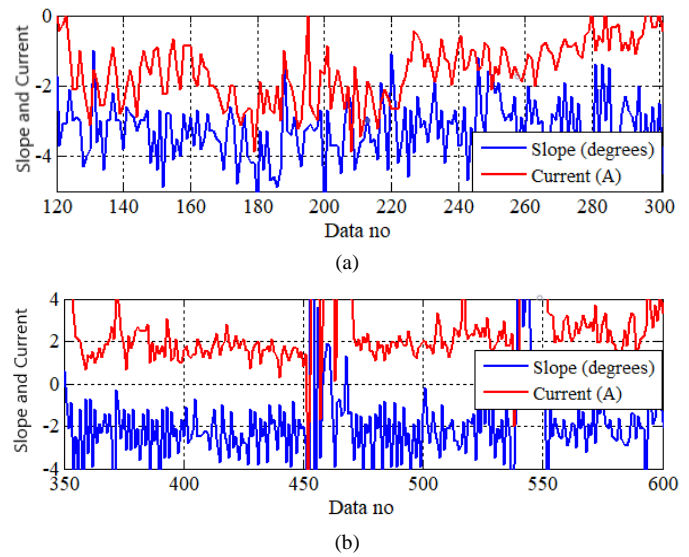


Fig. 16. Zooming of certain regions in the slope-current graphic of regenerative braking (a) 120th-300th data range, (b) 350th-600th data range (Time difference between two consecutive data 0.4859 s).

In Fig. 16.a, it can be seen that in regions where the average slope is less than -3°, the battery current has a significantly negative value, that is, the battery is charged. In Fig. 16.b, it is seen that the battery current has a positive value in the regions where the slope is -2° and larger, that is, the battery is discharged. However, in this region, the current drawn from the battery is quite small compared to the flat road and it varies in the range of 1.5 A and 2.5 A on average. This value is the current drawn to disable the vehicle's electromagnetic brake system.

The main component affected by the negative road slope is undoubtedly the battery. Therefore, examining the variation of the power drawn from the battery and the battery voltage throughout the test will enable us to see the effect of the negative slope from a different angle. The graphics related to this are given in Fig. 17.

When the graphic of power drawn from the battery in Fig. 17.a is examined, it is seen that the power flow is continuously from the motor to the battery in the 121th-299th data range. That is, the battery is continuously be charging due to the

negative road slope. After 299th data, it is noteworthy that, except for the transient regimes, the slope is always negative, but power is drawn from the battery.

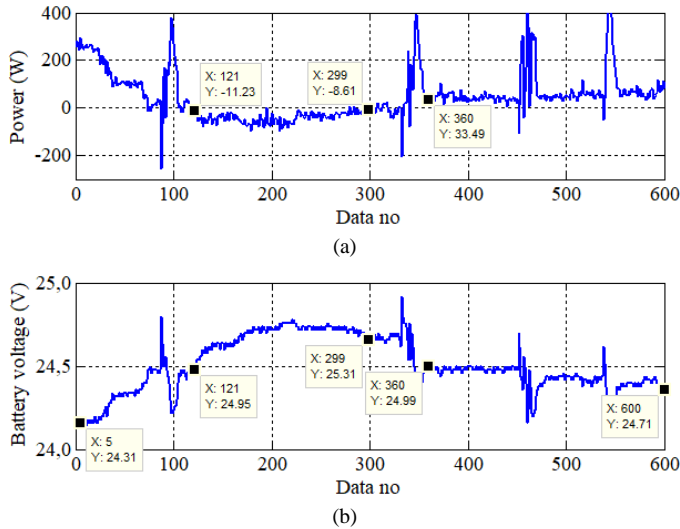


Fig. 17. Graphics of regenerative braking test (a) Power drawn from the battery, (b) Battery terminal voltage (Time difference between two consecutive data 0.4859 s).

When the battery voltage graphic in Fig. 17.b is examined, it is seen that the battery terminal voltage rises by 0.36 V (from 24.95 V to 25.31 V) among 121th-299th data where the battery is being charged continuously. When it is examined after the 360th data (the section after the speed bump), it is observed that the battery terminal voltage decreases gradually due to the gradual increase of the power drawn from the battery. When the entire test is taken into consideration, it is seen that the battery voltage, which was initially 24.31 V, increased to 24.71 V at the end of the test.

IV. DISCUSSION OF THE OBTAINED DATA

In the tests carried out, the following results were obtained when the road slope increased from 0.2° to 1.9° :

- The power drawn from the battery increased by 55.8%.
- Battery terminal voltage decreased by 1.29%.
- The current drawn from the battery increased by 57.8%.
- Vehicle speed decreased by 3.69%.

In order to see the effect of road slope on these parameters more clearly, it was made a test with the same user but on another route with more slope. When the slope of the road increased by 3.4° (from 2.1° to 5.5°), the following results were obtained:

- The current drawn from the battery increased by 46.27%. When the battery current in the 0.2° road slope is taken as reference, the 5.3° increase in the slope led to an increase of 134.63% in the current. Therefore, every 1° increase in road slope has caused an average increase of 25.40% in the battery current.
- Battery voltage decreased by 2.95%.
- Even when the vehicle reached a slope of 5.5° , PMDC motor used only 71.01% of its rated power (750 W).
- System efficiency increased from 44.83% to 74.40%.
- Vehicle speed decreased by 11.42%.
- Shaft power increased by 135.56%.

When the change in the shaft power and in the speed are evaluated together, it is seen that the increase in the road slope by 3.7° caused an increase in torque by 65.94%. Therefore, 1° increase in the road slope causes to produce 19.39% more torque of the motor. However, in the study conducted using basic formulas and given in [2], this value appears as 3.125%.

When the shaft power column in Table 2 is examined, it is seen that 532.6 W shaft power is sufficient even on an extreme road slope of 5.5° . If the disabled vehicle used in the experiment is used on roads where such extreme slopes will not exist, the motor shaft power it needs will remain below 300 W. Because even on a road slope of 2.1° , the motor shaft power of 226.1W was sufficient for the movement of the vehicle (Table 2). Therefore, using a motor with a shaft power of 750 W only increases the losses. The increase in losses means unnecessary energy withdrawal from the battery. This directly means that the battery charge is consumed unnecessarily and the battery is discharged in a shorter range.

In the tests performed, it was observed that the battery current had a clearly negative value if the average slope was less than -3° . In other words, regenerative braking occurred and the battery charged. If the road slope is -2° and bigger, it has observed that the battery current was positive, that is, the battery discharged. However, if the road slope is in the range of 0° to -2° , the current value drawn from the battery was observed to be small enough to disable the vehicle's electromagnetic brake system and its value ranged from 1.5 A to 2.5 A.

V. CONCLUSION

In this study, the effects of road slope on disabled vehicle speed, battery current, battery voltage, the power drawn from the battery, the motor shaft power, and the system efficiency were examined in the street tests of the disabled vehicle with PMDC motor used by 95 kg user. When the numerical data obtained from the experiments were evaluated, the following three main conclusions were reached.

- When determining the powers of the motors to be installed on the disabled vehicle, it should be taken into consideration whether the disabled vehicle will be used on inclined roads and unnecessarily high power motors should be avoided. Therefore, by choosing only a motor with suitable power, it will be possible for the disabled vehicle to travel longer with the same charge. In addition, because the frequency of charging the batteries will decrease, the life of the battery will increase.
- According to the formula-based study in [2], 1° increase in road slope requires to produce 3.15% more torque of the motor. However, in this experimental study, it was revealed that this value was 19.39%.
- It has been observed that regenerative braking in disabled vehicles does not contribute much to the charge of the battery, except that it reduces the vehicle speed. Because the battery does not start to charge unless the road slope falls below a certain value. In the test carried out on the road (Route-3) with a continuous negative slope for approximately 940m, it was observed that the battery was charged only when the road slope was less than -3° . On road slopes of -2° and greater, the battery continues to discharge.

ACKNOWLEDGMENT

This study is part of the project supported by Selçuk University Scientific Research Projects Coordination Office. The project number is 17101008.

REFERENCES

- [1] K. B. Kelly and H. J. Holcombe, "Aerodynamics for body engineers," SAE Transactions, pp. 570-578, 1964.
- [2] A. N. Gent, and J. D. Walter, "Pneumatic Tire", Mechanical Engineering Faculty Research. 854, the University of Akron, Ohio, USA, 2006.
- [3] A. Akpunar, "Ergonomik Bir Elektrikli Tekerlekli Sandalye İçin Dışsüz Ve Direkt Sürmeli Bir Elektrik Motor Tasarımı," Yüksek Lisans, Fen Bilimleri Enstitüsü, Muğla Üniversitesi, Muğla, 2007.
- [4] G. Marins, D. Carvalho, A. Marcato, and I. Junior, "Development of a control system for electric wheelchairs based on head movements," in Intelligent Systems Conference (IntelliSys), 2017, IEEE, pp. 996-1001.
- [5] S. Nasif and M. A. G. Khan, "Wireless head gesture controlled wheel chair for disable persons," in Humanitarian Technology Conference (R10-HTC), 2017 IEEE Region 10, 2017: IEEE, pp. 156-161.
- [6] S. Yokota, H. Hashimoto, Y. Ohyama, J.-H. She, D. Chugo, H. Kobayashi, and P. Blazevic, "Study on human body motion interface," in ICCAS-SICE, 2009, pp. 3264-3267.
- [7] D. Kupetz, S. Wentzell, and B. Busha, "Head motion controlled power wheelchair," in Proceedings of the 2010 IEEE 36th Annual Northeast Bioengineering Conference (NEBEC), 2010, pp. 1-2.
- [8] J. M. Ford and S. J. Sheredos, "Ultrasonic head controller for powered wheelchairs," Journal of Rehabilitation Research and Development, vol. 32, pp. 280-284, 1995.
- [9] F. A. Kondori, S. Yousefi, L. Liu, and H. Li, "Head operated electric wheelchair," in 2014 Southwest Symposium on Image Analysis and Interpretation, 2014: IEEE, pp. 53-56.
- [10] P. Ghule, M. Bhalerao, R. Chile, and V. G. Asutkar, "Wheelchair control using speech recognition," in 9th International Conference on Contemporary Computing (IC3), 2016: IEEE, pp. 1-6.
- [11] R. Chauhan, Y. Jain, H. Agarwal, and A. Patil, "Study of implementation of Voice Controlled Wheelchair," in 3rd International Advanced Computing and Communication Systems (ICACCS), 2016, pp. 1-4.
- [12] D. Wang and H. Yu, "Development of the control system of a voice-operated wheelchair with multi-posture characteristics," in 2nd Asia-Pacific Conference on Intelligent Robot Systems, 2017, pp. 151-155.
- [13] A. Škraba, R. Stojanović, A. Zupan, A. Koložvari, and D. Kofjač, "Speech-controlled cloud-based wheelchair platform for disabled persons," Microprocessors and Microsystems, vol. 39, pp. 819-828, 2015.
- [14] M. F. Ruzajij, S. Neubert, N. Stoll, and K. Thurow, "Design and testing of low cost three-modes of operation voice controller for wheelchairs and rehabilitation robotics," in IEEE 9th International Symposium on Intelligent Signal Processing (WISP) Proceedings, 2015, pp. 1-6.
- [15] M. F. Ruzajij and S. Poonguzhali, "Design and implementation of low cost intelligent wheelchair," in International Conference on Recent Trends in Information Technology, 2012, pp. 468-471.
- [16] M. A. Alim, S. Setumin, A. D. Rosli, and A. I. C. Ani, "Development of a Voice-controlled Intelligent Wheelchair System using Raspberry Pi," in 2021 IEEE 11th IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE), 2021, pp. 274-278.
- [17] M. S. Amin, S. T. H. Rizvi, S. Malik, Z. B. Faheem, and A. Liaqat, "Smart Wheelchair-An Implementation of Voice and Android Controlled System," in 2021 International Conference on Digital Futures and Transformative Technologies (ICoDT2), 2021, pp. 1-6.
- [18] P. S. Gajwani and S. A. Chhabria, "Eye motion tracking for wheelchair control," International Journal of Information Technology, vol. 2, pp. 185-187, 2010.
- [19] M. Jain, S. Puri, and S. Unishree, "Eyeball motion controlled wheelchair using IR sensors," International Journal of Computer and Information Engineering, vol. 9, pp. 906-909, 2015.
- [20] J. Sharma, M. Anbarasu, C. Chakraborty, and M. Shanmugasundaram, "Iris movement based wheel chair control using raspberry Pi—A state of art," in Power and Advanced Computing Technologies (i-PACT), 2017, pp. 1-5.
- [21] K. Arai and R. Mardiyanto, "Eyes based electric wheel chair control system," International Journal of Advanced Computer Science and Applications (IJACSA), vol. 2, 2011.
- [22] Q. X. Nguyen and S. Jo, "Electric wheelchair control using head pose free eye-gaze tracker," IET Electronics Letters, vol. 48, pp. 750-752, 2012.
- [23] M. Subramanian, N. Songur, D. Adjei, P. Orlov, and A. A. Faisal, "A. Eye Drive: Gaze-based semi-autonomous wheelchair interface," in 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2019, pp. 5967-5970.
- [24] M. Subramanian, S. Park, P. Orlov, A. Shafti, and A. A. Faisal, "Gaze-contingent decoding of human navigation intention on an autonomous wheelchair platform," in 2021 10th International IEEE/EMBS Conference on Neural Engineering (NER), 2021, pp. 335-338.
- [25] J. Ju, Y. Shin, and E. Kim, "Intelligent wheelchair using head tilt and mouth shape," IET Electronics letters, vol. 45, pp. 873-875, 2009.
- [26] E. Ko, J. S. Ju, E. Y. Kim, and N. S. Goo, "An intelligent wheelchair to enable mobility of severely disabled and elder people," in Digest of Technical Papers International Conference on Consumer Electronics, 2009, pp. 1-2.
- [27] J. S. Ju, Y. H. Shin, E. Y. Kim, and S. H. Park, "Intelligent wheelchair using face and mouth shape recognition," in Digest of Technical Papers-International Conference on Consumer Electronics, 2008, pp. 1-2.
- [28] V. Sankardoss and P. Geethanjali, "Design and low-cost implementation of an electric wheelchair control," IETE Journal of Research, pp. 1-10, 2019.
- [29] N. Tanohata, H. Murakami, and H. Seki, "Battery friendly driving control of electric power-assisted wheelchair based on fuzzy algorithm," in Proceedings of SICE Annual Conference, 2010, pp. 1595-1598.
- [30] T. Wang, J. i. Kaneko, and K. Kojima, "Study on relevance between electric wheelchair riding comfort and user exposure to whole-body vibration," in IEEE 6th Global Conference on Consumer Electronics, 2017, pp. 1-2.
- [31] H. M. Hondori, P. Q. Trung, and L. Shih-Fu, "Simultaneous sensing and actuating for path condition monitoring of a power wheel chair," in First RSI/ISM International Conference on Robotics and Mechatronics, 2013, pp. 343-346.
- [32] Y.-K. Kim, Y.-H. Cho, N.-C. Park, S.-H. Kim, and H.-S. Mok, "In-Wheel motor drive system using 2-phase PMSM," in 6th International Power Electronics and Motion Control Conference, 2009, pp. 1875-1879.
- [33] F. A. Kazan and R. Akkaya, "Use of PMSM in Electric Mobility Scooter Propulsion," in International Conference on Engineering Technologies (ICENTE'19), Konya, Turkey, 2019.
- [34] İ. Tarimer, A. Akpunar, and R. Gürbüzt, "Design of a direct sliding gearless electrical motor for an ergonomic electrical wheelchair," Elektronika IR Elektrotehnika, vol. 83, pp. 75-80, 2008.
- [35] B. Li, J. Zhang, H. Du, and W. Li, "Two-layer structure based adaptive estimation for vehicle mass and road slope under longitudinal motion," Measurement, vol. 95, pp. 439-455, 2017.
- [36] K. Jo, M. Lee, and M. Sunwoo, "Road slope aided vehicle position estimation system based on sensor fusion of GPS and automotive onboard sensors," IEEE Transactions on Intelligent Transportation Systems, vol. 17, pp. 250-263, 2015.
- [37] K. Jo, J. Kim, and M. Sunwoo, "Real-time road-slope estimation based on integration of onboard sensors with GPS using an IMM-PDA filter," IEEE Transactions on Intelligent Transportation Systems, vol. 14, pp. 1718-1732, 2013.
- [38] J.-H. Jang, S.-H. Bae, M.-W. Park, and J.-H. Kim, "Research of velocity control on the slope road for Unmanned Ground Vehicle," in International Conference on Control, Automation and Systems, 2010, pp. 1085-1088.
- [39] M. F. Ruzajij, S. Neubert, N. Stoll, and K. Thurow, "A speed compensation algorithm for a head tilts controller used for wheelchairs and rehabilitation applications," in IEEE 15th International Symposium on Applied Machine Intelligence and Informatics (SAMII), 2017, pp. 497-502.
- [40] S. Yang, M. Li, Y. Lin, and T. Tang, "Electric vehicle's electricity consumption on a road with different slope," Physica A: Statistical Mechanics and its Applications, vol. 402, pp. 41-48, 2014.
- [41] E. Ribar and J. Murin, "Road slope introduction in vehicle route modelling," in Cybernetics & Informatics (K&I), 2016, pp. 1-5.
- [42] F. A. Kazan and R. Akkaya, "The Effect of Increases in User Weight and Road Slope on Energy Consumption in Disabled Vehicle Driven

- with PMSM," *Balkan Journal of Electrical and Computer Engineering*, vol. 9, pp. 1-7, 2021.
- [43] I. Aden, H. Kahveci, and M. E. Şahin, "Single input, multiple output dc-dc buck converter for electric vehicles," *Turkish Journal of Electromechanics & Energy*, vol. 2, no. 2, pp. 7-13, 2017.
- [44] I. A. Aden, H. Kahveci, and M. E. Şahin, "Design and Implementation of Single-Input Multiple-Output DC-DC Buck converter for Electric Vehicles," *Journal of Circuits, Systems and Computers*, pp. 215-228, 2021.
- [45] S. Heydari, P. Fajri, M. Rasheduzzaman, and R. Sabzehgar, "Maximizing regenerative braking energy recovery of electric vehicles through dynamic low-speed cutoff point detection," *IEEE Transactions on Transportation Electrification*, vol. 5, no. 1, pp. 262-270, 2019.
- [46] R. Akkaya and F. A. Kazan, "Design and implementation of a test setup for electric mobility scooter for the disabled," *Measurement and Control*, vol. 52, no. 9-10, pp. 1434-1444, 2019.
- [47] ACS712 Datasheet, Allegro MicroSystems, 2017.

BIOGRAPHIES



FATİH ALPASLAN KAZAN was born in Amasya, Turkey. In 2002, he graduated from Electrical Education Department of Gazi University and started to work as a lecturer at Selçuk University. He completed his master's degree in Selçuk University, Institute of Science, Electrical and Electronics Engineering in 2009. He received his PhD degree from the Department of Electrical and Electronics Engineering of Konya Technical University in 2019. He has been working as Asst. Prof. in the Department of Aviation Electrical and Electronic of Selçuk University since 2020. His main interests are Avionic Systems, Microcontroller Applications, Measurement Systems, Circuit Design, and Motor Control.



RAMAZAN AKKAYA was born in Aksaray, Turkey, in 1963. He received the B.S., M.S., and PhD degrees from Electrical Engineering of Yıldız Technical University, Istanbul, Turkey in 1986, 1988, and 1995, respectively. He worked as a Research Assistant between 1987 and 1995 and as an Assistant Professor between 1996 and 1998 at Yıldız Technical University Electrical Engineering Department. In 1998, he started to work at Selçuk University Electrical and Electronics Engineering Department. He became Associate Professor in 2010 and Professor in 2016 at the same university. He has been working as a Prof. Dr. in the Department of Electrical and Electronics Engineering at Konya Technical University since 2018. His main interests are Power Electronics and Applications, Control of Electric Drives, Electric Power Quality, and Renewable Energy Systems.

Longitudinal Inter-vehicle Distance Control of Autonomous Vehicle Platoons Subjected to Internal and External Disturbances

Alex Gunagwera and Aydin Tarik Zengin*

Abstract—The idea of autonomous vehicle platoons presents a variety of social, economic, and safety benefits to the transportation industry. However, implementing and deploying autonomous vehicle platoons is still a challenge. In this paper, we present a PID-based computationally cost-efficient controller to aid in the longitudinal control of the inter-vehicle distance between successive platoon members. The proposed approach is facilitated by inter-vehicle communication. The algorithm was implemented using the Robotics Operating System and Gazebo simulation environment. In order to evaluate the performance and applicability of the proposed approach, meticulous simulations under numerous scenarios were performed using 3D vehicle models so as to mimic the real-world. The algorithm successfully maintains the longitudinal inter-vehicle distance within the desired range, ensures that no collisions occur among platoon members, and preserves the platoon formation.

Index Terms—Autonomous vehicles, autonomous vehicle platoons, intelligent transport systems, longitudinal platoon control, automated highway systems, PID.

I. INTRODUCTION

AUTONOMOUS VEHICLE PLATOONS (AVPs) is a trending multidisciplinary topic nowadays attracting attention from researchers, practitioners, and governmental bodies all over the world. AVPs are comprised of two or more vehicles mechanically or electronically connected and travelling closely together as a single unit with the same lateral and, or longitudinal motion control. Autonomous vehicle platooning promises a variety of social, economic and safety benefits ranging from saving employee time [1], optimal energy consumption [2], efficient road utilization [3]


to minimizing traffic accidents culminating from human error [4].

Owing to the multidisciplinary nature of AVPs, a multitude of work has been done pertaining to AVPs; their control, analysis, deployment to mention but a few. Work from areas such as control and analysis, the communication industry, and energy department have contributed significant work to the enhancement of AVP applications. For instance, studies on Adaptive Cruise Control (ACC) [5], Cooperative Adaptive Cruise Control (CACC) [6], [7], String Stability [8], [9]. The cruise control model proposed by [5] was implemented in MATLAB SIMULINK. They used the velocity and inter-vehicle distance deviation as inputs to their controller.


Communication plays a vital role in the overall success of autonomous vehicle platoons. In fact, AVPs incorporating communication in their architecture register better results in comparison to those that operate without communication among platoon members [10]. The quality, ease of communication and type of information shared significantly affects the performance of an AVP. Furthermore, the efficiency of the communication methodology employed, and the amount of information communicated in the platoon also have the ability to enhance the success of the entire platoon. Steven E. Shladover *et al.* [6] provided the essential definitions and distinctions among the different types of CACC and the various communication types employed by platoons. Robust Vehicle to Vehicle (V2V) communication, such as VANET, DSRC [11], [12], Vehicle to Infrastructure (V2I) or even both (V2X) communications facilitate the functioning of AV platooning. Shen, Z. *et al.* [13] discussed the effects of communication reliability and latency on the performance of vehicle systems using 5G V2X hardware prototypes and the 802.11 communication protocol. In our study, we utilize the Robot Operating System (ROS) [14] messages and topics for communication within the platoon. Inter vehicle communication is wireless and is based on Wi-Fi (IEEE 802.11) network.

String stability is another important feature of an autonomous vehicle platoon. Cremer, D. [9] provided a string stability criterion which only depends on the error in velocity of each vehicle in comparison to the velocity of the Leading Vehicle (LV). Cremer's standards did not rely on the inter-vehicle distance. Seiler *et al.* defined platoon stability as the error between the desired and the actual inter-vehicle spacing [15]. Swaroop *et al.* [16] presented string stability requirements that depend on the inter vehicle distance in two

ALEX GUNAGWERA, is with Department of Computer Engineering Istanbul Sabahattin Zaim University, Istanbul, Turkey, (e-mail: alex.gunagwera@izu.edu.tr).

 <https://orcid.org/0000-0002-0143-3743>

AYDIN TARIK ZENGİN, is with Department of Computer Engineering Istanbul Sabahattin Zaim University, Istanbul, Turkey, (e-mail: tarik.zengin@izu.edu.tr).

 <https://orcid.org/0000-0002-0860-4509>

Manuscript received Nov 1, 2021; accepted Jan 28, 2022.
DOI: [10.17694/bajece.1017623](https://doi.org/10.17694/bajece.1017623)

categories: the strong sense and the weak sense. In the strong sense, the presented string stability conditions requires that the maximum inter vehicle distance error of the i^{th} vehicle should either be equal or less to that of the $i - 1^{th}$ vehicle. String stability in a weak sense has a requirement that just the maximum inter-vehicle distance errors should be less than or equal to those of the first follower (F) vehicle.

The Global Positioning System (GPS) [17], is one of the most valuable sensors in look-ahead systems. For example, [18] illustrated a high integrity navigation system's development and implementation for usage in autonomous land vehicle applications. They mainly used GPS and the Inertial Measurement Unit (IMU) in their work. GPS is the backbone of the approach we present in this study as well.

Inter-vehicle distance (IVD) is an essential metric in autonomous vehicle platoons. How well it is kept and maintained ensures safe, comfortable, and more efficient road usage, among others. It is with this background that we propose this study.

In this paper, we propose a computationally cost efficient, PID-based algorithm for controlling the inter vehicle distance between successive members of an autonomous vehicle platoon. Our approach differs from other numerous studies mainly by using only the onboard GPS sensors of the vehicles, gazebo robot simulator and the Robot operating system (ROS). Furthermore, our PID controller requires just the current longitudinal inter vehicle distance to the preceding vehicle, unlike most PID approaches that require the velocity and acceleration information as well. The algorithm observes maintenance of platoon formation and makes sure no collisions occur amongst platoon members. Obtained results are presented and 3D simulations of the system further carried out using ROS and Gazebo platforms to demonstrate the performance of the proposed approach.

This paper is organized as follows. Section II presents the problem statement and followed platoon model. Section III explains the scenarios considered during simulations. Section IV presents the environment and conditions under which the simulations were performed. Section V illustrates the results obtained from the simulations, whereas Section VI discusses the obtained results, the limitations of the proposed approach, overall practicality, and applicability of the proposed algorithm. Furthermore, how the proposed algorithm differs from the current related works. Finally, in Section VII, we conclude this work and present directions for the future work.

II. PROBLEM STATEMENT AND PLATOON MODEL

The controlled platoon comprises of four vehicles in total. The platoon Leader Vehicle (LV) and three Follower (F) vehicles. We design a PID controller to aid the control of the distance between vehicles. It takes as input the current inter vehicle distance between vehicles and returns as output a velocity reference for the corresponding F_i vehicle in order to achieve the desired inter-vehicle distance, (D) to the preceding vehicle. We thus state the problem as:

$$Set(d_i) = D \pm |E_i| \tag{1}$$

$$E_i = D - d_i \tag{2}$$

$$\text{and } |E_i| \leq E_{thresh}.$$

Ultimately, the major purpose of our PID controller is to reduce the error, E_i , and drive it as close to $0m$ as possible. So, the best-case scenario at any point in the simulation is to have $E_i = 0m$, especially during the steady state. $\forall F_i, i \in \{1, 2, 3\}$ where d_i is the i^{th} inter-vehicle distance, D is the desired inter-vehicle distance, and E_i is the error between the i^{th} inter-vehicle distance and the desired distance, D . E_{thresh} is the maximum and minimum threshold value beyond which the error should not exceed in order to guarantee safety. This constraint ensures that F vehicles are not allowed to fall more than E_{thresh} behind the preceding vehicle, i.e., $E_i \leq E_{thresh}$. It also ensures that F vehicles do not get more than E_{thresh} closer to the preceding vehicle, i.e., $E_i \geq -E_{thresh}$. Every F vehicle runs its own instance of the PID control algorithm. The platoon model presented in this work is based on the Predecessor-Follower communication model [19]. In the simulations, the desired inter-vehicle distance, D , is set to $12m$ and E_{thresh} is considered as $5m$.

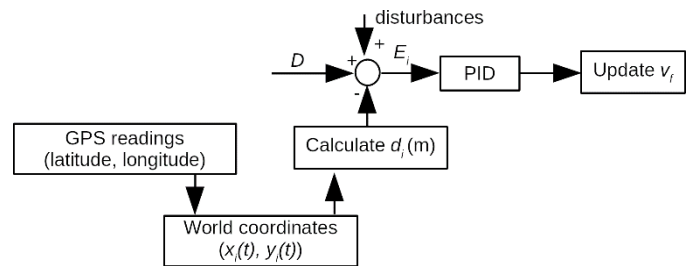


Fig.1. Controller Model

Fig.2. demonstrates the platoon setup. LV , is the Leading vehicle, also referred to as the root node of the platoon and is generally indexed as the first member of the platoon. The F labels depict the follower vehicles. The distance from one vehicle's center of mass to the preceding vehicle's center of mass is referred to as the inter-vehicle gap/distance in this study.

In this study, the control algorithm aims at ascertaining a constant inter-vehicle gap with all vehicles' velocity sufficiently approximately equal to that of the LV in the platoon using only the distance measure from the data provided by the onboard GPS sensors of the vehicles. Errors in the inter-vehicle gap should be bounded, and there should be always no collisions among platoon members in the worst-case scenario during the simulation. Following the definition of the platoon stability provided by [15], we can formulate the steady-state error transfer function as

$$H(s) = \frac{E_i}{E_{i-1}}. \tag{3}$$

Followingly, platoon stability is guaranteed, locally, if $\|H(s)\|_\infty \leq 1$, and $h(t) > 0$ where $h(t)$ gives the impulse response corresponding to $H(s)$ as per the ζ_2 norm [10]. ζ_∞ extends this notion throughout the whole platoon to ensure that overshoots do not occur as the signals propagate up the string, hence global stability.

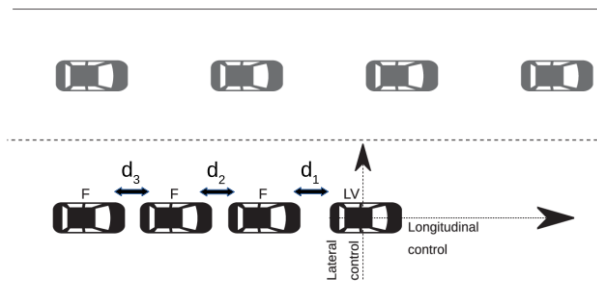


Fig.2. Illustration of inter-vehicle distance, d_i , LV, and F

At the start of the simulation, GPS data measurements are retrieved asynchronously if available from every vehicle's onboard GPS sensor. From this data, the relative inter-vehicle gap is calculated and forwarded to the PID algorithm which in turn returns the reference velocity with which the corresponding F vehicle's velocity is updated to achieve the desired inter-vehicle gap between F and the preceding vehicle. The desired inter-vehicle distance, D is set as the PID algorithm's *setpoint* whereas the inter-vehicle gap estimates, calculated from the data measurements provided by the GPS sensor, is provided as the *feedback* for the PID control algorithm. We repeat these steps throughout the simulation lifetime.

Fig.1. shows the general algorithm flow. Each F_i vehicle runs this algorithm with v_f being its effective reference velocity.

III. CONSIDERED SCENARIOS DURING SIMULATIONS

In reality, systems are exposed to various conditions and sensors do not perform as smoothly as intended. Due to natural and mechanical phenomena, sensor readings are generally, corrupted and affected by external noise. Furthermore, there is exhibited by the vehicle onboard sensors (sensor lag), lag within the vehicle control system and the delay in V2V communication. These are all inevitabilities that should be put into consideration if we wish to make the simulations as realistic, and accurate as possible. In case, the LV is either being manned by a human or is following a human-manned vehicle in traffic, the human-factor may, to some degree, not be negligible, i.e., the driver may suddenly brake or accelerate for any uncertain period of time. We shall, hereafter, refer to this phenomenon as the *human factor(HF)* - which may or may not be present in a particular scenario.

In this section we present the performance of the proposed controller with the above-mentioned criteria under consideration. The simulated scenarios are divided into four categories: In the first scenario, the platoon is *only* affected by sensor lag. *HF* and other forms of delays discussed are not present. In the second scenario, the platoon is subjected to both the sensor lag and random occurrences of the *HF*, only. In the third scenario, the platoon is subjected to the sensor lag, random V2V communication, and vehicle control delays with no *HF* occurrences. In the final scenario, the platoon is subjected to the random *HF*, sensor lag, and random V2V communication, and control delays. For all simulation scenarios presented in this section, the *LV* action at any given moment can be one of acceleration, deceleration or moving at

constant speed. The action is randomly generated with the following limitations and constraints:

- 1) Reverse vehicle motion is not permitted within the platoon. In case the LV velocity, during the deceleration phase period, were to drop below 0m/s , the LV is programmed to stop, i.e., LV velocity is set to 0m/s .
- 2) No action; acceleration, deceleration or constant speed should be executed more than once consecutively. This way, we ensure that the platoon performance under various uncertain scenarios is observed.
- 3) The duration of each phase/action is randomly generated and may last between 20 - 50 seconds with the exception of the final phase - the phase during which the LV decelerates to rest.

The first and final phases are a bit different. Since the platoon starts motion from rest, the first phase/action has to always be acceleration; the *LV* accelerates for a random period of time. The final phase is always the deceleration of the *LV* to rest, therefore its duration is not determined randomly. The vehicle simply uniformly decelerates to 0m/s .

Table 1 provides a categorical summary of the scenarios considered in this simulation. Follower vehicle controllers are data from sensors. The sensors are subjected to reductions in update frequency which leads to delays in the processed data. This helps to portray unfavorable real-world conditions [21].

V2V communication frequency was set in such a way that vehicles publish their information at an average of 33Hz , which is the proposed maximum frequency by [20]. We vary this frequency with a minimum being 20Hz and the maximum being 50Hz to account for variations in platoon member mechanical differences as well as miscellaneous occurrences such as natural phenomena such as variations in wind speeds. The time between messages varies with an average of 0.03s with a minimum of 0.02s and a maximum of 0.05s .

TABLE I
SUMMARY OF THE SCENARIOS OF THE SCENARIOS CONSIDERED DURING SIMULATIONS

Scenarios 1 and 2	Scenarios 3 and 4
sensor lag w/o <i>HF</i> sensor lag w/ <i>HF</i>	sensor lag, V2V communication delays, w/o <i>HF</i> sensor lag, V2V communication delays, w/ <i>HF</i>

IV. SIMULATION ENVIRONMENT

3D vehicles were designed and modeled using Gazebo robotics simulator version 9 integrated with ROS 1 (Melodic). Visuals of the 3D vehicle objects were designed using the gazebo platform, whereas vehicle motion was handled and controlled via topics and messages by the nodes implemented using the ROS framework [14]. The vehicle model is based on the Hyundai Genesis-2014 [20] which reproduces the major dynamic characteristics such as friction, acceleration, deceleration/braking, wheel radius, and weight making it possible to critically analyze the movement of vehicles in the platoon.

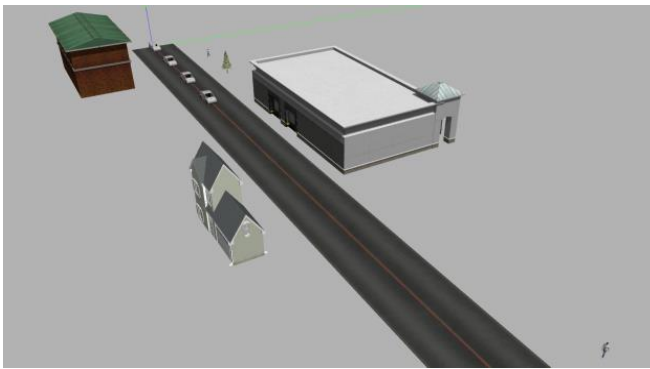


Fig.3. Gazebo simulation environment

Table 2 presents some of the major vehicle model parameters and their values.

TABLE II
VEHICLE INFORMATION

Vehicle Attribute	Value
Vehicle Mass	1823.0kg
Vehicle Length	5m
Vehicle Width	1.89m
Vehicle Height	1.480m
Wheel Radius	0.34m
Wheelbase	2.95m
Wheel Width	0.225m
Drag coefficient	0.27 cd

Fig.3. shows the environment in which the simulations were performed and monitored. The vehicles move forward along the road during simulation to preserve the presented platoon formation and the desired inter-vehicle distance. We performed simulations under the assumptions and constraints that:

- 1) The roads are straight and have got no slope so that longitudinal control of the platoon remains the focus of the platoon
- 2) Communication is wireless, that is, over a Wi-Fi (IEEE 802.11) and each vehicle is only allowed to communicate to the preceding vehicle.
- 3) Overtake and reversing manoeuvres are not allowed in the platoon.

V. RESULTS

In this section we present the results obtained from the performed simulations. For each simulation scenario, two study cases are presented. We manually tuned the PID controller. The parameter gains of the controller that yielded the presented results are listed in Table 3. We conclude from the observed results that the proposed approach indeed does guarantee the following effect while keeping the inter-vehicle distance within the desired range. The algorithm also ensures platoon formation preservation and no collisions amongst platoon members since the error, $E_i \geq -E_{thresh}$ at all times throughout the simulation. However, much as the vehicles in the platoon obtain vehicle stability [23], overall platoon stability is not guaranteed. The proposed controller also shows

best performance at steady state when the *LV* travels with uniform velocity – as can be most explicitly observed in the first scenario of the simulations. All error values presented include the transient errors.

TABLE III
PID CONTROLLER PARAMETERS

Parameter	Value
<i>P</i>	0.05
<i>I</i>	0.00001
<i>D</i>	0.03

Long et al. [24] proposed a distributed model predictive model for the longitudinal control of truck platoons. Their model considers the state of the *LV*. In their study, platoon members transition from cruise control (CC), adaptive cruise control (ACC), and cooperative cruise control (CACC). ACC and CACC constitute the most advanced platoon controllers that exist today. They use MATLAB to evaluate the performance of their model. We present a summarized comparison of the performance of our approach and the ACC phase of their study. Throughout the simulations, Long et al. had the *LV* accelerate to $\approx 16\text{m/s}$, then move with constant velocity after that. They experienced a maximum transient response error of $\approx 37\text{m}$ before the system finally settled and had the range error converge to 0m . Our proposed controller had a maximum transient error of 4.9m , after which it converged to or close to 0m at steady state especially whenever the *LV* was moving with constant speed.

A. Scenario 1

In this scenario, platoon members were only subjected to sensor lag. This lag is assumed to occur in the GPS sensors in this study. This scenario registered the best performance. In this scenario, global platoon stability can seldomly be observed at steady state especially when the *LV* travels with constant speed.

In the first case study of scenario 1 (Fig.4.), there is the *LV* velocity throughout the simulation illustrated in the first subfigure. Platoon real-time IVD (d_i) is presented in the second subfigure. The third subfigure presents the error (E_i) in the IVD in comparison to the desired IVD (D). The fourth subfigure presents platoon members' real-time velocity profiles during the simulation. In the first and second subfigures, blue denotes the IVD (d_1) and error (E_1) in IVD between the *LV* and F_1 , respectively. Similarly, red and green, denote the IVD (d_2 , d_3) and their corresponding errors (E_2 , E_3), respectively. In the fourth subfigure, v_1 , depicted by blue, shows the *LV* velocity while v_2 (red), v_3 (green) and v_4 (cyan) represent the velocities of the F vehicles (1, 2, and 3), respectively.

During this case study, a minimum error of -2.8m and a maximum error of 3.0m in IVD were obtained with maximum standard deviation and variance of 1.94m and 3.7m^2 , respectively. During this scenario, local platoon stability is not guaranteed since, for instance, $E_3/E_2 > 1$.

Similarly, in the second case study of the first scenario (Fig. 5.), there is the *LV* velocity throughout the simulation. Platoon real-time IVD (d_i) is presented in the second subfigure. The

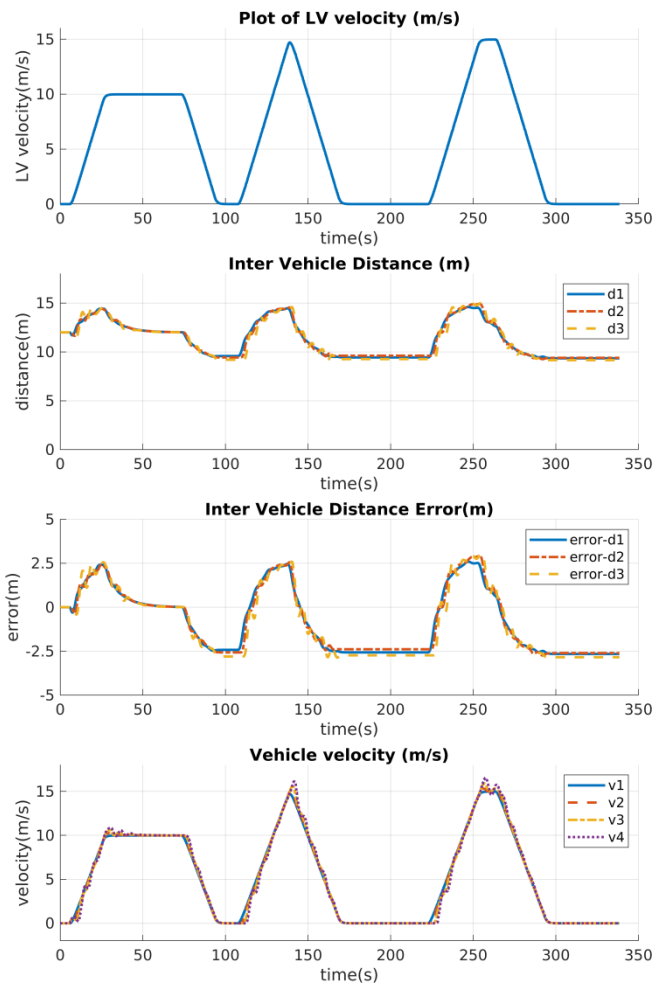


Fig.4. Scenario 1, first case study

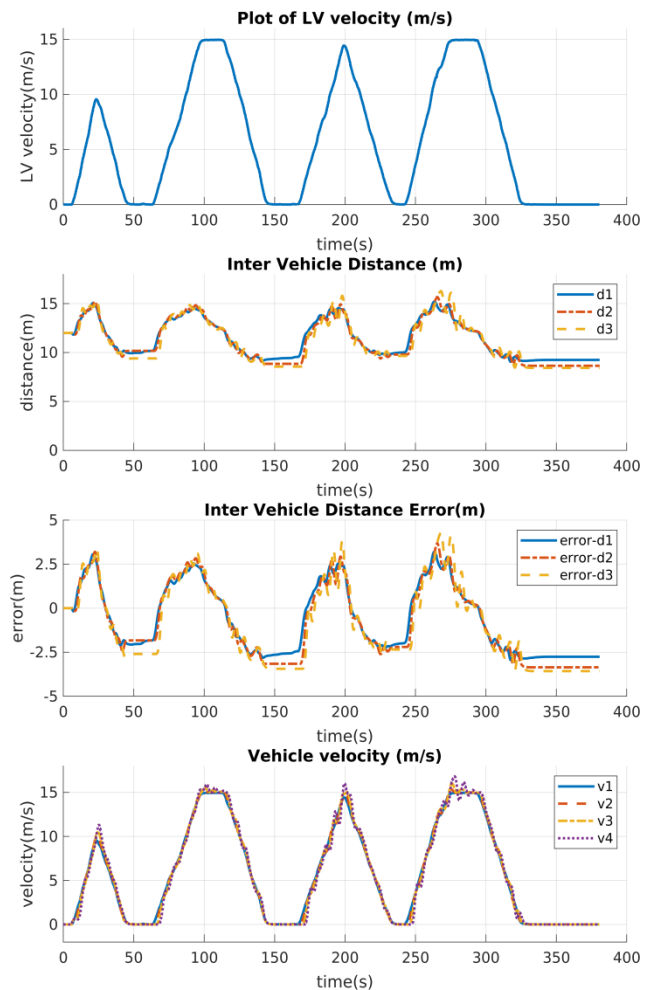


Fig.5. Scenario 1, second case study

third subfigure presents the error (E_i) in the IVD in comparison to the desired IVD (D). The fourth subfigure presents platoon members' real-time velocity profiles during the simulation. In the first and second subfigures, blue denotes the IVD (d_1) and error (E_1) in IVD between the LV and F_1 , respectively. Similarly, red and green, denote the IVD (d_2 , d_3) and their corresponding errors (E_2 , E_3), respectively. In the fourth subfigure, v_1 , depicted by blue, shows the LV velocity while v_2 (red), v_3 (green) and v_4 (cyan) represent the velocities of the F vehicles (1-3), respectively. In this particular case study, a minimum error of $-3.5m$ and a maximum error of $4.2m$ in IVD were obtained with maximum standard deviation and variance of $1.9m$ and $4.9m^2$, respectively. During this scenario, local platoon stability is not guaranteed since, for instance, $E_2/E_1 > 1$.

B. Scenario 2

In this scenario, the platoon members were subjected to sensor lag and the LV was subjected to random HF. In the first subfigure of the Scenario 2's first case study (Fig. 6.), we have the LV velocity profile during the simulation. Platoon real-time IVD(d_i) in the second subfigure followed by the error(E_i) in the IVD in comparison to the desired IVD (D),

in the third subfigure. Finally, the fourth subfigure shows the entire platoon's real-time velocity profiles during the simulation. Just like in, Fig.4. and Fig.5., in the first and second subfigures, blue denotes the IVD(d_1) and error (E_1) in IVD between the LV and F_1 , respectively. Similarly, red, and green, depict the IVD (d_2 , d_3) and their corresponding errors (E_2 , E_3), respectively. In the fourth subfigure, v_1 , depicted by blue, shows the LV velocity while v_2 (red), v_3 (green) and v_4 (cyan) represent the velocities of the F vehicles (1-3), respectively. In this case study, a minimum error of $-3.8m$ and a maximum error of $3.1m$ in IVD were obtained with maximum standard deviation and variance of $1.9m$ and $3.7m^2$, respectively. During this scenario, local platoon stability is not guaranteed since, for instance, $E_3/E_2 > 1$.

In the first subfigure of Scenario 2's second study (Fig.7.), the LV velocity throughout the simulation is presented followed by platoon real-time IVD(d_i) in the second subfigure. The third subfigure illustrates the error(E_i) in the IVD in comparison to the desired IVD (D). The fourth subfigure, on the other hand, real-time velocity profiles of the platoon during the simulation. In the first and second subfigures, blue denotes the IVD(d_1) and error (E_1) in IVD between the LV and F_1 , respectively. Similarly, red and green, denote the IVD (d_2 , d_3) and their corresponding errors (E_2 , E_3), respectively. In the

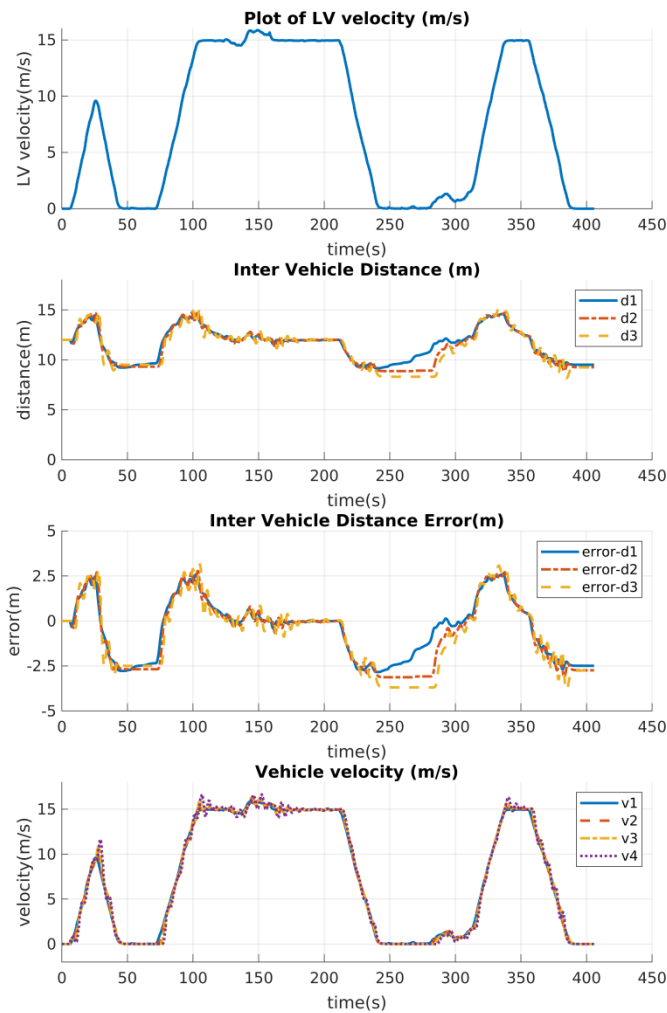


Fig.6. Scenario 2, first case study

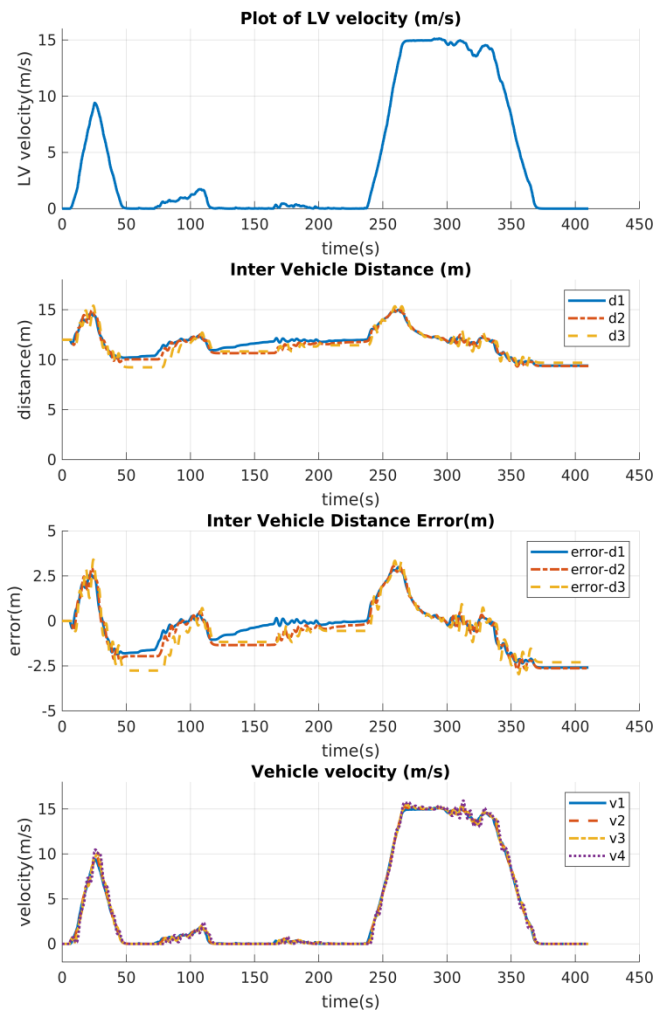


Fig.7. Scenario 2, second case study

fourth subfigure, $v1$, depicted by blue, shows the LV velocity while $v2$ (red), $v3$ (green) and $v4$ (cyan) represent the velocities of the F vehicles (1-3), respectively. A minimum error of $-2.9m$ and a maximum error of $3.4m$ in IVD were obtained with maximum standard deviation and variance of $1.4m$ and $2.0m^2$, respectively in this study case. During this scenario, local platoon stability is not guaranteed since, for instance, $E_3/E_2 > 1$.

C. Scenario 3

In this scenario, platoon members were subjected to random V2V communication delays, control delays within the vehicle, and sensor lags without the HF effect.

In the first case study of Scenario 3 (Fig.8.), there is the LV velocity throughout the simulation. Platoon real-time IVD(d_i) is presented in the second subfigure. The third subfigure presents the error(E_i) in the IVD in comparison to the desired IVD (D). The fourth subfigure presents platoon members' real-time velocity profiles during the simulation. In the first and second subfigures, blue denotes the IVD(d_1) and error (E_1) in IVD between the LV and F_1 , respectively. Similarly, red and green, denote the IVD (d_2 , d_3) and their corresponding errors (E_2 , E_3), respectively. In the fourth subfigure, $v1$, depicted by

blue, shows the LV velocity while $v2$ (red), $v3$ (green) and $v4$ (cyan) represent the velocities of the F vehicles (1-3), respectively.

During this case study, a minimum error of $-4.4m$ and a maximum error of $4.9m$ in IVD were obtained with maximum standard deviation and variance of $1.8m$ and $3.4m^2$, respectively. During this scenario, local platoon stability is not guaranteed since, for instance, $E_2/E_1 > 1$ almost everywhere.

In the second case study of this scenario (Fig.9.), we have the LV velocity throughout the simulation. Platoon real-time IVD(d_i) is presented in the second subfigure. The third subfigure presents the error (E_i) in the IVD in comparison to the desired IVD (D). The fourth subfigure presents platoon members' real-time velocity profiles during the simulation. In the first and second subfigures, blue denotes the IVD(d_1) and error (E_1) in IVD between the LV and F_1 , respectively. Similarly, red and green, denote the IVD (d_2 , d_3) and their corresponding errors (E_2 , E_3), respectively. In the fourth subfigure, $v1$, depicted by blue, shows the LV velocity while $v2$ (red), $v3$ (green) and $v4$ (cyan) represent the velocities of the F vehicles (1-3), respectively. During the second case study of scenario 3, presented in Fig.9., a minimum error of $-4.1m$ and a maximum error of $4.2m$ in IVD were obtained with maximum standard deviation and variance of $1.9m$ and $3.7m^2$,

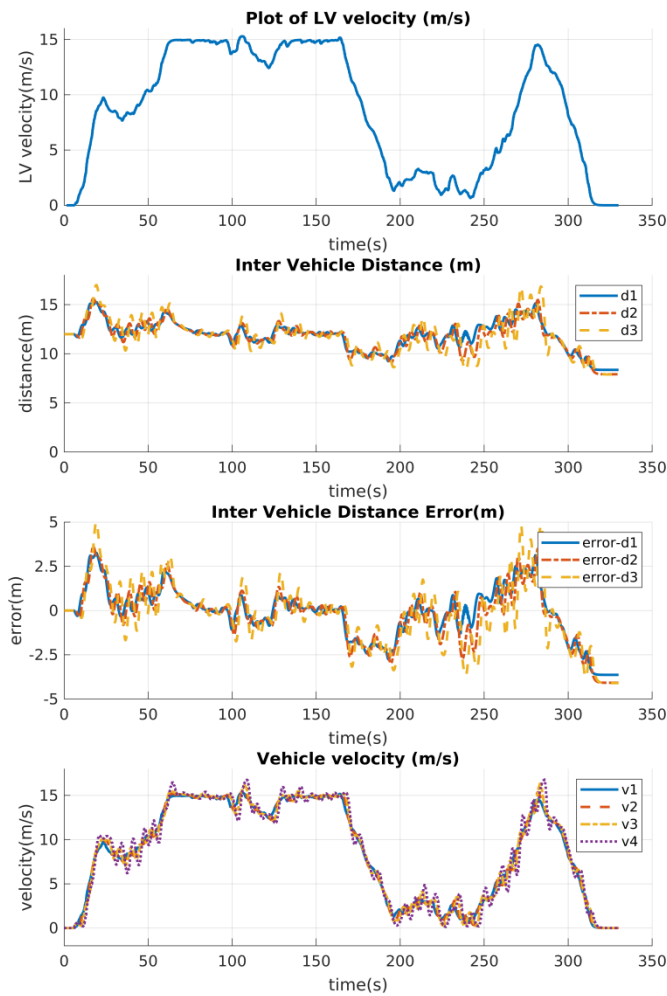


Fig.8. Scenario 3, first case study

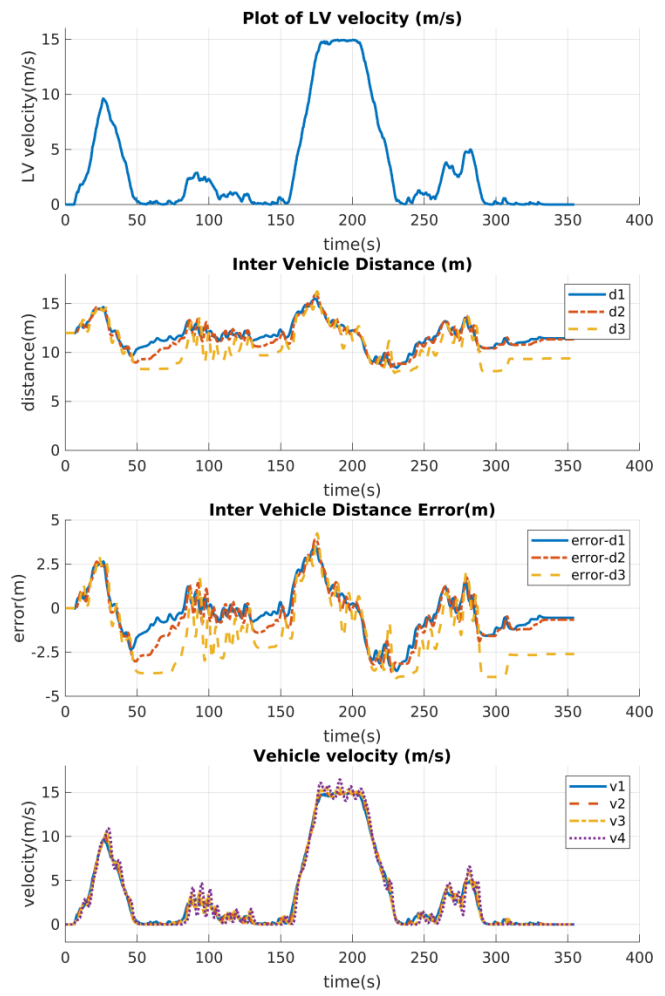


Fig.9. Scenario 3, second case study

respectively. During this scenario, local platoon stability is not guaranteed since, for instance, $E_3/E_2 > 1$.

D. Scenario 4

In this scenario, the platoon members were subjected to random V2V communication and vehicle control delays, sensor lag and random HF effect. In this scenario, platoon stability is not guaranteed, as well, either globally or locally since $H(s) > 1$ for all follower vehicles almost everywhere. In the first subfigure of Fig.10. we present the LV velocity in the simulation. Platoon real-time IVD(d_i) is presented in the second subfigure. The third subfigure presents the error(E_i) in the IVD in comparison to the desired IVD (D). The fourth subfigure presents platoon members' real-time velocity profiles during the simulation. In the first and second subfigures, blue denotes the IVD(d_1) and error (E_1) in IVD between the LV and F_1 , respectively. Similarly, red, and green, denote the IVD (d_2, d_3) and their corresponding errors (E_2, E_3), respectively. In the fourth subfigure, v_1 , depicted by blue, shows the LV velocity while v_2 (red), v_3 (green) and v_4 (cyan) represent the velocities of the F vehicles (1-3), respectively.

During the case study presented in Fig.10. a minimum error of $-3.9m$ and a maximum error of $3.9m$ in IVD were obtained

with maximum standard deviation and variance of $1.6m$ and $2.6m^2$, respectively. During this scenario, local platoon stability is not guaranteed since, for instance, $E_3/E_2 > 1$.

Likewise, the first subfigure of Fig.11. demonstrates the LV velocity throughout the simulation. Platoon real-time IVD(d_i) is presented in the second subfigure. The third subfigure presents the error(E_i) in the IVD in comparison to the desired IVD (D). The fourth subfigure presents platoon members' real-time velocity profiles during the simulation. In the first and second subfigures, blue denotes the IVD(d_1) and error (E_1) in IVD between the LV and F_1 , respectively. Similarly, red and green, denote the IVD (d_2, d_3) and their corresponding errors (E_2, E_3), respectively. In the fourth subfigure, v_1 , depicted by blue, shows the LV velocity while v_2 (red), v_3 (green) and v_4 (cyan) represent the velocities of the F vehicles (1-3), respectively. In the second case study of the fourth scenario (Fig.11.), a minimum error of $-4.5m$ and a maximum error of $4.1m$ in IVD were obtained with maximum standard deviation and variance of $2.0m$ and $3.9m^2$, respectively. During this scenario, local platoon stability is not guaranteed since, for instance, $E_3/E_2 > 1$.

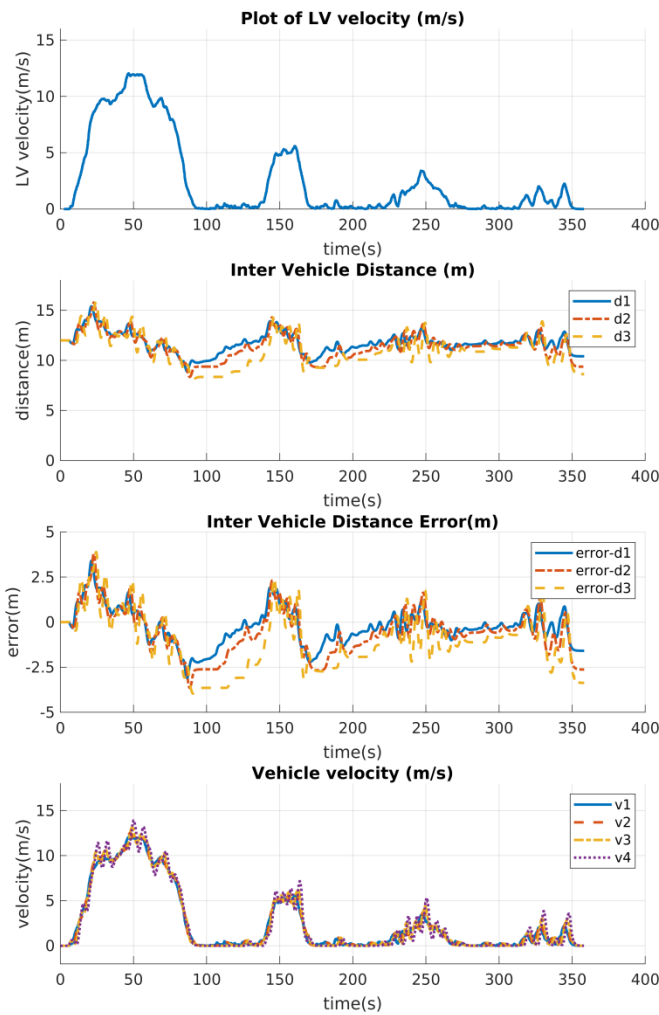


Fig.10. Scenario 4, first case study

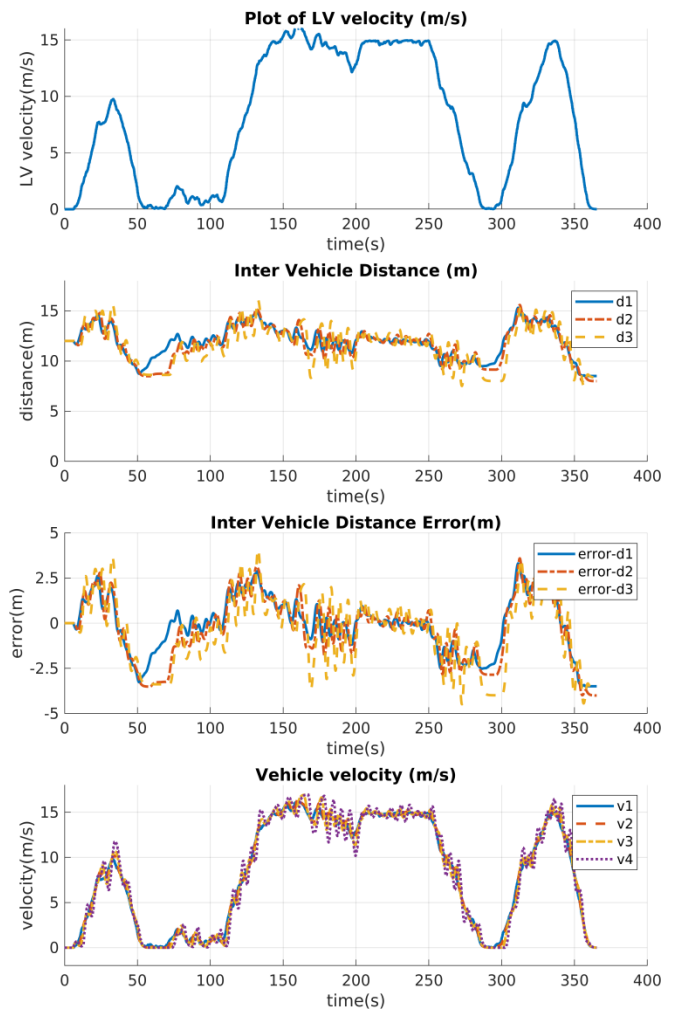


Fig.11. Scenario 4, second case study

VI. DISCUSSION

In this study, we provide a computationally less demanding longitudinal inter-vehicle distance control algorithm for a platoon of autonomous vehicles that only requires that vehicles be equipped with GPS sensors and have a connection to Wi-Fi. The algorithm computation cost is less than most control algorithms described in the literature. For example, physics-inspired control algorithms such as [22], most of the other PID-based algorithms with different approaches like taking as input to the PID the preceding vehicle's velocity and acceleration e.g., [5] and [23]. Furthermore, this study differs from those that only perform numeric simulations such as [5] by not only using generated GPS data, but also by applying the algorithm to the 3D models of the platoon to mimic the real world as closely as possible.

However, our approach does have limitations. The first limitation of this algorithm stems from the fact that it is mainly based on GPS sensors. GPS, in reality, is affected by high-frequency faults culminating from multipath errors that occur when signals bounce off surfaces before they can reach the sensor receivers. The position fix, therefore, gets affected as the signals are delayed. Another rarer cause of GPS faults happens when one of the satellites used by the sensor receiver

gets blocked and, as a result, has to be compensated by signals received from a different satellite. The position fix estimated by the GPS sensor is affected by the geometry of the satellites from which the sensor gets signals. So, such changes in configurations of the satellite observed by the sensor receiver affect the position fix finally reported by the GPS receiver.

High-frequency faults and multipath make the accuracy of GPS sensor, and ultimately, the efficiency of our algorithm heavily environment-dependent, making it more accurate and preferable in open space areas than in underground passages, enclosed environments, or places with tall buildings such as skyscrapers. The algorithm can be incorporated into indoor environment or closed environments by replacing the GPS technology with higher precision localization tools and, or sensors such as beacon technology illustrated by [27] and [28].

Autonomous vehicle platooning applications in urban areas involve high precision dependent maneuvers that require about $0.02m$ accuracy to guarantee safety, among other requirements - such as lane-keeping/changing on busy streets, overtaking operations, to mention but a few. In such applications, a $0.5m$ error is pretty significant. Thus, expanding the applicability of the proposed algorithm to all types of roads and environments requires fusing data from other sensors such as LIDAR, the Inertial Measurement Unit (IMU), and camera. Using sensor fusion to enhance the applicability of the proposed algorithm

and make it more ideal for even more complex environments is one of our future studies.

VII. CONCLUSION

In this study, a computationally-cost efficient algorithm for the control of the inter-vehicle distance of autonomous vehicle platoons is presented. The proposed approach takes as input to the PID controller, the updated inter-vehicle distance between a follower vehicle and the preceding vehicle. This distance is calculated from the data measured and provided by the vehicles' onboard GPS sensors. The controller returns the reference velocity with which the follower vehicle should move to achieve the desired inter-vehicle distance. 3D simulations using gazebo and ROS are additionally used to verify and monitor the performance of the system. The proposed approach guarantees the following effect of the platoon ensuring maintenance of platoon formation and no collisions among platoon members. Furthermore, after the transient response to the Leader vehicle's acceleration, the standard deviation of the inter-vehicle distance error was kept under 14.7% of the desired inter vehicle distance throughout the entire simulation period for all the scenarios. The system seldomly achieved a $0m$ error at the steady state when the leader moves with constant speed. However, the proposed method is mainly suitable for open environments since GPS accuracy is susceptible to High-Frequency errors resulting from multipath and collision of GPS signals with surfaces before they reach the receiver. Applicability of the approach can be extended to closed and underground environments if GPS is replaced with high precision localization equipment such as position beacons installed in the target environments. We are currently working on expanding the operability and applicability of the proposed approach on more road types, environments (urban, rural, to mention but a few). Incorporating more sensors and sensor fusion techniques to improve the accuracy of not only the inter-vehicle distance but also the velocity of the platoon members is another direction for our future work.

ACKNOWLEDGMENT

This work was supported by Istanbul Sabahattin Zaim University's Scientific Research Projects Programme (BAP) , under Project No: BAP-1001-61.

REFERENCES

- [1] R. Kimura, N. Matsunaga, H. Okajima, and G. Kotaki, "Design of virtual platoon control system using augmented reality to assist welfare vehicle users," *International Conference on Control, Automation and Systems*, vol. 2017-October, no. Iccas, pp. 330–335, 2017..
- [2] F. Luo, J. Larson, and T. Munson, "Coordinated platooning with multiple speeds," *Transportation Research Part C: Emerging Technologies*, vol. 90, pp. 213–225, 2018.
- [3] R. Janssen, H. Zwijnenberg, I. Blankers, and J. de Kruijff, "Truck platooning," *Driving The Future of Transportation*, TNO, 2015.
- [4] S. Belcher, E. Merlis, J. McNew, and M. Wright, "Roadmap To Vehicle Connectivity," funded by Crown Castle, Tech. Rep. September, 2018.
- [5] V. V. Sivaji and M. Sailaja, "Adaptive cruise control systems for vehicle modeling using stop and go manoeuvres," *International Journal of Engineering Research and Applications*, vol. 3, no. 4, pp. 2453–2456, 2013. [Online]. Available: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.433.9908&r=ep1&type=pdf>
- [6] S. E. Shladover, C. Nowakowski, X. Y. Lu, and R. Ferlis, "Cooperative adaptive cruise control: Definitions and operating concepts," *Transportation Research Record*, vol. 2489, no. November 2014, pp. 145–152, 2015.
- [7] G. Naus, R. Vugts, J. Ploeg, R. Van De Molengraft, and M. Steinbuch, "Co-operative adaptive cruise control, design and experiments," *Proceedings of the 2010 American Control Conference, ACC 2010*, no. 1, pp. 6145–6150, 2010.
- [8] D. Swaroop, "String Stability of Interconnected Systems – Automatic Control, IEEE Transactions on," *IEEE Transactions on Automatic Control*, vol. 41, no. 3, pp. 349–357, 1996.
- [9] M. Cremer, "On Convoy-Stable Control Laws for Automatically Driven Vehicle Clusters," 1992.
- [10] S. Öncü, N. Van de Wouw, W. M. H. Heemels, and H. Nijmeijer, "String stability of interconnected vehicles under communication constraints," in *2012 IEEE 51st IEEE conference on decision and control (cdc)*. IEEE, 2012, pp. 2459–2464.
- [11] T. L. Willke, P. Tientrakool, and N. F. Maxemchuk, "A survey of inter-vehicle communication protocols and their applications," *IEEE Communications Surveys Tutorials*, vol. 11, no. 2, pp. 3–20, 2009.
- [12] M. Jain and R. Saxena, "Overview of VANET: Requirements and its routing protocols," in *2017 International Conference on Communication and Signal Processing (ICCSP)*. IEEE, 2017, pp. 1957–1961.
- [13] Z. Shen, X. Zhang, D. Yang, I. I. S. Enzor, and S. H. N. Etwork, "Performance Analysis of Extended Sensor Sharing in Vehicular Ad Hoc Networks," *2018 International Symposium on Antennas and Propagation (ISAP)*, pp. 1–2, 2018.
- [14] M. Quigley, "ROS: an open-source Robot Operating System," in *ICRA 2009*, 2009.
- [15] P. Seiler, A. Pant, and K. Hedrick, "Disturbance propagation in vehicle strings," *IEEE Transactions on automatic control*, vol. 49, no. 10, pp. 1835–1842, 2004.
- [16] D. Swaroop, J. K. Hedrick, C. Chien, and P. Ioannou, "A comparison of spacing and headway control laws for automatically controlled vehicles," *Vehicle system dynamics*, vol. 23, no. 1, pp. 597–625, 1994.
- [17] S. Edelkamp, S. Jabbar, and T. Willhalm, "Geometric travel planning," *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*, vol. 2, no. 1, pp. 964–969, 2003.
- [18] S. Sukkarieh, E. M. Nebot, and H. F. Durrant-Whyte, "A high integrity IMU/GPS navigation loop for autonomous land vehicle applications," *IEEE Transactions on Robotics and Automation*, vol. 15, no. 3, pp. 572–578, 1999.
- [19] O. Karoui, M. Khalgui, A. Koubâa, E. Guerfala, Z. Li, and E. Tovar, "Dual mode for vehicular platoon safety: Simulation and formal verification," *Information Sciences*, vol. 402, pp. 216–232, 2017.
- [20] E. T. S. Institute, "ETSI TR 103 299 V2.1.1 Intelligent Transport Systems (ITS); Cooperative Adaptive Cruise Control (CACC);," Tech. Rep., 2019. [Online]. Available: <https://cdn.standards.iteh.ai/samples/45854/7b3e6d0d3c014cb995408e4a8a0a87d5/ETSI-TR-103-299-V2-1-1-2019-06-.pdf>
- [21] R. Rajamani and S. E. Shladover, "An experimental comparative study of autonomous and co-operative vehicle-follower control systems," *Transportation Research Part C: Emerging Technologies*, vol. 9, no. 1, pp.15–31, 2001.
- [22] I. Edmunds. (2014) Used 2014 hyundai genesis specs & features. Used 2014 Hyundai Genesis Specs & Features. [Online]. Available: <https://www.edmunds.com/hyundai/genesis/2014/features-specs/>
- [23] R. Rajamani, *Vehicle dynamics and control*. Springer Science & Business Media, 2011
- [24] M. Long, G. Tian, and H. Cheng, "Longitudinal control for truck platooning," in *2020 4th CAA International Conference on Vehicular Control and Intelligence (CVCI), 2020*, pp. 418–423.
- [25] S.-Y. Yi and K.-T. Chong, "Impedance control for a vehicle platoon system," *Mechatronics*, vol. 15, no. 5, pp. 627–638, 2005. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0957415805000279>
- [26] Z. Ali Memon, S. Jumani, and J. Larik, "Longitudinal Control of a Platoon of Road Vehicles Equipped with Adaptive Cruise Control System," vol. 31, no. 3, pp. 475–494, 2012. [Online]. Available: <http://oaji.net/articles/2016/2712-1454749706.pdf>
- [27] D. Surian, V. Kim, R. Menon, A. G. Dunn, V. Sintchenko, and E. Coiera, "Tracking a moving user in indoor environments using

Bluetooth low energy beacons,” Journal of Biomedical Informatics, vol. 98, p. 103288, 2019. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1532046419302072>

[28] R. Siegwart and I. R. Nourbakhsh, Introduction to Autonomous Mobile Robots. USA: Bradford Company, 2004.

BIOGRAPHIES



GUNAGWERA ALEX received his M.S. degree in computer science from Istanbul S. Zaim University, Istanbul, Turkey in 2017 where he is currently pursuing the Ph.D. degree in computer engineering.



ZENGIN A. TARIK received his electrical and electronics engineering B.S. degree from Ege University, Turkey, in 2007. Later on, received M.E. and Ph.D. degrees from Department of computer science and electrical engineering, Kumamoto University, Japan, in 2010 and 2013, respectively. Currently, he is an Asst. Prof. at Istanbul S. Zaim University. His research interests include Autonomous systems and control theory.

Cognitive Based Electric Power Management System

Tahir Cetin Akinci and Alfredo A. Martinez-Morales


Abstract—An electric power network can be evolved into smart grids, which are measured by providing energy efficiency and improving the available resources. With the development of software and hardware elements, the decision-making mechanism of existing smart grids is transformed into more robust uninterrupted and economical energy management systems. In this study, a cognitive-based algorithm using dynamic energy management flexibility, storage and energy management algorithm and cloud computing architecture is proposed. Using this approach, an uninterrupted and economical energy management system can be planned. In addition, the proposed approach provides the optimization of supply and demand sides.

Index Terms—Cognitive Management, Electric Power Management, Electric Power Price Management System, Cognitive Network System.


I. INTRODUCTION

COGNITIVE engineering is an area that provides decision support, human-automation interaction by redesigning existing knowledge in cognitive science and engineering [1]. This area is used in all risky areas from health to energy. Cognitive engineering; cognitive science, data science, data analysis methods and analytical and software-based methods are combined. For example, in control engineering applications, controllers in the role of human beings who manage semi-autonomous systems can be considered an application for cognitive systems. These semi-autonomous systems reduce the physical workload of people while reducing human-induced failures. Almost all of these applications are computer-based systems. Thus, the development of the experience of computer systems and software technology will enable the creation of more cognitive technology environments in the future [1-4].

TAHIR CETIN AKINCI, is with Department of Electrical Engineering University of Istanbul Technical University (ITU), Istanbul, Turkey, (e-mail: akincitc@itu.edu.tr). And University of California Riverside, WCGEC, Riverside, CA, USA. (tahircetin.akinci@ucr.edu)

 <https://orcid.org/0000-0002-4657-6617>

ALFREDO A. MARTINEZ-MORALES, is with CE-CERT, University of California Riverside (UCR), Riverside, CA, USA, (e-mail: alfmart@ece.ucr.edu).

 <https://orcid.org/0000-0003-4204-2228>

Manuscript received September 12, 2021; accepted January 08, 2022.
DOI: [10.17694/bajece.1060998](https://doi.org/10.17694/bajece.1060998)

Cognitive Engineering methods allow you to solve natural complex problems in the workspace by giving them simple attributes. The results of the analyses obtained as a result of Cognitive Engineering methods can be put into action and can be designed [3-7].

A. Cognitive Approach to Electrical Power Engineering

Computer and software technology are also closely related to energy management systems. Intelligent systems with bidirectional communication capability, which provide the energy supply and the coordination between the demanders, are developing every day [8]. One of the most important features of the smart grid is to provide energy efficiency and management of existing resources [9]. Smart Electricity Networks are focused on providing the most economical way to ensure the continuity of energy transmission and distribution in the coordination of all the elements within the system. Energy quality, continuity and energy economy are the three most important elements of energy transmission and distribution systems. The integration of these elements ensures the flexibility of the energy systems. Energy security is another of the elements that are considered in terms of energy security both by supply and demand [8-11]. In this context, smart grid systems enable the communication of information between consumers and producers. In Figure 1, this flow in energy architecture is given [12].

This mutual communication in the intelligent network architecture helps control the supply and demand balance between producers and consumers by checking every moment of the detection and measurement components on energy power systems. Nowadays, public service companies use smart meter technology to gain access to the requested information to monitor the supply-demand balance. Besides, the supply or demand side of the energy can be measured by wired or wireless methods by developing sensor technology. The obtained information is combined and system response is prepared by using decision-making algorithms. The Smart Grid market is developing in a direction that will encourage both the customer and the manufacturer with the development of data analysis application methods. Recently, information and communication technologies are used in electrical power systems. Electrical energy management algorithms provide important solutions to power decision-making in efficient ways. Especially successful in reducing consumer electricity energy bills is achieved.

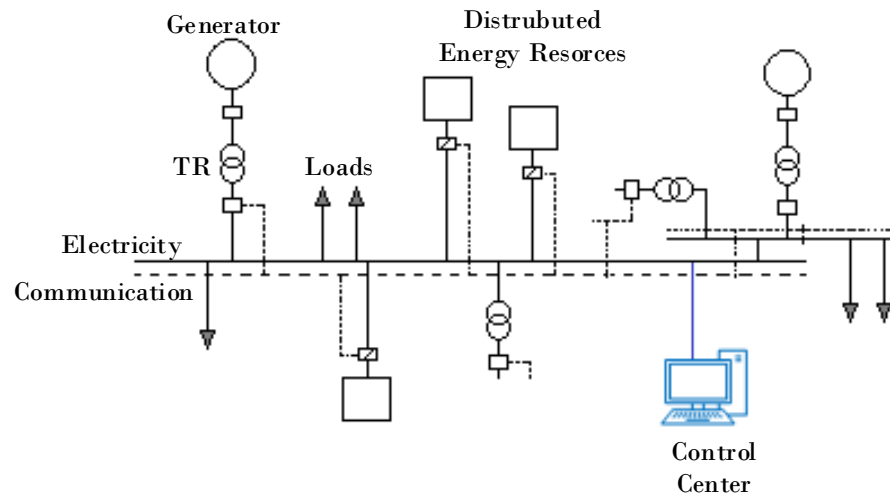


Fig.1. Architecture of a Typical Energy System

In addition, energy management, autonomy practices in controlling different power grid subsystems (energy) allow consumers to choose suppliers. In the literature, many definitions of energy management have been associated with the energy strategy [14-19]. In this sense, it would not be wrong to define and associate energy management with planning, strategy and policy. These definitions may reveal the idea of controlling the management approach in the distribution and planning of energy transmission with a Cognitive-based approach.

In this study, a cognitive-based electrical energy management system was researched.

II. ENERGY MANAGEMENT REQUIREMENTS AND ELECTRICITY MARKETS

The increase in the population and the development of the industry increase the demand for electrical energy. Energy providers have to provide steady, high quality, continuous and economic energy to meet the growing demands. However, the fact that energy cannot be supplied from a single source and from a single producer can lead to complex problems, including environmental, commercial and sometimes international agreements. While the demand response is required for the stability of the electricity grid operation, the energy supply side forces consumers to offer tariffs that motivate them to save energy and/or shift loads into intense periods. In addition, unbalanced energy usage demands reveal the problems of buffering and storing energy.

III. DEMAND RESPONSE FOR ELECTRIC POWER DISTRIBUTION

In the energy markets, the Demand Response (DR) could be considered as a stimulating opportunity for energy consumers to participate in these electricity markets. Smart grids are important for the integration of DR, both by demand and by energy supply organizations. The main role of DR in the electricity market is that it economically optimizes supply-demand balance during peak hours. Without changing the

energy consumption during the high price hours, it shifts some of the high demand processes to non-peak periods and makes optimum adjustments [8,13,14].

IV. CLOUD COMPUTING

Many authors used cloud computing in smart grid energy management. In this section, we will present some of the proposed algorithms. Cloud systems are hierarchically placed into three basic groups. These are Hybrid, Public and Private. (Figure 2). Within this group, users have the possibility to select Cloud that is suitable for data processing capacity [8,18].

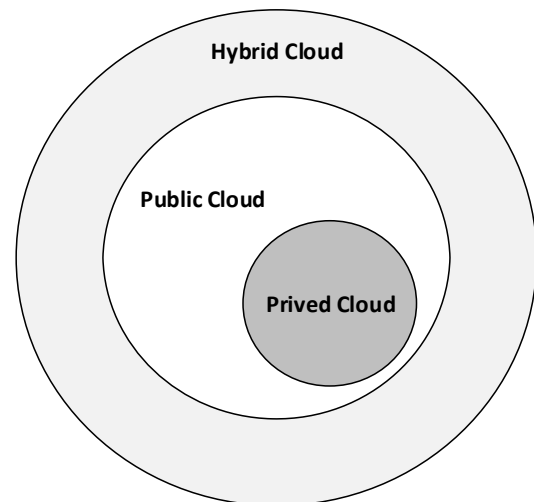


Fig.2. Cloud Systems Groups [8,18].

Cloud Computing (CC) is defined as a data sharing based model. CC enables independent operating systems, data warehouses and hardware infrastructures to be shared via software via servers. It provides users with virtual machines, software, and storage physical virtual resource infrastructure. CC allows users to run applications without installing software

on their own computers and have a wide range of uses [18] (Figure 3).

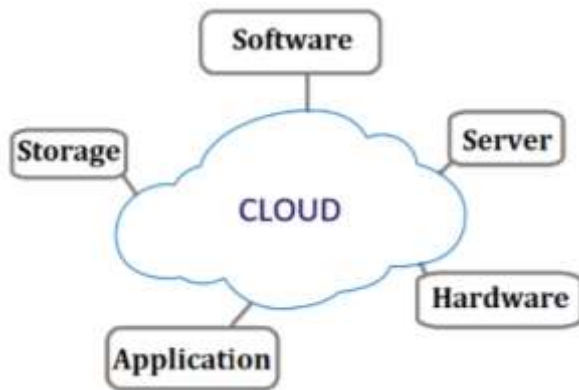


Fig. 3. General Uses of Cloud Systems

A. Fog Computing

Fog Computing is an extension or part of the Cloud network. However, it is more accessible and less extensive [21,22]. With Fog Computing, end users can carry out data processing, storage, and service between the traditional Cloud systems. Fog Computing has many advantages. These; fast, high geographic distribution, a large number of real-time interaction. In terms of quality, Fog Devices have three levels of hierarchy. This hierarchy; Cloud Level can be classified as Fog Level and Smart Devices. Fog Computing very suitable for intelligent grid applications in electrical power systems. In the networks, energy load balancing, smart meters and microgrid applications are the most common applications. It also offers practical algorithm approaches for planning and management for power systems [22].

V. COGNITIVE SYSTEMS

Human-based expert systems are defined as cognitive systems. Establishing a knowledge-based system by transferring human knowledge and expertise to learning systems with programming systematic is defined as Cognition. Cognitive systems engineering has recently played a key role in science and technology. Since cognitive systems are the systems that learn and think of human beings, a human-based expert enables cognition in the system through learning algorithms and logic [23]. Cognition terminology covers mental processes such as perception, memory, thought, and learning [24]. Today, cognition has very popular applications in social sciences and science. In engineering, the term Cognitive Systems finds application and is considered a set above artificial intelligence methods. The increase of cognitive-based applications in industrial and engineering systems makes the performance of the systems determine [6, 25].

When the cognitive decision-making mechanism is considered as the supercluster of artificial intelligence, the relationship between the concept of artificial intelligence and the cognitive system can be expressed as in Figure 4. In this sense, cognitive system engineers take part in the design of

human processes, communication systems, teams and management systems that coordinate various activities within the specified processes. Cognitive System engineering will be the key engineering role in the control of many critical and complex systems in the future.

Information processing is performed by processors in electronic equipment. In fact, processors are not just hardware on computers. Many electronic systems now have a processor, which allows control with the processor. The first processor was the beginning of a new era in the hardware world, thus enabling the development of the computers we use today. However, the most important information processor is the human brain. The human brain system attempts to achieve a match between models and objects by using real-life objects or mathematical models. They are stored in the processor's memory, creating inputs and outputs to generate symbolic expressions represented by some symbols. In an immense computing system such as the human brain, the mathematical or abstract model is provided by the capability of the processing units. This structure can be explained by Cognitive Systems or Cognitive engineering [26,27].

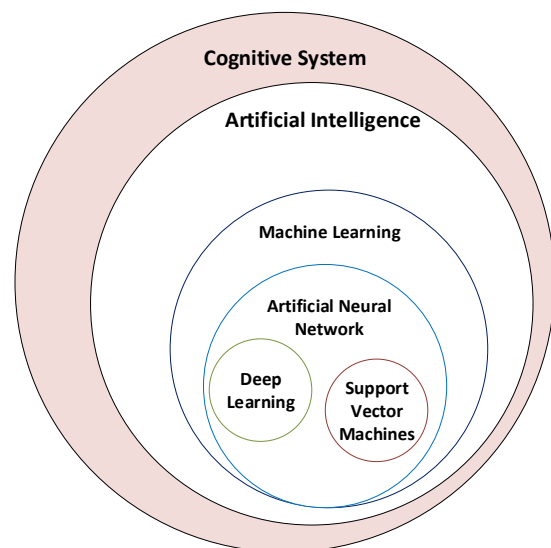


Fig.4. Relationship between Cognitive Systems and Artificial Intelligence Methods

A. The framework of Cognitive Learning Systems

Cognitive Systems (CS) is a new discipline in which brain and natural intelligence are shared by psychology, philosophy, mathematics, statistics and engineering disciplines. This discipline, covers many topics, from social psychology to industrial autonomous systems. In the field of engineering, artificial intelligence, statistics and software engineering are widely used. It is possible to explain the mathematical background representing the cognitive systems with the Cognitive Information Model (CIM). The concept of examining the four basic sections of the CIM makes it easier to explain. These; Knowledge (K), Behaviour (B), Experience (Exp), and Skill ($Skil$). Learning (L_{learn}) is expressed here;

CIM model of knowledge (\mathcal{L}_{know}), experience (\mathcal{L}_{exp}) and skill (\mathcal{L}_{skil}) is expressed by [15-17].

$$\mathcal{L}_{Learn} = \{\mathcal{L}_{know}, \mathcal{L}_{beh}, \mathcal{L}_{exp}, \mathcal{L}_{skil}\} \quad (1)$$

It can also be expressed here with Cognitive Learning (\mathcal{L}_{cog}). In fact, Cognitive Learning is a form of Transitive Learning.

$$\begin{aligned} \mathcal{L}_{cog} &\triangleq \{\mathcal{L}_{know}, \mathcal{L}_{beh}, \mathcal{L}_{exp}, \mathcal{L}_{skil}\} \\ \mathcal{L}_{cog} &\triangleq f_{cog}: (STM) \rightarrow K(LTM) \end{aligned} \quad (2)$$

Perceptive Cognitive Behavioural;

$$\begin{aligned} \mathcal{L}_{pcogbeh} &\triangleq \mathcal{L}_p \rightarrow \mathcal{L}_{cog} \rightarrow \mathcal{L}_{beh} \\ f_{pcogbeh}: D(SBM) &\rightarrow I(STM) \rightarrow K(LTM) \rightarrow B(ABM) \end{aligned} \quad (3)$$

Cognitive Introspective;

$$\begin{aligned} \mathcal{L}_{cogint} &\triangleq \mathcal{L}_{cog} \rightarrow \mathcal{L}_{int} \\ f_{cogint}: I(STM) &\rightarrow K(LTM) \rightarrow I(STM) \end{aligned} \quad (4)$$

Equations 3 and 4 are shown in Cognitive Models of Memory (CMM), Sensory Buffer Memory (SBF), Short Term Memory (STM), Long Term Memory (LTM) and Action Buffer Memory (ABM) [15-17]. In this sense, it can be used in the planning and pricing of the mentioned learning systems, energy management systems.

VI. CONCEPT OF COGNITIVE ENERGY MANAGEMENT SYSTEM

Cognitive systems are topologically related to a lot of information and data. Therefore, it may be possible to explain the cognitive system model obtained by human-based expert and neural network with this procedure.

In this algorithm, logical coding and data structure is applied to train information through the neural network. As a result, this newly trained neural network plays the role of logic modelling. And such a system can be used in the electrical power systems management system. The algorithm flow chart of this system is shown in Figure 5.

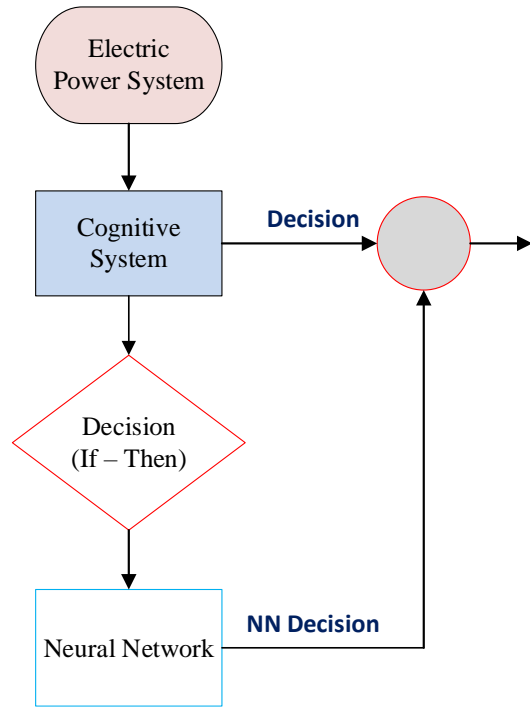


Fig. 5. Cognitive Electric Power Flowchart

As shown in the flowchart shown in Figure 5, the second block has a cognitive or human-based (expert) operator. The block with the if-then structure is associated with the neural network block for the decision.

Although the decision structure is given symbolically as a neural network, it works as a machine learning decision-maker in this block.

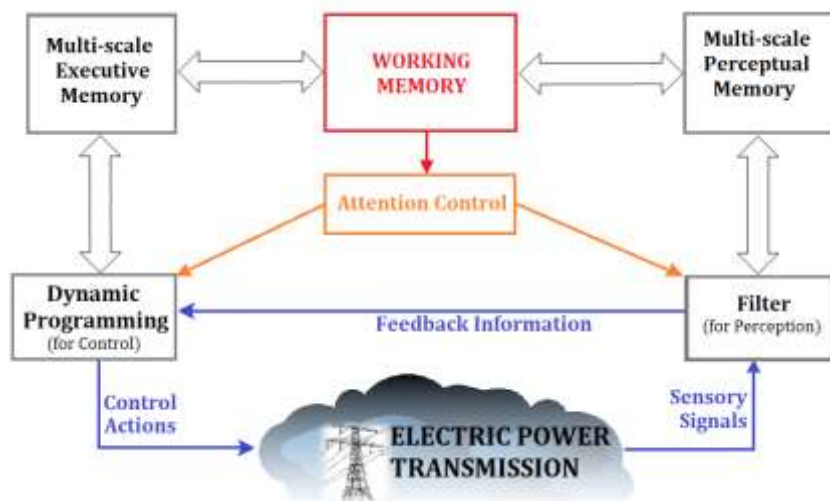


Fig. 5. Block diagram of Electric Power Transmission Management System [12].

The Cognitive Energy Management System (CEMS) controls and plans all stages of energy transmission from generation to end-user through the cloud. It includes hardware infrastructure as well as software and market research. Working Memory controls the operation of collecting all data. Below it is Attention Control. Dynamic Control and Perception Filters balances the system at every stage of the organization (Fig.5). The main purpose of Working Memory is to control the data on the Electric Power Transmission Line using sensors.

Block diagram of Cognitive base Electric Power Price Management System is shown in Figure 6. Here, the decision to determine the price in the energy markets is determined by the Cognitive system over the cloud.

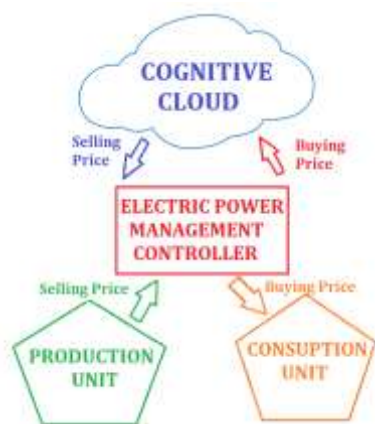


Fig. 6. Block diagram of Cognitive base Electric Power Price Management System

VII. CONCLUSION

The integration of electric power systems is becoming more and more important. The main objective of the integration is to ensure the transmission of quality and uninterrupted energy to the users by using flexible integration strategies of energy sources. In this study, the management and planning of cognitive-based energy systems and the organization of the energy economy are examined. In addition, pricing management has been examined under the control of Electric Power Systems. Relationships between working memory and control actions are given in energy transmission lines. In this study, a cognitive-based algorithm that uses a storage and energy management algorithm and cloud computing architecture is proposed. This algorithm also has dynamic energy management flexibility.

ACKNOWLEDGMENT

This study was supported by TUBITAK (The Scientific and Technological Research Council of Turkey) 2219-International Postdoctoral Research Scholarship Program, with project number 1059B192001347 in the 2020/2 application period.

The authors would like to thank TUBITAK for their support.

REFERENCES

- [1] A.Z. Hettinger, E.M. Roth, A.M. Bisantz, "Cognitive Engineering and Health Informatics: Applications and Intersections", *Journal of Biomedical Informatics*, Vol.67, pp.21-33. 2017.
- [2] Y. Wang, "The Theoretical Framework of Cognitive Informatics", *International Journal of Cognitive Informatics and Natural Intelligence*, Vol.1, No.1, 2007, pp. 1–27.
- [3] V.L. Patel, T.G. Kannampallil, "Cognitive Informatics in Biomedicine and Healthcare", *Journal of Biomedical Informatics*, Vol.53, 2015, pp. 3-14.
- [4] M.R. Endsley, R. Hoffman, D. Kaber, E. Roth, "Cognitive Engineering and Decisionmaking: an Overview and Future Course", *Journal of Cognitive Engineering and Decision Making*, Vol.1, No.1, 2007, pp.1-21.
- [5] D. A. Norman and S. W. Draper, *User Centered System Design: New Perspectives on Human-Computer Interaction*, CRC Press, 1986.
- [6] E. Hollnagel, and D.D. Woods, "Cognitive Systems Engineering: New Wine in New Bottles", *International Journal of Man-Machine Studies*, 1983, Vol.18, pp.583–591.
- [7] E. Hollnagel, D.D. Woods, *Joint Cognitive Systems: Foundations of Cognitive Systems Engineering*, CRC Press, 2005, 113-155.
- [8] J. Klaimi, R. Rahim-Amoud, L. Merghem-Boulahia and A. Jrad, *Energy Management Algorithms in Smart Grids: State of the Art and Emerging Trends*, *International Journal of Artificial Intelligence and Applications (IJAA)*, Vol.7, No.4, 2016, pp.25-45.
- [9] G. Yuanxiong, M. Pan, Y. Fang, "Optimal Power Management of Residential Customers in the Smart Grid", *IEEE Transactions on Parallel and Distributed Systems*, Vol.23, No.9, 2012, pp.1593-1606.
- [10] R. Belkacemi, A. Bababola, *Experimental Implementation of Multi-Agent System Algorithm for Distributed Restoration of a Smart Grid System*, *IEEE SOUTHEASTCON*, 2014.
- [11] T. Nagata, Y. Ueda, M. Utatani, *A multi-agent approach to smart grid energy management*, *10th International Power & Energy Conference (IPEC)*, pp.327-331, Dec 2012.
- [12] Y. Xue, S. Sirouspour, and A. Emadi, *Towards Cognitive Energy Management System of Microgrid in Enabling Transportation Electrification*, *2012 IEEE Transportation Electrification Conference and Expo (ITEC)*.
- [13] B. Bitzer, and E.S. Gebretsadik, *Cloud Computing Framework for Smart Grid Applications*, *48th International Universities' Power Engineering Conference (UPEC) 2013 48th International Universities'*, 2013, pp.1-5.
- [14] S. Pierluigi, "Demand response and Smart Grids-A Survey", *Renewable and Sustainable Energy Reviews*, Vol.30, 2014, pp.461-478.
- [15] Y. Wang, *Using Process Algebra to Describe Human and Software System Behaviors*. *Brain and Mind: A Transdisciplinary Journal of Neuroscience and Neurophilosophy*, Vol.4, No.2, 2003, pp.199–213.
- [16] Y. Wang, *The Theoretical Framework and Cognitive Process of Learning*, *Proc. 6th IEEE International Conference on Cognitive Informatics (ICCI'07)*, 2007, pp.470-479.
- [17] L. Ogiela, "Towards Cognitive Economy", *Soft Computing*, 2014. Vol.18, No.9, pp.1675-1683.
- [18] D.C. Wyld, *Cloud Computing and the Public Sector Around the World*, *International Journal of Web & Semantic Technology*, Vol.1, No.1, 2010, pp.1-20.
- [19] H.F. Atlam, R.J. Walters and G.B. Wills, "Fog Computing and the Internet of Things: A Review", *Big Data Cognitive Computing*, Vol.2, No.10, 2018, pp.2-18.
- [20] O. Aizpurúa, R. Galán, A. Jiménez, *A New Cognitive-Based Massive Alarm Management System in Electrical Power Administration*, *Proceedings of the 7th International Caribbean Conference on Devices, Circuits and Systems*, Mexico, Apr. 28-30, 2008.
- [21] E.M. Rantanen, J.A. Winkle, T.J. Overbye, *Cognitive Task Analysis of Electric Power System Control Center Operations*, *Human Factors and Ergonomics Society Annual Meeting Proceedings*, Vol.52, No.24, pp.1934-1938 · September 2008.
- [22] Ni, J.; Zhang, K.; Lin, X.; Shen, X. *Securing Fog Computing for Internet of Things Applications: Challenges and Solutions*. *IEEE Communication Surv. Tutor.* 2017, 20, 601–628.

- [23] Miller, G. The cognitive revolution: a historical perspective, Trends in Cognitive Sciences, 2003, Vol.7, No.3, pp.141-144.
- [24] Seker, S, Dikun, J. Physical Limits for Self-Contained Systems: An Example of the Human Brain as a Cognitive System. The Journal of Cognitive Systems, 2018, Vol.3, No.2, pp.28-29.
- [25] Militello, L.G., Dominguez, C.O., Lintern, G., Klein, G. The Role of Cognitive Systems Engineering in the Systems Engineering Design Process, System Engineering, 2010, Vol.13, No.3, pp.261-273.
- [26] Alafi, B . A New Cognitive System Modeling with the Combination of Expert and Neural Network Systems. The Journal of Cognitive Systems, 2018, Vol.3, No.2, pp.34-35.

BIOGRAPHIES



TAHIR CETIN AKINCI received bachelor's degree in electrical engineering in 2000. Dr. Akinci received Master's and PhD degrees from Istanbul Marmara University in 2005 and 2010, respectively. He works as a full professor in the Department of Electrical Engineering at Istanbul Technical University (ITU). At the same time, he works as a visiting scholar at the University of California Riverside (UCR) to do his postdoctoral research.

His research studies includes on Cognitive systems, Artificial intelligence methods, Machine learning, Electric power systems and data analytics.



ALFREDO A. MARTINEZ-MORALES is the Managing Director of the Southern California Research Initiative for Solar Energy (SC-RISE) and Research Professor at the Bourns College of Engineering Center for Environmental Research and Technology (CE-CERT). Dr. Martinez-Morales received his Ph.D., M.S. and B.S. degrees in electrical engineering from University of California Riverside (UCR).

His current research includes solar cells, alkali metal-ion batteries, highly integrated renewables, energy storage systems, and microgrids. Dr. Martinez-Morales is a principal investigator in the Sustainable Integrated Grid Initiative (SIGI) at UCR, and has contributed in the engineering, permitting, and deployment of the SIGI smartgrid testbed system and multiple microgrids throughout Southern California.

Development of a Python-Based Classification Web Interface for Independent Datasets


İpek Balikci Cicek, İlhami Sel, Fatma Hilal Yagin and Cemil Colak


Abstract—Classification; biomedical, bioinformatics, medicine, engineering etc. It is a fundamental approach that is frequently used in many research areas, such as especially in the field of health; it has become common to classify diseases with machine learning methods using risk factors of these diseases and to determine the effect levels of these risk factors on the related disease. There are both commercial and free software tools that researchers can analyze their data with classification methods. The aim of this study is to develop a user-friendly web-based software for classification analysis. Python sklearn and Dash libraries were used during the development of the software. Among the classification algorithms in the developed software; Logistic regression, Decision trees, Support vector Machines, Random Forest, LightGBM, Gaussian Naive Bayes, AdaBoost and XGBoost methods are available. In order to show how the software works, a classification model was created with the Random forest algorithm using the cervical cancer data set. Different metric values were evaluated for the models. Obtained from a random forest classification model; accuracy, sensitivity, specificity, negative predictive value, matthews correlation coefficient, and F1 score values obtained from the model were 94.44%, 100%, 93.33%, 100%, 83.67%, and 94.44 respectively. It is thought that the classification software developed in this study will provide great convenience to clinicians and researchers in the field of medicine, in terms of applying predictive classification algorithms for the disease without any software knowledge.


Index Terms— Classification, machine learning, web based software.


I. INTRODUCTION

THE NUMBER of data produced in parallel with the developing technology is increasing day by day.

İPEK BALIKCI CICEK, Inonu University Department of Biostatistics and Medical Informatics, Faculty of Medicine, Malatya, Turkey, (ipek.balikci@inonu.edu.tr)  <https://orcid.org/0000-0002-3805-9214>

İLHAMİ SEL, Inonu University Computer Engineering Department, Engineering Faculty, 44280 Malatya, Turkey, (ilhamisel23@gmail.com)  <https://orcid.org/0000-0003-0222-7017>

FATMA HILAL YAĞIN, Inonu University Department of Biostatistics and Medical Informatics, Faculty of Medicine, Malatya, Turkey, (hilal.yagin@inonu.edu.tr)  <https://orcid.org/0000-0002-9848-7958>

CEMİL COLAK, Inonu University Department of Biostatistics and Medical Informatics, Faculty of Medicine, Malatya, Turkey, (cemil.colak@inonu.edu.tr)  <https://orcid.org/0000-0001-5406-098X>

Manuscript received June 9, 2021; accepted Jan 26, 2022.

DOI: [10.17694/bajece.949935](https://doi.org/10.17694/bajece.949935)

It is of great importance to store, manage and make useful the enormous amount of data produced. Therefore, it is of great importance to be able to use techniques that can process large amounts of data. Data mining is the process of discovering patterns and trends hidden in large data sets [1].

Data mining is a technique that attempts to identify previously unknown hidden relationships among data in databases [2]. Data mining is a multidisciplinary field that bridges many technical fields such as database technology, statistics, artificial intelligence, machine learning, pattern identification and data visualization [1].

Models used in data mining are examined under two main headings: predictive and descriptive. In predictive models, it is aimed to develop a model based on the data with known results and to estimate the result values for datasets with unknown results by using this established model. In descriptive models, patterns in existing data that can be used to guide decision making are defined [3].

It is possible to examine data mining models under three main headings, namely classification and regression, clustering and association rules, according to their functions. Classification and regression models are predictive models, clustering and association rules models are descriptive models [4]. It is of great importance that the patterns in the heap datasets are extracted by data mining/machine learning techniques and used as prediction and decision support components. In this context, one of the frequently applied data mining/machine learning topics is classification [5].

Classification is a very common process in scientific studies because of the benefits it provides in solving problems. Especially in the field of medicine, the classification of diseases and the development of treatment methods according to this classification are among the most prominent examples. In addition to medicine, the functionality of classification can be seen in other branches of science. Classification is an estimation process that assigns the observations that make up the data set to previously determined classes within the framework of certain rules [6]. Each observation in the data set has a feature and these features are divided into classes. Creates a model with the observations determined to belong to which class. The success of the model is measured with the observations that are not included in the training set. Quite different algorithms are used in classification methods. Different results have been obtained due to the fact that there are many algorithms. Each algorithm works with different parameters and has more than one version. The studied algorithms are for different purposes, the data source used is different, the algorithms support different data types, and the

preprocessing on the data depends on the practitioner. Knowing in which areas and in which types of variables these algorithms give more effective results increases the success of the methods. For this reason, it is important to apply classification methods by comparing them [7].

In this research, we aimed to develop a new user-friendly web-based software developed with the Python Dash library that will allow the comparison of Logistic regression, Decision trees, Support vector Machines, Random Forest, LightGBM, Gaussian Naive Bayes, AdaBoost and XGBoost data classification methods.

II. MATERIAL AND METHODS

A. Dataset

The UCI data repository's open access dataset "Cervical Cancer Behavior Risk Data Set" was used to demonstrate how the software works. The dataset includes 72 cervical cancer samples with 18 predictive variables and one outcome variable. 50 (69.45%) of the samples tested negative for cervical cancer, while 22 (30.55%) tested positive for cervical cancer [8].

B. Methods

One of the most widely used data mining methods, which is used to classify large data sets and reveal important data classes, or to predict future data trends, is classification models [9].

Classification is used to reveal hidden patterns in databases. It is used to estimate the class of the data whose class has not been determined by using the existing classed data, or to determine whether the previously classified data is classified correctly and if there is a misclassification, it is used to assign the data to the correct group [4].

In this web-based software, there are algorithms such as Logistic regression, Decision trees, Support vector Machines, Random Forest, LightGBM, Gaussian Naive Bayes, AdaBoost and XGBoost, which are classification methods. The classification algorithms included in the software are described below. In addition, an application was made on the cervical cancer dataset in order to evaluate the outputs of the software.

B.1. Logistic Regression Analysis

The main purpose of logistic regression analysis is to model in order to define the relationship between the dependent variable and the independent variable without being subject to a certain precondition, when the dependent variable can be categorical and the independent variables can be both categorical and continuous. In other words, it is an analysis method that investigates the cause and effect relationship between the dependent variable and the independent variables. The relationship between the variables need not be linear. It can also be an exponential or binomial distribution relationship [10].

In the medical applications of logistic regression models, independent variables are risk variables or variables that determine whether a disease will occur or not. In short, logistic regression is a regression method that helps to assign

and classify the expected value of the dependent variable according to the independent variables [11].

B.2. Decision Trees

Decision trees in data mining are the most widely used technique among classification models because they are cheap to set up, easy to interpret, easily integrated with database systems, and have good reliability. Decision tree, as the name suggests, is a predictive technique in a tree view. It is the most popular classification technique that can create easy-to-understand rules and integrates easily with information technology processes with its tree structure [12].

Decision trees create tree-based classification models. They classify records into groups or make an estimation of the target (dependent) variable value, which depends on the values of the independent variables [13].

To create a classification tree, there is a feature that best determines the examples in the learning set. With this feature, the so-called branch and leaves of the tree are separated and a new sample set is created. A new defining attribute is found from the instances on this parsed branch and new branches are created. If all instances in each sub dataset, that is, on the branch, belong to the same class, there are no other attributes to parse the instances, and there are no other instances with the value in the remaining attributes, the branching process ends. Otherwise, there is a respecifying attribute to parse the sub dataset [14].

B.3. Support Vector Machines

Support vector Machines is a machine learning model developed by Vapnik-Chervonenkis, used in clustering and regression problems, especially in classification [15].

The SVM method has been used frequently in recent years, especially in data mining, for classification problems in data sets where the patterns between variables are unknown. This method was originally thought of as a linear classifier for solving two-class problems, then generalized to the solution of nonlinearly separable or multi-class classification problems, and started to be widely used in solving these problems [16].

Support vector machine models is an algorithm that has become popular recently. The main purpose of the SVM model is to determine the hyperplane that will best separate the classes of the target variable from each other. In other words, it is to maximize the distance between support vectors belonging to different classes. Support vector machines use an iterative training algorithm used to minimize the error function to create an optimal hyperplane [15].

B.4. Random Forest

The random forest (RF) classifier is made up of a number of tree classifiers, each of which is constructed using a random vector sampled separately from the input vector, and each tree casts a unit vote for the most popular class in order to classify an input vector. A decision tree's design necessitated the selection of a feature selection measure as well as a pruning procedure. There are several techniques for choosing features for decision tree induction, and most of them assign a quality measure to the feature directly. The Information Gain Ratio criterion and the Gini Index are the most commonly utilized feature selection measures in decision tree induction. The Gini

Index is used by the random forest classifier as a feature selection measure, which measures the impurity of a feature in relation to the classes [17].

B.5. Gaussian Naive Bayes

Gaussian naive Bayes (GNB) classification is a supervised learning approach that employs Bayes' theorem as a framework for categorizing observations into one of a pre-defined set of classes based on predictor variables' information. GNB classifiers estimate the conditional probability that an observation belongs to a certain class based on the values of the predictor variables, assuming that the predictor variables are class-conditionally independent, and hence (naively) ignore predictor variable covariance. GNB classifiers beat other, more sophisticated classifiers in classification tasks, even when assumptions aren't met [18].

B.6. XGBoost

Using gradient-boosted decision trees, XGBoost was primarily developed for speed and performance. It represents a method for machine boosting, or applying boosting to machines, pioneered by Tianqi Chen and adopted by a large number of developers. It's a part of the Distributed Machine Learning Community's toolkit (DMLC). For tree boosting methods, XGBoost (eXtreme Gradient Boosting) aids in maximizing memory and hardware resources. It has the advantages of improving the algorithm and modifying the model, and it can also be used in computing settings. Gradient Boosting, Regularized Boosting, and Stochastic Boosting are the three major gradient boosting techniques that XGBoost can perform. It also distinguishes itself from other libraries by allowing the addition and adjustment of regularization parameters. The approach is very efficient in lowering computation time and making the best use of memory resources [19].

B.7. LightGBM

Microsoft's LightGBM is a free and open source Gradient boosting algorithm. The parallel voting decision tree approach, which uses the histogram-based method to speed up the training process, minimize memory consumption, and combine advanced network connectivity to maximize parallel learning, employs the histogram-based method. In each cycle, divide the training data into different machines and make a local voting choice to select the top-k attributes and a global voting choice to receive the top-2k attributes. To locate the leaf with the highest splitter gain, LightGBM employs a leaf-by-leaf approach [20].

B.8. AdaBoost

The AdaBoost algorithm creates strong classifiers from weak ones. The AdaBoost algorithm's weak classifiers are members of the ensemble classifier. By adaptively modifying the weights in each cycle, AdaBoost develops a committee of member weak classifiers. The weights of the training samples that a current weak classifier misclassified are increased, whereas the weights of the training samples that a current weak classifier successfully classified are dropped [21].

C. Model Validation and Performance Evaluation

For model validation, the data set was splitted into training (75%) and testing (25%) datasets. In the evaluation of classification results for all classification methods available in the software; Performance criteria of accuracy, sensitivity, specificity, negative predictive value, false positive rate, false negative rate, matthews correlation coefficient, positive likelihood ratio, negative likelihood ratio and F1 score are given. Obtained from a random forest classification model; accuracy is 94.44%, sensitivity 100%, specificity 93.33%, and F1 score value of 94.44%. The random forest model created according to the relevant performance criteria successfully classifies cervical cancer.

III. RESULTS

D. Data Classification Software

The data classification software user interface was created using Python Dash and Html codes. The data set file can be loaded from the data loading menu, and the predictive and predicted (class variable) variables can be selected in the data selection menu. The next menu is the training menu, which includes comprehensive classification algorithms. The main menu of the software is as in Figure I.

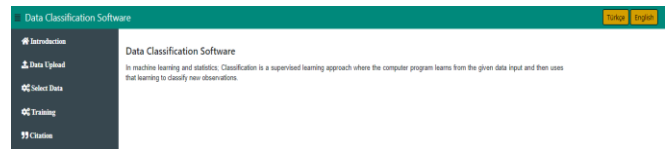


Figure I. The main menu of the software

The values of the performance criteria for the generated Random forest model are shown in Figure II. Obtained from a random forest classification model; Accuracy is 94.44%, Sensitivity 100%, Specificity 93.33%, and F1 score value of 94.44%. The random forest model created according to the relevant performance criteria successfully classifies cervical cancer.

Metric	Value
Accuracy	0.9444
Precision	0.75
Sensitivity	1
Negative Predictive Value	1
MCC	0.8367
False Positive Rate	0.0667
False Negative Rate	0
Specificity	0.9333
Positive Likelihood Ratio	14.9025
Negative Likelihood Ratio	0
F1 Score	0.9444

Figure II. Results for Performance Metrics for Random Forest Model Generated with Data Classification Software

IV. CONCLUSIONS

Classification analysis is one of the basic machine learning methods and is used by a large scientific community. There are many analysis tools used to guide researchers in this type of analysis. There are both commercial and free software tools that users can analyze their data with classification methods. In general, easy-to-use and well-designed interfaces are offered by commercial software packages [22].

Generally, well and comprehensively designed interfaces are offered by commercial software packages. One of these packages, Stata, is command-based and dependent on its commercial environment, the operating system. One of the free software packages that provides advanced possibilities for data classification analysis is the R environment. However, the R environment, like Stata, is OS dependent and command based. There are also free, open-source tools such as Weka that provide advanced analysis techniques. However, the Weka environment is an analytics tool that works as a desktop application. The fact that Weka is a desktop application can be difficult and time-consuming, especially for users (physicians, etc.) who do not have the Weka program installed on their computer. In addition, performing analysis with the Weka interface can be complex for most physicians [23-25].

Stata, R environment, Python, Rapidminer and WEKA programs can be both time consuming and difficult for researchers when data analysis needs to be evaluated quickly as they have to install these programs on their computers. Also, for most researchers, performing their analysis with such software can become more complex. On the other hand, the web-based software developed by this study is free, user-friendly and can perform data classification analysis from any device with internet access without writing any code, and provides comprehensive performance criteria and outputs of the classification results.

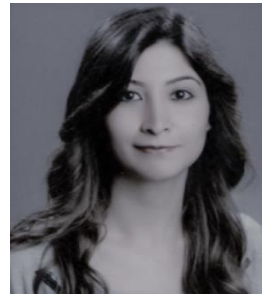
The software developed in this study, on the other hand, offers researchers a new user-friendly web-based software where they can easily perform data classification analysis and the analysis results can be easily understood. As a result, it is expected that this web-based software will allow them to compare comprehensive classification methods in disease prediction, particularly when compared to other analysis tools for physicians and healthcare professionals.

REFERENCES

- [1] S. Özekes, "Veri Madenciliği Modelleri ve Uygulama Alanları," 2003.
- [2] S. Y. B. Dalı, "Veri Madenciliği ve Müşteri İlişkileri Yönetiminde (Crm) Bir Uygulama."
- [3] N. Zhong and L. Zhou, Methodologies For Knowledge Discovery And Data Mining: Third Pacific-Asia Conference, PAKDD'99, Beijing, China, April 26-28, 1999, Proceedings: Springer, 2003.
- [4] H. Akpınar, "Veri Tabanlarında Bilgi Keşfi ve Veri Madenciliği, İstanbul Üniversitesi, İşletme Fakültesi Dergisi, C," ed: XXIX, 2000.
- [5] G. Akgül, A. A. Çelik, Z. E. Aydın, and Z. K. Öztürk, "Hipotiroidi Hastalığı Teşhisinde Sınıflandırma Algoritmalarının Kullanımı," Bilişim Teknolojileri Dergisi, vol. 13, pp. 255-268, 2020.
- [6] B. Gülmez, "Yapay Sinir Ağlarının Yeni Metasezgisel Algoritmalar ile Eğitimi ve Veri Madenciliğinde Sınıflandırma Alanında Kullanımı," Erciyes Üniversitesi, Fen Bilimleri Enstitüsü Endüstri Mühendisliği Anabilim Dalı.
- [7] Y. E. Kuyucu, "Lojistik Regresyon Analizi (LRA), Yapay Sinir Ağları (YSA) ve Sınıflandırma ve Regresyon Ağaçları (C&RT) Yöntemlerinin

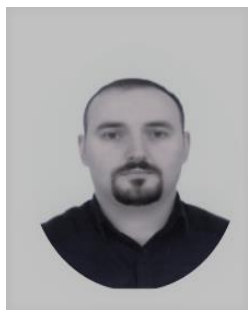
- Karşılaştırılması ve Tıp Alanında Bir Uygulama," Gaziosmanpaşa Üniversitesi, Sağlık Bilimleri Enstitüsü, 2012.
- [8] R. Machmud and A. Wijaya, "Behavior Determinant Based Cervical Cancer Early Detection With Machine Learning Algorithm," Advanced Science Letters, vol. 22, pp. 3120-3123, 2016.
- [9] F. Köktürk, H. Ankaralı, and V. Sümbüloğlu, "Veri Madenciliği Yöntemlerine Genel Bakış/Overview to Data Mining Methods," Türkiye Klinikleri Biyoistatistik, vol. 1, p. 20, 2009.
- [10] N. Bayram, "Multinomial Lojistik Regresyon Analizinin İstihdamdaki İşgücüne Uygulanması," İstanbul Üniversitesi İktisat Fakültesi Mecmuası, vol. 54, p. 61, 2004.
- [11] H. Bircan, "Lojistik Regresyon Analizi: Tıp Verileri Üzerine Bir Uygulama," Kocaeli Üniversitesi Sosyal Bilimler Dergisi, pp. 185-208, 2004.
- [12] G. Ulusoy, "Karar Ağacı Analizi ile AB Genişleme Kriterlerinin Değerlendirilmesi," 2013.
- [13] G. Silahtaroglu, "Veri Madenciliği," Papatya Yayınları, İstanbul, 2008.
- [14] E. Akçetin And U. Çelik, "İstenmeyen Elektronik Posta (Spam) Tespitinde Karar Ağacı Algoritmalarının Performans Kıyaslaması," İnternet Uygulamaları ve Yönetimi Dergisi, vol. 5, pp. 43-56, 2014.
- [15] V. Vapnik, The Nature Of Statistical Learning Theory: Springer science & business media, 2013.
- [16] Ö. Y. Akşehirli, H. Ankaralı, D. Aydın, and Ö. Saraçlı, "Tıbbi Tahminde Alternatif Bir Yaklaşım: Destek Vektör Makineleri," Türkiye Klinikleri Journal of Biostatistics, vol. 5, 2013.
- [17] M. Pal, "Random Forest Classifier For Remote Sensing Classification," International journal of remote sensing, vol. 26, pp. 217-222, 2005.
- [18] J. C. Griffis, J. B. Allendorfer, and J. P. Szaflarski, "Voxel-Based Gaussian Naïve Bayes Classification Of Ischemic Stroke Lesions In Individual T1-Weighted MRI Scans," Journal Of Neuroscience Methods, vol. 257, pp. 97-108, 2016.
- [19] S. D. Dhaliwal, A.-A. Nahid, and R. Abbas, "Effective intrusion detection system using XGBoost," Information, vol. 9, p. 149, 2018.
- [20] D. Wang, Y. Zhang, and Y. Zhao, "LightGBM: An Effective Mirna Classification Method In Breast Cancer Patients," in Proceedings of the 2017 International Conference on Computational Biology and Bioinformatics, 2017, pp. 7-11.
- [21] T.-K. An and M.-H. Kim, "A New Diverse Adaboost Classifier," in 2010 International conference on artificial intelligence and computational intelligence, 2010, pp. 359-363.
- [22] İ. Perçin, F. H. Yağın, A. K. Arslan, and C. ÇOLAK, "An Interactive Web Tool for Classification Problems Based on Machine Learning Algorithms Using Java Programming Language: Data Classification Software," in 2019 3rd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT), 2019, pp. 1-7.
- [23] T. C. Sharma and M. Jain, "WEKA Approach For Comparative Study of Classification Algorithm," International Journal of Advanced Research in Computer and Communication Engineering, vol. 2, pp. 1925-1931, 2013.
- [24] L. StataCorp, "Stata Data Analysis and Statistical Software," Special Edition Release, vol. 10, p. 733, 2007.
- [25] R. RStudio Team, "RStudio: Integrated Development for R," RStudio, Inc., Boston, MA URL <http://www.rstudio.com>, vol. 42, p. 14, 2015.

BIOGRAPHIES



İPEK BALIKÇI ÇİÇEK obtained her BSc. degree in mathematics from Çukurova University in 2010. She received MSc. degree in biostatistics and medical informatics from the Inonu University in 2018. She currently continues Ph.D. degrees in biostatistics and medical informatics from the Inonu University. In 2014, she joined the Department of Biostatistics

and Medical Informatics at Inonu University as a researcher assistant. Her research interests are cognitive systems, data mining, machine learning, deep learning.



İLHAMİ SEL obtained his BSc. degree in Computer Education from Firat University in 2007 and Computer Engineering from Inonu university in 2018. He received MSc. degree in Electronic and Computer Education from Firat University in 2013. He currently continues his Ph.D. studies in Computer Engineering at the Inonu University.

He has been working as a computer teacher in the Ministry of Education since 2007. His research interests are Natural Language Processing, data mining, machine learning, deep learning.



FATMA HİLAL YAĞIN obtained her BSc. degree in Statistics from Gazi University in 2017. She received MSc. degree in biostatistics and medical informatics from the Inonu University in 2020. She currently continues Ph.D. education in biostatistics and medical informatics from the Inonu University. In 2019, she joined the

Department of Biostatistics and Medical Informatics at Inonu University as a researcher assistant. Her research interests are cognitive systems, data mining, machine learning, deep learning, and image processing



CEMİL ÇOLAK obtained his BSc. degree in Statistics from Ondokuz Mayıs University in 1999. He received MSc. degree in Biostatistics from the Inonu University in 2001, and Ph.D. degree in the Graduate Department of Biostatistics and Medical Informatics of Ankara University in

2005. His research interests are cognitive systems, data mining, reliability, and biomedical system, genetics, and bioengineering. In 2016, he joined the Department of Biostatistics and Medical Informatics at Inonu University as a Professor, where he is presently a professor. He is active in teaching and research in the general image processing, artificial intelligence, data mining, and analysis.

Real-Time Monitoring, Analysis and Control of Power Parameters in Residential Houses Using LabVIEW

Behcet Kocaman* and Yunus Yiğit

Abstract— Electrical energy consumed in residential has an important value in total consumption. For this reason, monitoring and control of the electricity consumed in the residential offers a significant potential for energy saving. In this study, the power parameters of the linear and nonlinear loads used in the residential have measured, analyzed and controlled in real time. Thanks to real-time monitoring, analysis and control of power parameters ensure efficient use of energy. For this, power parameters of the loads used have transferred to the computer thanks to the data acquisition card. A visual interface has been created with the software designed in the Laboratory Virtual Instrument Engineering Workbench (LabVIEW) program so that the data transferred to the computer can be monitored and controlled instantly.

Index Terms— Control, energy monitoring, energy efficiency, LabVIEW, power system analysis.

I. INTRODUCTION

ELECTRIC ENERGY demand is increasing day by day. To reduce this increasing demand, it is necessary to use energy economically and efficiently. Nowadays, studies on energy efficiency and saving are ongoing. Various systems have been developed to deliver the energy produced in the electricity generation facilities to the end consumer in an efficient and quality manner. One of these systems is the energy monitoring system. Thanks to energy monitoring systems, energy efficiency can be increased in enterprises, transmission lines, production facilities, distribution facilities and buildings. For this, it is necessary to measure energy consumption, analyze the measured data and take measures for energy efficiency and quality according to the results of

this analysis [1]. In addition, energy monitoring system which is helps to monitor the stability and consumption of electrical energy is an essentiality for electricity powered buildings at the present time.


In energy monitoring systems, parameters such as current, voltage, power factor, frequency, apparent power, active power, reactive power are shown. Sometimes the current and voltage in the electrical grid can be high or low. In this case, negative results such as malfunction or distortion of electrical and electronics devices occur in businesses, residential, offices, industrial facilities. These problems create high cost and large operational problems for the user consuming electricity and all the elements on the power system [2]. To prevent these problems, real-time monitoring, analysis and control of power parameters should be reported and energy management should be planned according to the report results. Some work has been done on energy monitoring systems, analysis and control. With these studies, it has been aimed to make electrical energy better quality, more efficient, faster and lower cost.

A real-time grid energy consumption profile and an energy-sensitive grid/interface selection tool are proposed for android-based smartphones [3]. The proposed system makes use of actual package measurements and exact calculations. Thus, it reports the energy consumption levels of different grid interfaces such as Wi-Fi and cellular and facilitates the horizontal / vertical transition of devices to increase the energy efficiency. General analyzes have been made to show the accuracy of the tool proposed for this.

The presented analysis of the attitudinal and socio-economic determinants of the willingness of residential consumers to monitor their energy consumption. In their study, they were able to demonstrate how consumers' knowledge and experience of monitoring environmental attitudes, energy use and demographics affect their desire to monitor energy consumption, both in the electricity bill and smart metering information platforms [4].


The suggested that energy monitoring systems reduce energy consumption. In particular, real-time feedback to consumers has been noted to cause a significant behavioral change in energy consumption. The energy monitoring system is crucial for emerging businesses and changing industries to effectively achieve their aims. Integrating cloud / Internet of Thing (IoT) solutions can result in remarkable energy savings and improvements [5].

BEHÇET KOCAMAN, is with Department of Electrical-Electronics Engineering University of Bitlis Eren University, Bitlis, Turkey, (e-mail: bkocaman@beu.edu.tr).

 <https://orcid.org/0000-0002-1432-0959>

*Corresponding Author.

YUNUS YIĞİT, is with Graduated from Science Institute, Department of Electrical-Electronics Engineering University of Bitlis Eren University, Bitlis, Turkey, (e-mail: yunis4963@gmail.com).

 <https://orcid.org/0000-0002-1991-1635>

Manuscript received July 13, 2021; accepted October 26, 2021.

DOI: [10.17694/bajece.970685](https://doi.org/10.17694/bajece.970685)

The developed an approach to identify home appliances and classify household activities in daily life using home smart meter data. It also introduced a deep learning approach based on multi-layered, feed-forward neural grids that can identify common household electrical appliances from a typical home smart meter metering. The resulting daily life classifying activities can give an idea about the attitude of the family, which has a number of applications in the field of energy and other fields [6].

The proposed the management problem of dissimilar types of Egyptian residential loads, including daily operating performance and time shifting operation capacity, has been formulated and a new smart load management method. The proposed method represents the programming section required for load management control centers and accepts the preferred working times of consumers as input data approving maximum customer satisfaction [7].

Several study proposed real-time home energy monitoring and automation system using the Arduino microcontroller board and the LabVIEW program. The proposed study was designed to be capable of measuring different electricity parameters, determining parameters of power quality, simulating and analysing signals, recording parameters and detecting faults[8-12].

The presented a real time energy monitoring system that is cost-effective and reliable. The system can be used to analyze and evaluate the household appliance's output voltage or generated energy. A hardware device was used to collect energy data that is transferred and stored to a database via cloud-based RESTful API resources. The data obtained was then used by the mobile web application to view real-time and historical energy readings[13].

A low-cost energy monitoring system has been proposed using ZigBee wireless communication with the Arduino controller to monitor real-time electrical parameters. The ZigBee and Arduino interface technique has been used to reduce more coding in the system and to take energy meter reading with less delay time. The proposed system has been developed to save energy and to raise awareness of daily power usage for residential, commercial and industrial consumers [14,15].

An experimental comparison among IoTs devices with energy gathering features using wireless technology such as ZigBee, WiFi 2.4 GHz and Long Range Wireless Area Network (LoRaWAN) in order to monitoring of agricultural proposed[16].

Application of an energy monitoring device using Arduino Nano microcontroller to measure the energy consumed by any electrical appliance presented in residential homes. Besides, designed with the ability to send measured electrical power parameters to owners via SMS when requested. Proposed device is useful for homeowners who want to monitor the energy consumption of some appliances to reduce their electricity bill. Besides, It can also be used for residential energy control [17].

A smart building real-time energy monitoring system

presented. The proposed system aimed at improved safety and user awareness of energy consumption at low cost. The energy consumed can be measured instantly. In case of increased power consumption, the system can be programmed and start to turn off some unimportant appliances. With the system made using a low-cost general purpose microcontroller (Arduino), the power consumption and cost are reduced [18].

In the system where the intelligent voltage and current monitoring system technique is proposed, an arduino platform is used as a microcontroller to read the voltage and current from the sensors. With the system, the measured data is sent wirelessly in monitoring a three-phase electrical system. A new android application is used for this [19].

In literature reviews on the subject, methods that help monitor and use energy efficiently such as using android smartphones [3], using arduino microcontroller [10], DFRduino Uno [13], Using Zigbee and Arduino [15], using internet of things devices [16] are presented. In this study, real time power parameters and energy consumption are monitored. In addition, the relay and the software prepared in the LabVIEW program were used to prevent the vehicles connected to the single phase electricity grid from being damaged by excessive current / voltage. In the event that the adjustable current / voltage lower and upper limit values are exceeded with the designed system, the electricity of the relevant device is cut and controlled. Thus, damage to household appliances is prevented. In addition to monitoring the novelty of the study, power parameters and energy, the control of electrical home appliances was provided. The system we proposed; the tablet performs real-time measurement, monitoring and control with mobile applications.

The rest of the paper has structured as follows. In Section 2, material and method have explained. In other words, hardware and software features of the device prototype have described. Results and discussion have investigated in Section 3. Conclusion has evaluation in Section 4.

II. MATERIAL AND METHOD

The proposed system consists of experimental and software. Experimental mainly includes analog/digital input and digital output circuits, myDAQ data acquisition card, test cables and computer. MyDAQ produced by National Instruments has been used to collect and process the data. SCT-13 30A is used as current sensor. This sensor offers easy use due to its non-invasive nature. 220V-9 V, AC-AC adapter is used as voltage sensor. A voltage divider circuit is used to adapt the output signal of the voltage sensor to the data acquisition board. A switch is installed in the system to control current and voltage manually or automatically. In addition, a 5V controlled 220 V relay has installed to protect the system. Fault warning system has realized with visual and audible warning. The block scheme of the experimental setup is shown in Fig. 1.

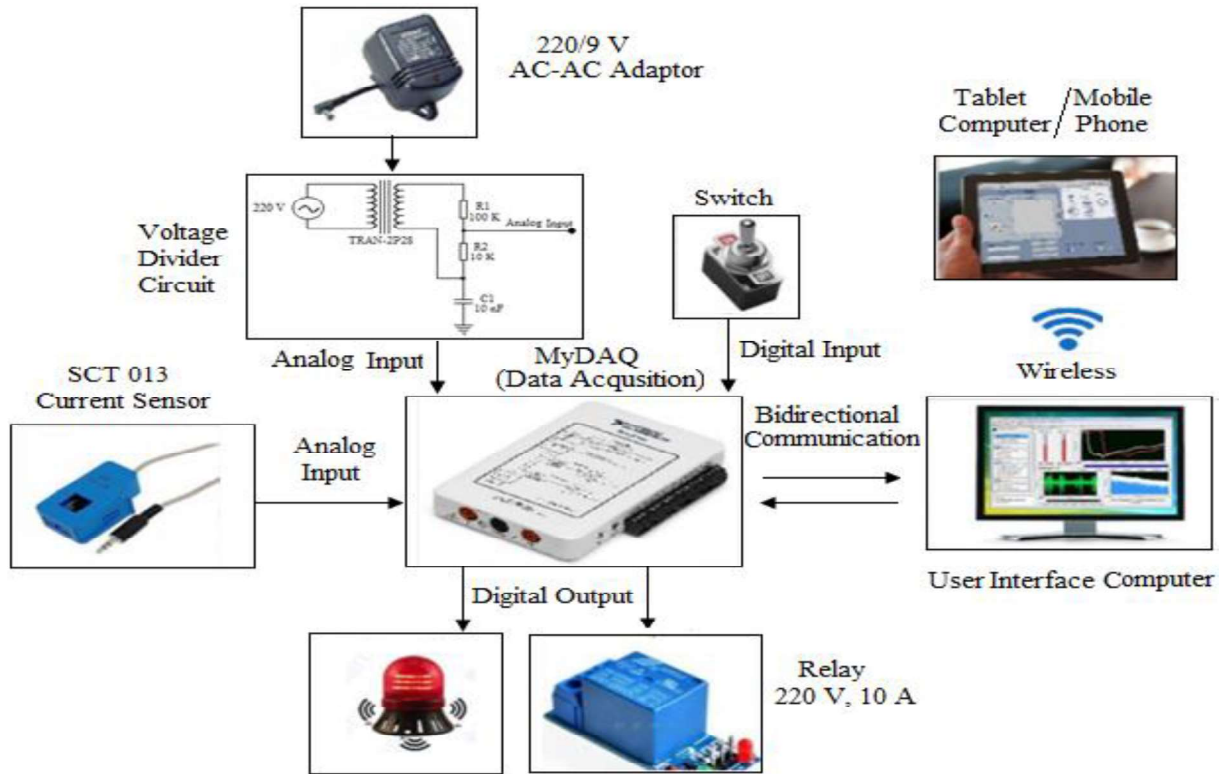


Fig. 1. Block scheme of the experimental setup.

Software is based on LabVIEW platform in computer. LabVIEW is a system-design platform and development environment for a visual programming language from National Instruments. The total cost of all components used in the proposed system was approximately \$ 281,35. However, devices that are used only for measuring and monitoring purposes (energy analyzer) without control feature are sold at prices of approximately \$ 1200 and above [20,21]. Data processing in LabVIEW is used to calculate and analyze the collected data. Photo of the experimental setup is shown in Fig. 2. System structure block diagram is shown in Fig. 3.

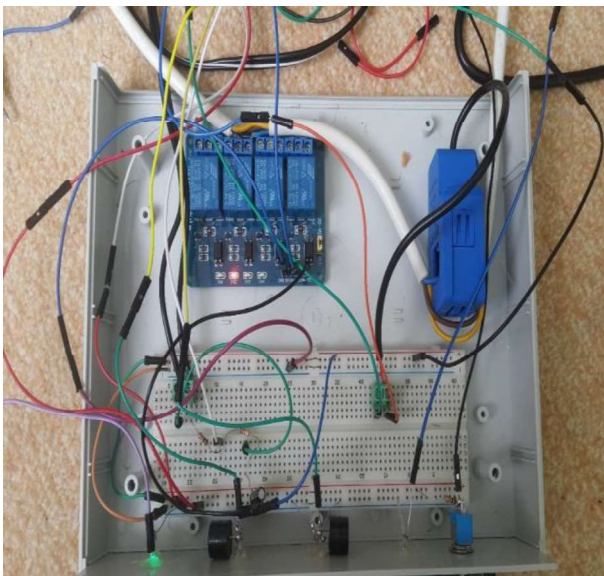


Fig. 2. Photo of the experimental setup.

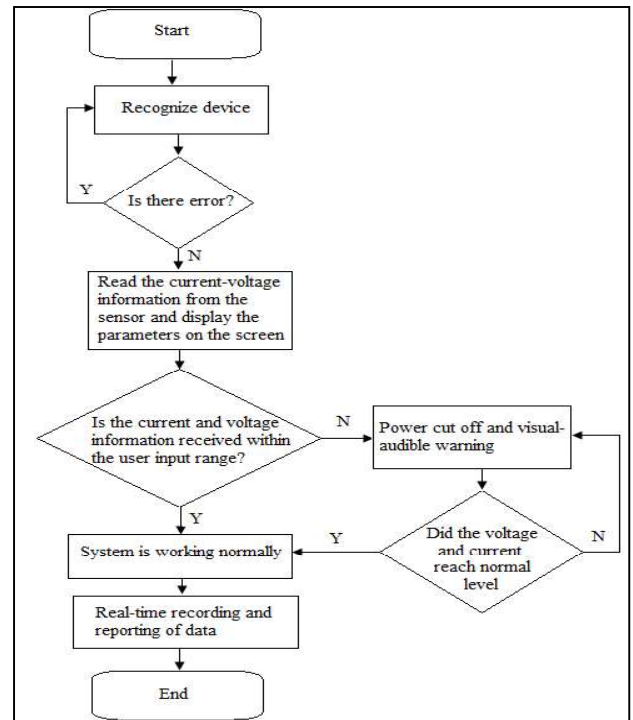


Fig. 3. System structure block diagram.

With connected linear and nonlinear loads to the system, the power parameters of the load have monitored in real time on a computer and tablet, has analyzed and controlled. The analog signal output from current and voltage sensors is integrated into the inputs of the data acquisition card. In addition, in order to monitor the data and system status, necessary designs have been made in

the LabVIEW program and a visual interface has been created. The interface consists of two parts. In the first part, the waveform of current and voltage signals is shown on the graph. In the second part, the settings of the power parameters have been displayed numerically, system

status, contact information, fault information and the recording time required for reporting parameters. The first part of the visual interface is shown in Fig. 4 and the second part is shown in Fig. 5.

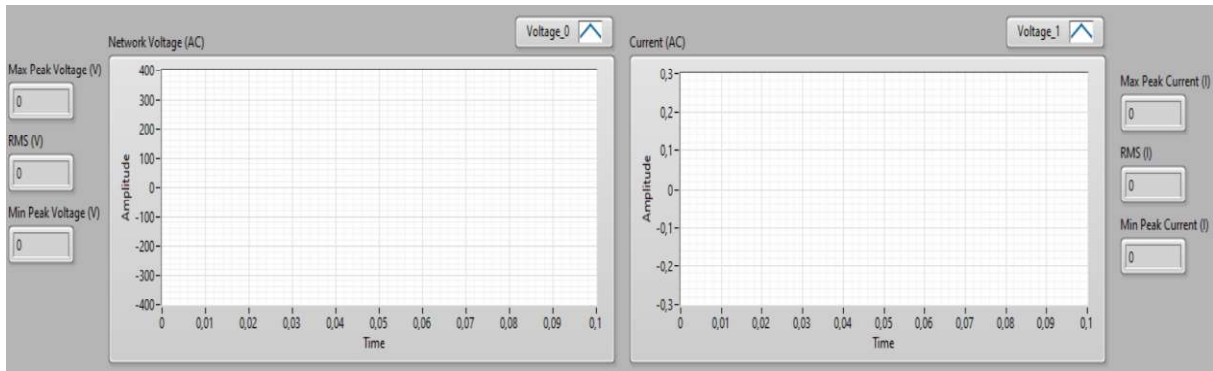


Fig. 4. The first part of the visual interface

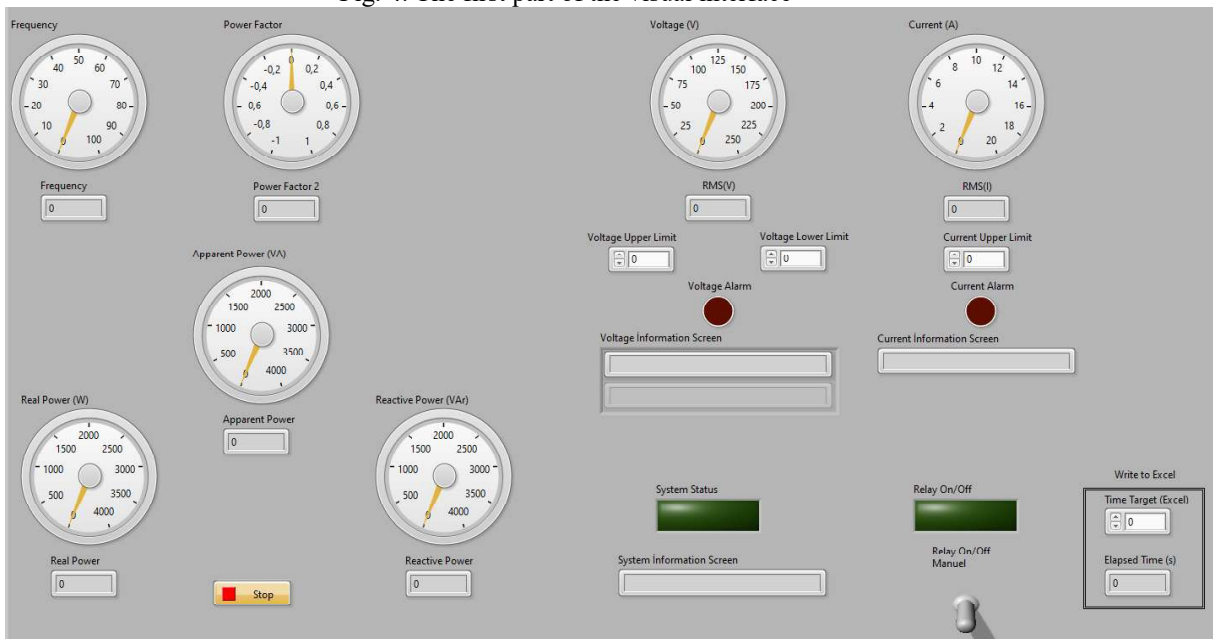


Fig. 5. The second part of the visual interface

The block diagrams used in the LabVIEW program for operating the designed system are given in Figure 6-10. The block diagram of sending alarm information to myDAQ is given in Fig. 6. The block diagram of analogue reading of the voltage signal is given in Fig. 7. An analog reading of the current signal block diagram is given in Fig. 8.

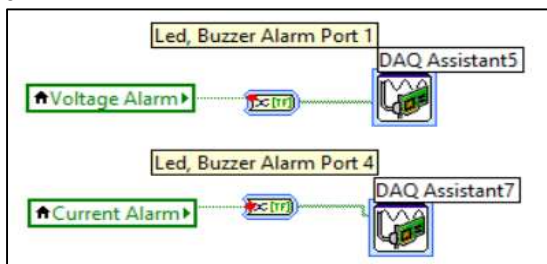


Fig. 6. Sending alarm information to myDAQ block diagram

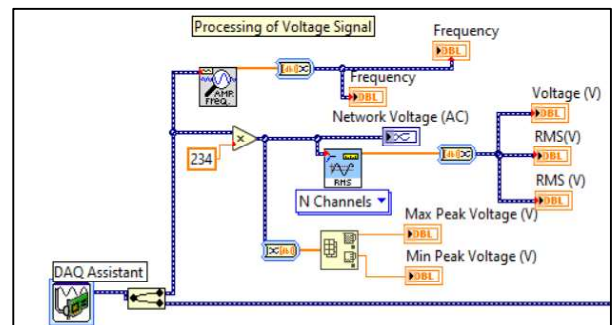


Fig. 7. Analogue reading of the voltage signal block diagram

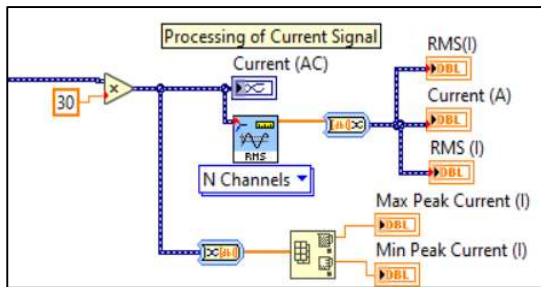


Fig. 8. Analogue reading of the current signal block diagram

Harmonic and THD analysis of the analog voltage signal received from the grid has been performed. The voltage harmonic measurement block diagram used for this is given in Fig. 9 and the current harmonic measurement block diagram in Fig. 10.

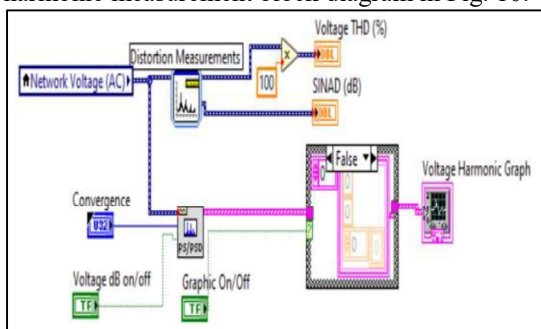


Fig. 9. Voltage harmonic measurement block diagram

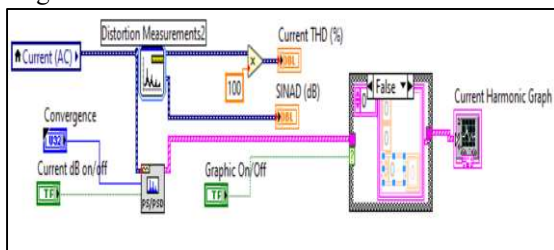


Fig. 10. Current harmonic measurement block diagram

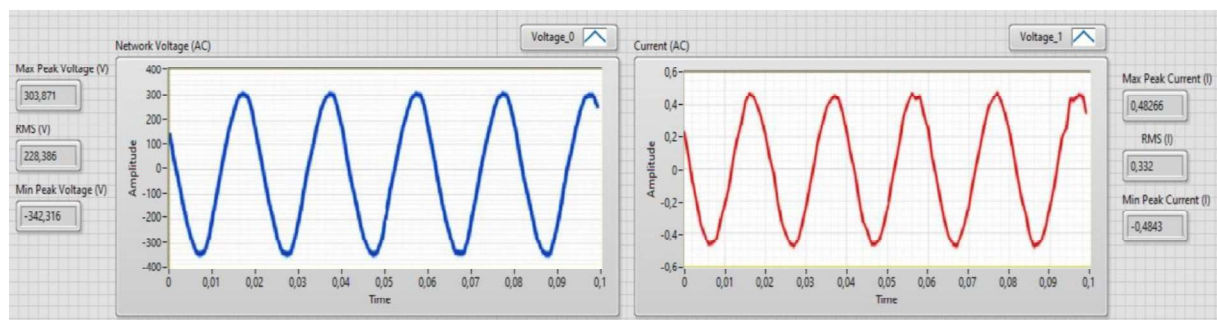


Fig. 11. 75W linear load incandescent lamp current and voltage waveforms

In case of voltage increase/drop and current increase, protection of the loads fed from single phase mains according to the limit values to be determined by the user can be done automatically or manually in this panel. Thus, electronic devices are prevented from being damaged. In Transformer Substations, relays are set to a maximum of 1-1.2 seconds [22], the opening (operation) time of the

The power parameter values taken from the data acquisition card have monitored graphically and numerically via the LabVIEW program. The data obtained are recorded by the designed system. Power parameters can be monitored within the specified time interval when wanted. In addition, according to the measured data, an automatic or manual control switch is used in order to prevent failures that may occur in the loads due to voltage drop/rise or current rise. In this way, the times when the more energy consumption of residential can be detected and electrical device failures can be prevented.

III. RESULT AND DISCUSSION

The designed system has been tested on linear and nonlinear loads. The received analog signals have been measured at a rate of 10.000 samples/second/channel and communicated with the computer. Measured analog signals were calibrated in LabVIEW program and power parameter values were monitored in computer environment. Signals created using linear and nonlinear loads have been monitored and analyzed. The proposed system has been tested by using linear and nonlinear loads separately and together. Current and voltage waveforms formed by connecting 75 W incandescent lamp to the system are given in Fig. 11. Also, the maximum and minimum peaks and root mean square (RMS) values of the signals have measured. Since the load is linear, the current waveform is close to the sine wave.

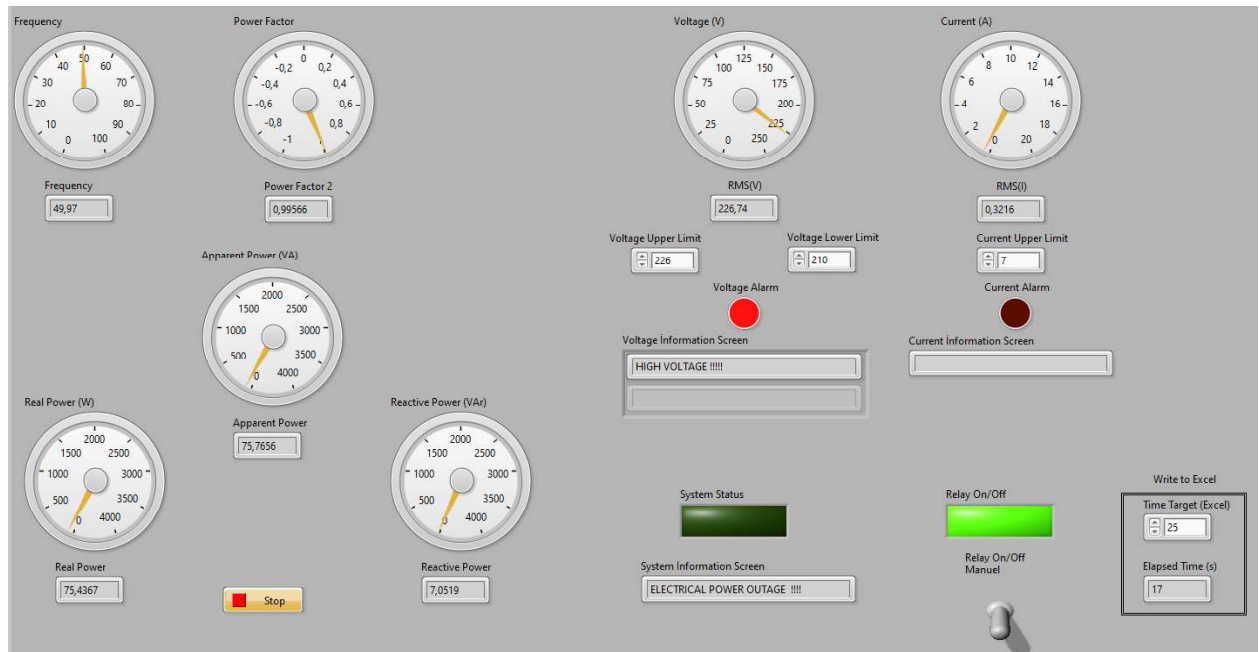


Fig. 12. System information screen and protection of devices due to voltage increase

The instantly received power parameters can record in report file in real time according to a time determined by

the user. A real time report of 75 W incandescent lamp is given in Table 1.

TABLE I
REAL TIME REPORT OF 75 W INCANDESCENT LAMP

Time	Voltage (V)	Current (A)	Apparent Power (VA)	Real Power (W)	Reactive Power (VAr)	Power Factor	Frequency
05.10.2019 12:06:48,80	226,98353	0,33395	75,80167	75,49347	6,82859	0,99593	49,94426
05.10.2019 12:07:13,90	226,98421	0,33150	75,24664	75,01415	5,91057	0,99691	49,95765
05.10.2019 12:07:38,99	226,83153	0,33016	74,89102	74,63330	6,21554	0,99655	49,97370
05.10.2019 12:07:45,20	226,83325	0,33015	74,88899	74,63330	6,18834	0,99658	49,97210

The data received instantly in Table 1 has been set to 60 seconds as the recording time to be determined by the user.

Data is written to the report file every 60 seconds. With this writing process, old data cannot be deleted, new data can be overwritten with old data and historical analyzes can be checked. 23 W energy saving lamp is used as a

nonlinear load in the system. Current and voltage waveforms drawn by nonlinear load are given in Fig. 13. The data are taken instantly and in real time. The current signal does not resemble the full sine signal due to the nonlinear load used. The maximum, minimum and RMS values of the signals are shown numerically.

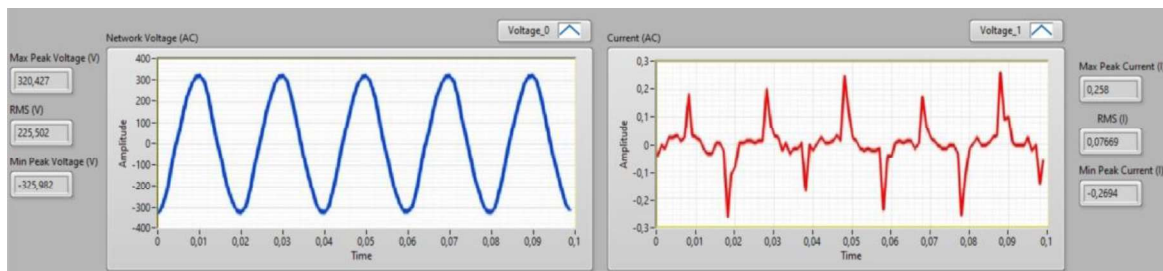


Fig. 13. 23 W nonlinear load energy saving lamp current and voltage waveforms

In case of rising / falling voltage and rising current, protection of the loads fed from single phase mains according to the limit values to be determined by the user can be performed automatically or manually in this panel.

In addition, in order to inform the user, failure conditions can be seen from the voltage and current information screen of the system. Instant acquisition of power parameters is shown in Fig. 14.

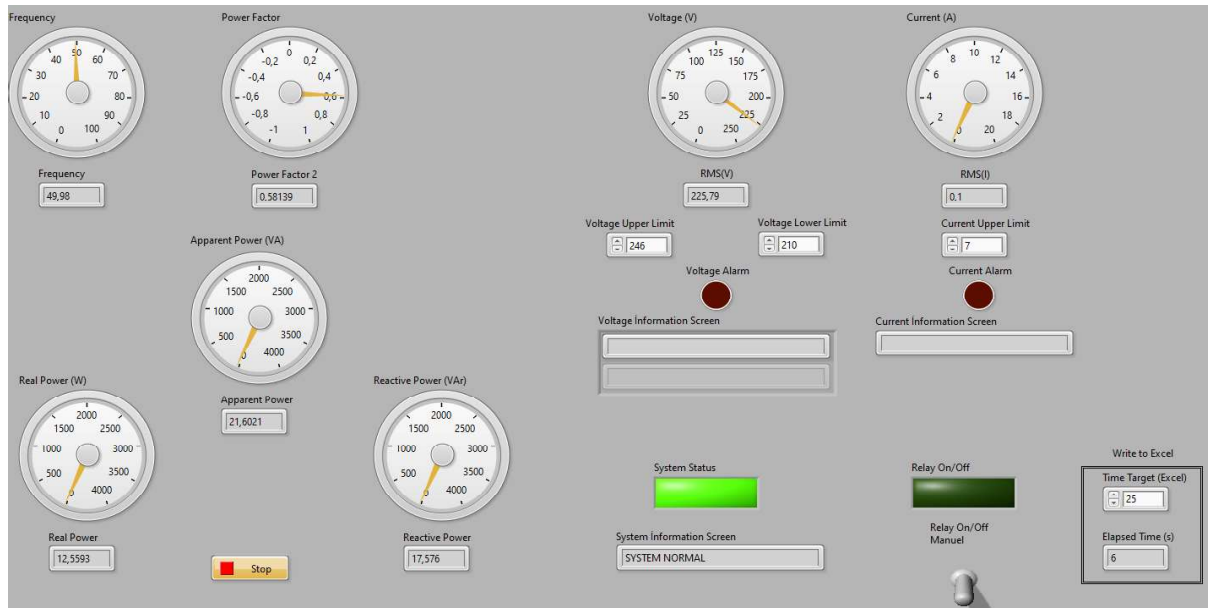


Fig. 14. Nonlinear load power parameters, system status, automatic-manual control and reporting

The instantly received power parameters can record in report file in real time according to a time determined by the user. The power factor was obtained by taking the cosine of the phase difference angle of the current and voltage signals transferred to the graph with 10.000 samples per second with the adjustment made in the

myDAQ card in the LabVIEW program. Apparent power is the total amount of power drawn by the load connected to the system. The product of the apparent power by the $\cos \phi$ gives the real power in theory. A real time report of 23 W energy saving lamp is given in Table 2.

TABLE II
REAL TIME REPORT OF 23 W ENERGY SAVING LAMP

Time	Voltage (V)	Current (A)	Apparent Power (VA)	Real Power (W)	Reactive Power (VAr)	Power Factor	Frequency
05.10.2019 12:12:15,55	225,87835	0,06857	15,48848	9,11141	12,52498	0,58827	49,96669
05.10.2019 12:12:40,65	225,65836	0,10112	22,81857	12,71314	18,94896	0,55714	49,96709
05.10.2019 12:13:05,74	225,54856	0,09032	20,37155	11,84299	16,57538	0,58135	49,96951
05.10.2019 12:13:30,84	225,33637	0,07586	17,09402	10,09094	13,79776	0,59032	49,97946

The data received instantly in Table 2 is set to 25 seconds as the recording time to be determined by the user. Data is written to the report file every 25 seconds. With this writing process, old data cannot be deleted, new data can be overwritten with old data and historical analyses can be checked. The power factor has been found to be around 0.58 by using a nonlinear load. Linear and nonlinear loads are used together in the residential. Linear

loads such as incandescent lamp iron and hair dryer and nonlinear loads such as energy saving lamp and laptop have been used in order to test the proposed system. Current and voltage waveforms of using linear and nonlinear loads are given in Fig.15. As shown in Fig.15 current waveforms is include harmonic. This is due to the nonlinear loads in the system. The real time report of linear and nonlinear loads is given in Table 3.

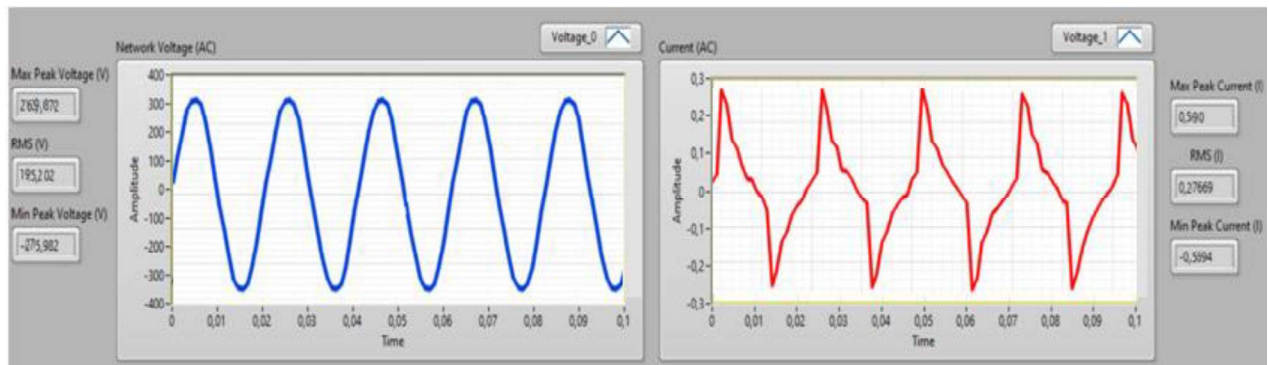


Fig. 15. Voltage and current waveforms of using together linear and nonlinear loads

TABLE III
REAL TIME REPORT OF USING TOGETHER LINEAR AND NONLINEAR LOADS

Time	Voltage (V)	Current (A)	Apparent Power (VA)	Real Power (W)	Reactive Power (VAr)	Power Factor	Frequency
18.12.2019 18:22:15,55	194,53046	0,25016	48,66374	41,27269	25,75886	0,84800	49,98189
18.12.2019 18:22:40,65	194,05919	0,23494	45,59227	38,63397	24,20889	0,84738	49,97802
18.12.2019 18:23:05,74	194,41001	0,25256	49,10019	41,50341	26,23539	0,84528	49,98101
18.12.2019 18:23:30,84	194,53049	0,23315	45,35478	39,20876	22,79758	0,86449	49,97527

In the proposed system, voltage, current, apparent power, real power, reactive power, power factor and frequency parameters are given in the report. However, if desired, parameters such as apparent energy, active energy, reactive energy, current THD and voltage THD can also be displayed. In addition, the consumed energy can be sent to the consumer as a message. According to this consumption data, the customer can use energy more efficiently by planning energy usage. The system is in a structure that can be developed in line with user requests.

IV. CONCLUSION

In this study, a real-time system is designed by processing the current and voltage values into the LabVIEW program with myDAQ data acquisition card. In this way, the power parameters of the linear and nonlinear loads used in the residential houses have measured, analyzed and controlled in real time. Real-time monitoring, analysis and control of the consumed energy is important for the efficient use of energy. Thanks to the measured and analyzed power parameters, the consumer energy will be used more carefully. Thus, the energy consumed in the residential will be saved. In addition, energy control can be provided within 4-10 ms to prevent damage to electrical household appliances due to malfunctions in the grid.

In order to the current and voltage information received by the system to be more than the nominal values of the devices, a protection circuit has been made to prevent damage to the devices connected to the system. The protection circuit can be realized both with the interface designed on the computer and with the help of a switch from the outside. In addition,

visual and audio alarm circuits have been installed on the computer and test circuit to inform the user of any failure. As a result of the tests, the system worked quickly and accurately. It is possible to make a good energy planning for future times by recording the data in real time and analyzing the historical data. The proposed system can be preferred in residential houses due to the features mentioned below.

It can provide both wired and wireless communication,

The relay used in the system is activated within 4 -10 ms in case of failure. So it is fast in case of failure,

It can be designed according to the user's request,

It can be easily transported and integrated into electrical energy systems more easily,

Performing both energy analysis and protection together,

It is cheaper than energy analyzers that can only measure and monitor without a control mechanism. So it is low cost,

It can be controlled automatically or manually: In case of an error failure in the grid, the energy can be controlled automatically with the LabVIEW program and the used relay. In addition, the energy can be manually controlled with the switch in the system.

Data storage by taking reports and according to these data better energy planning is made; more efficient energy use can be achieved by sharing energy consumption data with the customer.

By using energy monitoring, analysis and control systems, electrical energy is ensured to be efficient, high quality, low cost, uninterrupted and sustainable. Energy management policies can be directed by monitoring the data instantly and recording and reporting power parameters to the database. Energy analysis applications designed according to the user's desire in energy consumption can be tailored according to the

need. The proposed system can be used not only to monitoring energy, but also to control the energy received or delivered from the system.

REFERENCES

- [1] R. Govindarajan, S. Meikandasivam, and D. Vijayakumar, "Performance Analysis of Smart Energy Monitoring Systems in Real-time", *Engineering, Technology & Applied Science Research*, vol. 10.3, 2020, pp. 5808–5813.
- [2] C.T.Yang, S.T. Chen, J.C. Liu, et al. "On construction of an energy monitoring service using big data technology for the smart campus". *Cluster Computing*, vol. 23, 2020, pp.265–288.
- [3] M. F. Tuysuz, M. Ucan, R. Trestian. "A real-time power monitoring and energy-efficient network/interface selection tool for android smartphones". *Journal of Network and Computer Application*, vol.127, 2019, pp.107-121.
- [4] P. A. Kowalska, K. Byrka. "Determinants of the Willingness to Energy Monitoring by Residential Consumers: A Case Study in the City of Wroclaw in Poland". *Energies*, vol.12.5, 2019,907.
- [5] F. Al-Turjman, C. Altrjman, S. Din, A. Paul. Energy monitoring in IoT-based ad hoc networks: An overview. *Computers & Electrical Engineering*, 2019, vol.76, pp.133–142.
- [6] M.A. Devlin, B.P. Hayes. "Non-Intrusive Load Monitoring and Classification of Activities of Daily Living using Residential Smart Meter Data". *IEEE Transactions on Consumer Electronics*, vol 65.3, 2019, pp.339-348.
- [7] A.M. Elsayed, M.M.Hegab, S.M. Farrag. "Smart residential load management technique for distribution systems' performance enhancement and consumers' satisfaction achievement". *International Transactions on Electrical Energy Systems*, vol. 29.4,2019, pp.1-23.
- [8] P.K. Neelima, S.Suvarna, M.Roja, D.Madhusudhan. "Smart Energy and Overload Tracking Using Labview". *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering*, vol.7.6, 2018, pp. 2828-2834.
- [9] C. Nutthaka, L. Monthon, B. Suttipong, C. Chow. "Design and Implementation of a smart monitoring system of a modern renewable energy micro-grid system using a low-cost data acquisition system and LabVIEWTM program". *Journal of International Council on Electrical Engineering*, vol 7.1, 2017, pp. 142-152.
- [10] S.S.Mohamed, A.A.Ahmad, A.M. Abdulwadoud, O.E.Mohamed. "Design and Implementation of a Real-Time Smart Home Automation System Based on Arduino Microcontroller Kit and LabVIEW Platform". *International Journal of Applied Engineering Research*, vol.12.18,2017, pp. 7259-7264.
- [11] K.G.Samit, P.B.Natarajan, B.P.Sankata. "Intelligent Smart Home Automation System based on LabVIEW". *International Journal of Pure and Applied Mathematics*, Vol. 120. 6,2018, pp.339-348.
- [12] P. Ali, R.B. Hamid, E.I. Mohamad, B.G. Gevork, J.M. Guerrero. "Real-time simulator and ofine/online closed-loop test bed for power system modeling and development". *Electrical Power and Energy Systems*, vol.122.106203, 2020, pp.1-16.
- [13] K.M.E. Galera, O.E. Llantos. "Mobile Web Energy Monitoring System Using DFRduino Uno". *Procedia Computer Science*, vol. 124, 2017,706–713.
- [14] N. Hnin, Z.M.T. Thaug, M.T. Hla. "Automatic Energy Control And Monitoring System For Building". *International Journal of Scientific & Technology Research*, vol. 5.6, 2016, pp.125-129.
- [15] R. Govindarajan, S. Meikandasivam, D.Vijayakumar. "Energy Monitoring System Using Zigbee and Arduino". *International Journal of Engineering & Technology Sciences*, vol.7.4, 2018, pp.608-611.
- [16] S. Sadowski, P. Spachos. "Wireless technologies for smart agricultural monitoring using internet of things devices with energy harvesting capabilities". *Computers and Electronics in Agriculture*, vol.172.105338, 2020, pp.1-8.
- [17] O.S.Shereefdeen, O.O. Kehinde, M.Ismail. "Design and Implementation of Home Appliance Energy Monitoring Device". *International Journal of Electrical, Energy and Power System Engineering (IJEPPSE)*,vol. 2.2,2019, pp. 1-6.
- [18] S.K. Omar, K.A. Ahmed. "Real-Time Power Measurements In Smart Building Monitoring System", *Journal of University of Duhok (Pure and Eng. Sciences)*, vol. 20.1, 2017, pp.142-151.
- [19] M.M. Jabbar, A. Van den Bossche, C.R. Farhood. "A Smart Voltage and Current Monitoring System for Three Phase Inverters Using an Android Smartphone Application". *Sensors (Basel)*, vol.17.4, 2017, pp.2-16.
- [20] Fluke: Power Quality, (2020). [Online] Available: <https://www.fluke.com/en-us/product/electrical-testing/power-quality/434-435> [Accessed: 2020].
- [21] E-commerce: (2020) [Online] Available: https://www.alibaba.com/product-detail/High-quality-Handheld-Digital-Energy-Meter_62315836812.html?spm=a2700.7735675.normalList.97.3bd63971atn6Xv&s=p [Accessed: 2020].
- [22] R. Akkan, "Relay Coordination And Selection of Breaker Example of Ihsaniye". *Afyon Kocatepe University Graduate School of Natural and Applied Sciences, M.Sc. Thesis, Department of Electrical and Electronics Engineering*,2019.
- [23] Chint: Brief Katalog [Online] Available: http://www.chint.com.tr/images/content/files/ChintBriefKatalog_2014.pdf [Accessed: 2020].

BIOGRAPHIES



BEHÇET KOCAMAN was born in Bitlis, in 1972. He received the B.S. degree in Electrical Engineering from Yıldız Technical University, (Turkey), in 1993, the M.S. and PhD degrees in Electrical Engineering from Kocaeli University, (Turkey), in 1997 and 2015 respectively. Since 2015, he has been an Associate Professor with the Electrical and Electronics Engineering department, Bitlis Eren University, (Turkey). His areas of interest include energy efficiency, illumination, renewable energy sources, energy management and new transmission and distribution technologies.



YUNUS YİĞİT was born in Şanlıurfa, in 1993. He received the B.S. degree in Electrical and Electronics Engineering from Sakarya University (Turkey) in 2016. He graduated M.S. degree with the Electrical and Electronics Engineering department, Bitlis Eren University (Turkey) in 2019.

RSSI Based Indoor Localization with Reduced Feature Dimension

Mustafa Eren Yildirim

Abstract—Wifi based indoor localization gains the interest of researchers for several purposes. Among various techniques, fingerprinting based on Wifi received signal strength indicator (RSSI) is a widely used feature in indoor localization because of its simplicity in implementation and minimal hardware requirement conditions. However, the amount of access points (AP) at which the RSSI is measured from in the network increases the computational load. This paper presents an alternative approach for dimension reduction in RSSI based indoor localization. We focus on recognizing the building and floor of the test user which is a multi-class problem for both cases. In a multiple class problem, inter-class differences are obtained by Manhattan distance in pair-wise manner. From each pair calculation, top-25 and top-50 features with the largest variances are chosen and merged to generate the final feature set. The proposed algorithm is implemented and evaluated on UJIIndoorLoc dataset. According to the outcomes, our method provides 99.1% accuracy for building and 82.8% accuracy for floor estimation

Index Terms—Received signal strength indicator, dimension reduction, indoor localization, UJIIndoorLoc dataset.

I. INTRODUCTION

THE localization in indoor and outdoor environments is an essential but also challenging task for researchers for several years. The location information can be used for several purposes such as routing enhancement, object tracking, smart home and smart hospital applications. The localization in outdoor mediums is achieved by using Global Positioning System (GPS). However, employment of GPS is not efficient and successful for indoor localization due to several reasons. One of them is the high attenuation and scattering by roofs and walls. Other reason is that the error rate of GPS chips can be larger than indoor areas. In case of indoor localization, different technologies have been used such as Wifi, Bluetooth, radio-frequency identification (RFID) and visual features [1], [2], [3], [4]. Wifi is considered as the most popular and practical solution in indoors [5]. Indoor localization methods can be grouped as fingerprint-based methods which rely on received-signal-strength indicator (RSSI) [6] and range-based [7] methods. This paper proposes a RSSI based indoor localization method.

RSSI fingerprinting algorithm consists of offline and online phases. In offline phase, data samples are collected. These samples are the signal powers measured at certain reference points (RP) relative to all of the access points (AP) in the network. Thus, a database of RSSI values is generated. In

the online stage, the location of a test point is estimated by using the generated database. One of the major advantages of the WLAN fingerprint based methods is that they do not require the installation of any additional hardware since they use the existing WLAN infrastructure. Therefore, the location of the user can be obtained without additional infrastructures and costs. However, WLANs were not natively designed to support a positioning function. It becomes even more difficult to capture the spreaded radio signals when we take the existing obstacles introduced by the indoor environment including reflections and multipath interference into considerations [8].

The proposed method is tested on a publicly available dataset called UJIIndoorLoc [8]. There are several studies which have used this dataset. In [9], authors implemented a deep neural network (DNN) to decrease the computation load. Convolutional neural network (CNN) based approach for indoor localization using RSSI time-series is presented in [10]. The study [11] estimates the node building, floor and location coordinates using a single DNN. Authors in [12] proposed a confidence measure to reflect the uncertainty of the positioning prediction.

In this paper, we present a scheme for dimension reduction by examining the Manhattan differences of classes as pairs. From each pair, top N features with the largest amount of variance are selected. Then, these features are gathered and common features among the pairs are removed. We estimated the building and floor of users in the dataset. For this purpose, Random Forest (RF) is used in WEKA. This paper is organized as follows; in Chapter 2, RF is described. Proposed method and experiments are given in Chapter 3 and 4 respectively.

II. RANDOM FOREST

A random forest multi-class classifier consists of a number of trees, with each tree using some form of randomization. The leaf nodes of all trees are labeled by estimations of the posterior distribution over the image classes. Each internal node contains a test that best splits the space of data to be classified. A test data is classified by sending it down each tree and aggregating the reached leaf distributions [13]. Randomness can be injected at two points during training: in subsampling the training data so that each tree is grown using a different subset; and in selecting the node tests. The trees here are binary and are constructed in a top-down manner. The binary test at each node can be chosen in one of two ways: (i) randomly, for example data independent; or (ii) by a greedy algorithm which picks the test that best separates the given training examples. The best one is measured by the information gain

MUSTAFA EREN YILDIRIM is with the Department of Electrical and Electronics Engineering, Bahcesehir University, Istanbul, 34353 TURKEY e-mail: mustafaeren.yildirim@eng.bau.edu.tr

<https://orcid.org/0000-0000-0662-2770>

Manuscript received June 24, 2021; accepted Jan 28, 2022.

DOI: 10.17694/bajece.956866

$$\Delta E = -\frac{|Q_i|}{Q} E(Q_i) \quad (1)$$

caused by partitioning the set Q of examples into two subsets Q_i according to the given test. Here $E(q)$ is the entropy $-\sum_{j=1}^N p_j \log_2(p_j)$ with p_j the proportion of examples in q belongs to class j and $|\cdot|$ the size of the set. The process of selecting a test is repeated for each nonterminal node, using only the training examples falling in that node. The recursion is stopped when the node receives too few examples or when it reaches a given depth.

III. PROPOSED METHOD

Before using the data for model training or testing, we replace +100 values with -110 to indicate a very weak signal, as recommended by work in [9]. In the proposed method, we aim to decrease the computation time and load by reducing the feature dimension while obtaining a high hit rate. In case of buildingID in the given dataset, a row appears as $[RSSI_1, RSSI_2, RSSI_{520}, buildingID]$, where the first 520 elements are the measured signal levels and the last element is the building number. For each building, the average value for all 520 features are obtained. Since we have 3 buildings in our dataset, results is as below where the first 520 features are the expected values of features and last one is the building number again. \bar{A} vs \bar{A}

$$\begin{aligned} & [\overline{RSSI_1}, \overline{RSSI_2}, \dots, \overline{RSSI_{520}}, 0] \\ & [\overline{RSSI_1}, \overline{RSSI_2}, \dots, \overline{RSSI_{520}}, 1] \\ & [\overline{RSSI_1}, \overline{RSSI_2}, \dots, \overline{RSSI_{520}}, 2] \end{aligned} \quad (2)$$

To obtain the most informative features of each class, we find the absolute inter-class difference between features set of each class pairs of 0-1, 0-2 and 1-2. Three arrays are obtained and then sorted in descending order. The greater difference value refers to a more distinguishing feature. In each difference array, we eliminate the features which have low difference. Top N features with highest difference are selected from each array and merged together to generate the final feature set. The size of final set is $3 \times N$ decreased by the number of repeated features.

In case of floorID in the given dataset, similar to how it appeared in building case, a row appears as $[RSSI_1, RSSI_2, RSSI_{520}, floorID]$, where the first 520 elements are the measured signal levels and the last element is the floor number. For each floor, the average value for all 520 features are obtained. Since we have 5 floors in our dataset, results are as below where the first 520 features are the expected values of features and last one is the floor number.

$$\begin{aligned} & [\overline{RSSI_1}, \overline{RSSI_2}, \dots, \overline{RSSI_{520}}, 0] \\ & [\overline{RSSI_1}, \overline{RSSI_2}, \dots, \overline{RSSI_{520}}, 1] \\ & [\overline{RSSI_1}, \overline{RSSI_2}, \dots, \overline{RSSI_{520}}, 2] \end{aligned} \quad (3)$$

When we calculate the interclass differences between 5 floors, it gives 10 arrays coming from the pairs of 0-1, 0-2, 0-3, 0-4, 1-2, 1-3, 1-4, 2-3, 2-4, 3-4. In each difference array, we eliminate the features which have low difference.

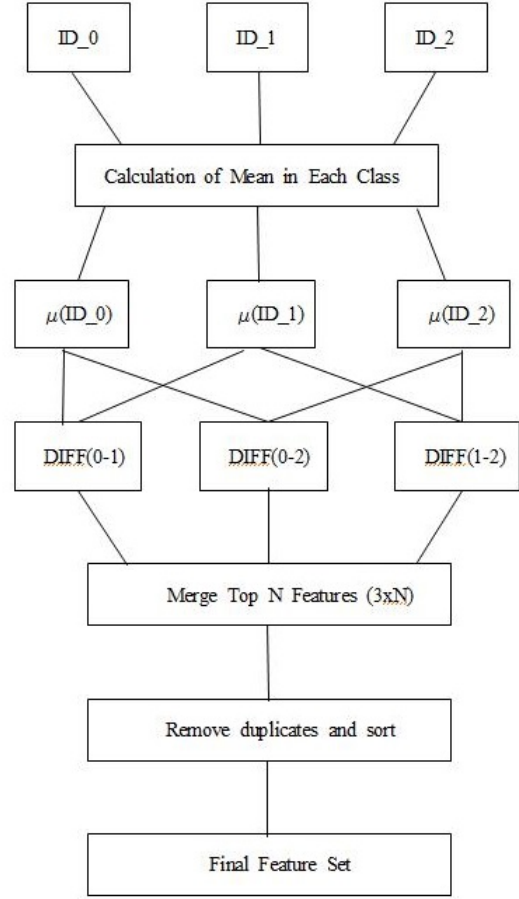


Fig. 1. Block diagram of the proposed method.

The top N features with highest difference are selected from each array and merged together to generate the final feature set. The size of final set is $10 \times N$ decreased by the number of repeated features. The value of N is heuristically obtained and it directly affects the performance of the algorithm.

Block diagram of the proposed feature selection method for a three class example is as shown in Fig. 1.

IV. EXPERIMENTS

The tests were conducted in WEKA 3.8.2 tool. We approached the problem of estimating floor and building independently from each other. Random forest algorithm is used for the measurement method in both experiments.

A. Dataset

We used UJIIndoorLoc dataset which is a publicly available dataset to test the performance of our proposed method. It consists of multiple buildings and multiple floors to build indoor positioning systems that rely on WiFi RSSI fingerprints. It was generated in 2013 at the University of Jaume I, Spain by 20 users with 25 Android devices in a WLAN with 520 AP. It covers an area of 110,000 m^2 . It has 19,937 training fingerprints and 1111 testing fingerprints. The RSSI values are

TABLE I
CHARACTERISTICS OF THE UJINDOORLOC DATASET

	Training	Testing
Sample Number	19,937	1111
Building Number	3	3
Floor Number	5	5

in range of -104 dBm to 0 dBm indicating the worst and best signal levels. RSSI is equal to +100 dBm if AP is not detected. However, 96% of the measured values are in the range of -95 dBm and -45 dBm. Figure-2 shows the measured RSSI distribution of all the intensity values.

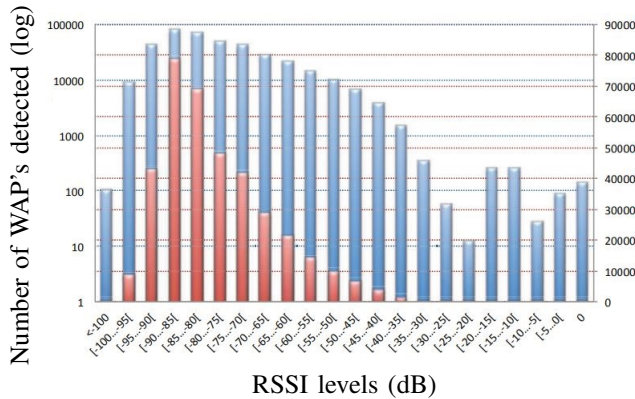


Fig. 2. Frequency distribution of the number of times that a RSSI value appears in the proposed database. Red bars stand for the values in linear scale (right scale) and blue bars stand for the values in logarithmic scale (left scale) [8].

There are three multifloor buildings. There are 4 floors in building0, 4 floors in building1 and 5 floors in building2. The floors are represented as floor0, floor1, floor2, floor3 and floor4. Table 1 shows the characteristics of the used dataset.

B. Hit rate evaluation for floor and building

In both building and floor estimation, tests are conducted for N=25 and N=50. After merging the top N features coming from all three pairs and removing the duplicates, feature set size becomes 50 and 97 for N=25 and N=50, respectively. This means that there is approximately 66% similarity in eature characteristics among each pair. The comparison of the proposed method is done with unreduced feature set and a state-of-art study. According to the results in Table. 2, on the testing dataset with 1111 samples 99.1% and 97.6% hit rates are obtained when N=50 and N=25 rectively.

TABLE II
BUILDING HIT RATE (%)

N=520	N=50	N=25	[12]
100	99.1	97.6	99.4

Training durations of each test are given in Table. 3. According to the results, proposed method decreases the training time by 39% and 60% for N=50 and N=25, respectively.

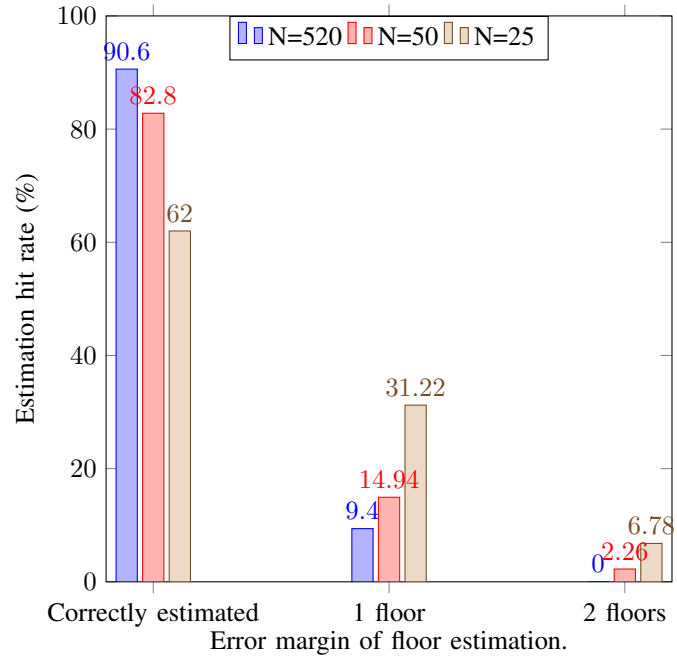


Fig. 3. Distribution of floor estimation results.

Unfortunately, time information is not available for the study [12].

TABLE III
COMPUTATION TIME FOR MODEL TRAINING FOR BUILDING ESTIMATION (s).

N=520	N=50	N=25
23.26	14.29	10.6

In case of floor estimation, model with N=50 obtained 82.8% hit rate while the model using all features obtained 90.6%. The reason of lower accuracy in floor estimation is that the floor classes do not depend on buildings. Building0 and building1 have 4 floors and building2 has 5 floors. The floor id numbers do not differ among the buildings.

Although the accuracy is low in case of N=25, most of the misclassifications are only 1 floor up or down from the actual floor. According to Fig. 3, when N=25, out of 38% of misclassification rate, 31.22% arises from being misclassified with 1 floor up or down. The same parameter has the values of 9.4% and 14.94% for N=520 and N=50, respectively.

When N=520, there is no misclassification of floor with error margin of two. The misclassification with two floors error margin occurs only with the reduced sample size. The maximum error in floor estimation is two floors.

TABLE IV
FLOOR HIT RATE (%)

N=520	N=50	N=25	[12]
90.6	82.8	62	78.2

Training durations of each floor estimation test are given in Table. 5. According to the results, in case of floor estimation,

proposed method decreases the training time by 45% and 54% for N=50 and N=25, respectively.

TABLE V
COMPUTATION TIME FOR MODEL TRAINING FOR FLOOR ESTIMATION (S).

N=520	N=50	N=25
47.35	26.4	21.8

V. CONCLUSION

In this paper, a new dimension reduction approach is proposed for RSSI based indoor localization. The most distinguishing features of each class are extracted by pair-wise analysis with other classes.

We applied our method for the estimation and building estimation in a publicly available dataset. The proposed approach significantly decreased the training time for the model whereas leading a slightly decrease in the estimation accuracy compared to the case in which the total feature set is used.

The drawback of the system is that the value of N is heuristic and it depends on the dataset size. For future study, we aim to employ several deep learning models to increase the accuracy.

REFERENCES

- [1] W. Cui, L. Zhang, B. Li, J. Guo, W. Meng, H. Wang, and L. Xie, "Received signal strength based indoor positioning using a random vector functional link network," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 5, pp. 1846–1855, 2018.
- [2] K. Lee, Y. Nam, and S. D. Min, "An indoor localization solution using bluetooth rssi and multiple sensors on a smartphone," *Multimedia Tools and Applications*, vol. 77, pp. 1–20, 05 2018.
- [3] F. Seco and A. R. Jiménez, "Smartphone-based cooperative indoor localization with rfid technology," *Sensors*, vol. 18, no. 1, 2018. [Online]. Available: <https://www.mdpi.com/1424-8220/18/1/266>
- [4] Z. Liu, L. Zhang, Q. Liu, Y. Yin, L. Cheng, and R. Zimmermann, "Fusion of magnetic and visual sensors for indoor localization: Infrastructure-free and more effective," *IEEE Transactions on Multimedia*, vol. 19, no. 4, pp. 874–888, 2017.
- [5] H. Zhang, K. Liu, F. Jin, L. Feng, V. Lee, and J. Ng, "A scalable indoor localization algorithm based on distance fitting and fingerprint mapping in wi-fi environments," *Neural Computing and Applications*, vol. 32, 05 2020.
- [6] I. Alshami, N. Ahmad, and S. Sahibuddin, "Automatic wlan fingerprint radio map generation for accurate indoor positioning based on signal path loss model," vol. 10, pp. 17930–17936, 01 2015.
- [7] S.-Y. Jung, S. Hann, and C.-S. Park, "Tdoa-based optical wireless indoor localization using led ceiling lamps," *IEEE Transactions on Consumer Electronics*, vol. 57, no. 4, pp. 1592–1597, 2011.
- [8] J. Torres-Sospedra, R. Montoliu, A. Martínez-Usó, J. P. Avariento, T. J. Arnau, M. Benedito-Bordonau, and J. Huerta, "Ujiindoorloc: A new multi-building and multi-floor database for wlan fingerprint-based indoor localization problems," in *2014 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, 2014, pp. 261–270.
- [9] M. Nowicki and J. Wietrzykowski, "Low-effort place recognition with wifi fingerprints using deep learning," 11 2016.
- [10] M. Ibrahim, M. Torki, and M. ElNainay, "Cnn based indoor localization using rssi time-series," in *2018 IEEE Symposium on Computers and Communications (ISCC)*, 2018, pp. 01 044–01 049.
- [11] K. S. Kim, S. Lee, and K. Huang, "A scalable deep neural network architecture for multi-building and multi-floor indoor localization based on wi-fi fingerprinting," *Big Data Analytics*, vol. 3, 04 2018.
- [12] K. A. Nguyen, "A performance guaranteed indoor positioning system using conformal prediction and the wifi signal strength," *Journal of Information and Telecommunication*, vol. 1, no. 1, pp. 41–65, 2017.
- [13] A. Bosch, A. Zisserman, and X. Munoz, "Image classification using random forests and ferns," in *2007 IEEE 11th International Conference on Computer Vision*, 2007, pp. 1–8.



Mustafa Eren Yıldırım received his BS degree in Electrical Engineering from Bahcesehir University, Istanbul, Turkey, in 2008 and his MS and PhD degrees in Electronics Engineering from the Graduate School of Electrical and Electronics Engineering, Kyungshung University, Pusan, Rep. of Korea, in 2010 and 2014, respectively. He worked as a researcher and lecturer for Kyungshung University until August 2015. He is currently holding two assistant professor positions in the Department of Electrical and Electronics Engineering, Bahcesehir University and the Department of Electronics Engineering, Kyungshung University. His research interests include image processing, computer vision, and pattern recognition.

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: January 2022
Volume: 10
Issue: 1

CONTENTS

Ş. Sivritaş, A. Tekin; Autonomous 3D Target Tracking via UAV,.....	1 - 9
E. Kuşkan, M.A. Sahraei, M.Y. Çodur; Classification of Aviation Accidents Using Data Mining Algorithms,.....	10 - 15
T. Ö. Onur, G. U. Kaya; Application of Binary Genetic Algorithm for Holographic Vascular Mimicking Phantom Reconstruction,.....	16 - 22
H. Mutlu, C. Akuner, G. Akgün; An Intelligent Machine Condition Monitoring Model for Servo Systems,.....	23 - 29
T. Tut; Broadband Low Reflection Surfaces with Silicon Nano-pillar Square Arrays for Energy Harvesting,.....	30 - 34
H. Firat, M. E. Asker, D. Hanbay; Hybrid 3D Convolution and 2D Depthwise Separable Convolution Neural Network for Hyperspectral Image Classification,.....	35 - 46
N. Noğay, H. S. Noğay; Estimation of Survival According to Body Mass Index (BMI), Hypertension, Diabetes and Heart Disease with Optimizable Decision Trees,.....	47 - 52
H. Sezer, A. Güven, O. B. Tör, M. E. Cebeci, S. Teimourzadeh; A Cost-Based Approach to Elicit Ancillary Service Tariffs from Unified Capacity and Energy Tariffs,.....	53 - 63
F. A. Kazan, R. Akkaya; The Effect of Road Slope on Parameters of Disabled Vehicle Driven by PMDC Motor,.....	64 - 74
A. Gunagwera, A. T. Zengin; Longitudinal Inter-Vehicle Distance Control of Autonomous Vehicle Platoons Subjected to Internal and External Disturbances,.....	75 - 84
T. C. Akinci, A. A. Martinez-Morales; Cognitive Based Electric Power Management System,.....	85 - 90
I. B. Çiçek, I. Sel, F. H. Yağın, C. Çolak; Development of a Python-Based Classification Web Interface for Independent Datasets,.....	91 - 96
B. Kocaman, Y. Yiğit; Real-Time Monitoring, Analysis and Control of Power Parameters in Residential Houses Using LabVIEW,.....	97 - 105
M. E. Yildirim; RSSI Based Indoor Localization with Reduced Feature Dimension,.....	106 - 109

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

