



# Balkan Journal of Electrical & Computer Engineering

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

[www.bajece.com](http://www.bajece.com)

Vol : 11

No : 4

Year : 2023

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, University of California Riverside, US

**Vice Editor**  
Hamidreza Nazarpouya, Oklahoma State University, US

**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: 11  
No : 4  
Year: October 2023

#### **CONTENTS**

1. Research Article  
İsa Ataş, Cüneyt Özdemir, Musa Ataş, Yahya Doğan; Forensic Dental Age Estimation Using Modified Deep Learning Neural Network, 298-305
2. Research Article  
Mustafa Can Bingol, Ömür Aydoğmuş; Development of a Human-Robot Interaction System for Industrial Applications, ..... 306-315
3. Research Article  
Abdülkadir Albayrak; Vision Transformer Based Photo Capturing System, ..... 316-321
4. Research Article  
Kadir Yılmaz, Taner Dindar, Murat Ayaz, Serkan Aktaş, Serkan Sezen; Finite Element Analysis of SMC Core Magnetic Gear for Vehicle Powertrain Systems, ..... 322-328
5. Research Article  
Yasin Bektaş; African Vultures Optimization Algorithm-Based Selective Harmonic Elimination for Multi-level Inverter, ..... 329-339
6. Research Article  
Cüneyt Özdemir; Designing Effective Models for COVID-19 Diagnosis through Transfer Learning and Interlayer Visualization, ..... 340-345
7. Research Article  
Olayemi Olaniyi, Shefiu Ganiyu; Intelligent Video Surveillance System Using Faster Regional Convolutional Neural Networks, ..... 346-351
8. Research Article  
Cihan Bilge Kayasandık; Automated Detection of Collagen Bundles in Second Harmonic Generation Microscopy Images, ..... 352-363
9. Research Article  
Ahmet Saim Yılmaz, Haydar Cukurtepe, Emin Kugu; Geo-Location Spoofing on E-Scooters; Threat Analysis and Prevention Framework, ..... 364-372
10. Research Article  
Övül Eski, Kemal Şahin, Sevilay Çetin; A Modular Multilevel Converter-Based Pulsed Electric Field Generator Design for Electroporation Applications, ..... 373-379
11. Research Article  
Çiğdem Bakır; Different Induction Motor Faults by New Proposed Random Forest Method, ..... 380-386
12. Research Article  
Abdülmelik Bekmez, Kadir Aram; Three Dimensional Formation Control of Unmanned Aerial Vehicles in Obstacle Environments, ..... 387-394
13. Research Article  
Mahir Kaya; Bayesian Optimization-based CNN Framework for Automated Detection of Brain Tumors, ..... 395-402

#### **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](mailto:bajece@hotmail.com)




# Forensic Dental Age Estimation Using Modified Deep Learning Neural Network

İsa Ataş\*, Cüneyt Özdemir, Musa Ataş, and Yahya Doğan


**Abstract**—Dental age is one of the most reliable methods to identify an individual's age. By using dental panoramic radiography (DPR) images, physicians and pathologists in forensic sciences try to establish the chronological age of individuals with no valid legal records or registered patients. The current methods in practice demand intensive labor, time, and qualified experts. The development of deep learning algorithms in the field of medical image processing has improved the sensitivity of predicting truth values while reducing the processing speed of imaging time. This study proposed an automated approach to estimate the forensic ages of individuals ranging in age from 8 to 68 using 1332 DPR images. Initially, experimental analyses were performed with the transfer learning-based models, including InceptionV3, DenseNet201, EfficientNetB4, MobileNetV2, VGG16, and ResNet50V2; and accordingly, the best-performing model, InceptionV3, was modified, and a new neural network model was developed. Reducing the number of the parameters already available in the developed model architecture resulted in a faster and more accurate dental age estimation. The performance metrics of the results attained were as follows: mean absolute error (MAE) was 3.13, root mean square error (RMSE) was 4.77, and correlation coefficient R2 was 87%. It is conceivable to propose the new model as potentially dependable and practical ancillary equipment in forensic sciences and dental medicine.

**Index Terms**—Dental age, deep learning, dental panoramic radiograph, forensic odontology, inceptionV3, regression.


İSA ATAŞ\*, is with Computer Technologies Department, Diyarbakır Vocational School of Technical Sciences, Dicle University, Diyarbakır, Turkey, (e-mail: [isa\\_atas@dicle.edu.tr](mailto:isa_atas@dicle.edu.tr)).

 <https://orcid.org/0000-0003-4094-9598>


CÜNEYT ÖZDEMİR, is with Computer Engineering Department, Engineering Faculty, Siirt University, Siirt, Turkey, (e-mail: [cozdemir@siirt.edu.tr](mailto:cozdemir@siirt.edu.tr)).

 <https://orcid.org/0000-0002-9252-5888>

MUSA ATAŞ, is with Computer Engineering Department, Engineering Faculty, Siirt University, Siirt, Turkey, (e-mail: [musa.atas@siirt.edu.tr](mailto:musa.atas@siirt.edu.tr)).

 <https://orcid.org/0000-0002-1214-3127>

YAHYA DOĞAN, is with Computer Engineering Department, Engineering Faculty, Siirt University, Siirt, Turkey, (e-mail: [yahyadogan@siirt.edu.tr](mailto:yahyadogan@siirt.edu.tr)).

 <https://orcid.org/0000-0003-1529-6118>

Manuscript received Aug 28, 2023; accepted October 22, 2023.

DOI: [10.17694/bajece.1351546](https://doi.org/10.17694/bajece.1351546)

## I. INTRODUCTION

Age determination is a critical research topic in forensic science [1]. Age range based on facial features is considered the most common approach due to its applicability [2]. In forensic units, in cases where the person's age is uncertain and disputed, a medical age assessment may be requested by government agencies for legal action. Therefore, the age estimation process may provide guidance on whether an individual should be considered a child, teenager, or adult [3]. Despite variances in many countries, the legal age range for criminal liability is typically between 14 and 21 [4–6]. The radiographic images of hand bones (metacarpus) and wrist bones (pisiform) and the DPR images are utilized frequently in forensic sciences to identify the age range [7].

This study employed the DPR images for age range determination. After puberty, tooth development decelerates in individuals, and their distinctive tooth properties deteriorate. Therefore, age determination from DPR images becomes challenging after a particular age range [8]. For dental age determination, traditional manual processes are commonly used [9]. Although these manual techniques have been employed properly in diverse populations, there are still certain limitations in clinical applications, such as the technique's subjectivity and measurement bias. These procedures are also monotonous and time-consuming [10]. Therefore, estimating the automatic dental age is crucial to improve the age range accuracy [7].

Traditional automated dental age estimation procedures involve phases, such as image-preprocessing, segmentation, feature extraction, and classification (categorical) or regression (numerical). In the case of classification, these procedures seek to determine individuals' age groups, whereas the regression phase seeks to pinpoint their precise ages. The success of each step in the methods highly depends on its compatibility with the previous ones [11].

Deep learning approaches (artificial vision, object detection, recognition, etc.), which have recently supplanted traditional methods, efficiently address numerous crises in diverse scientific fields [12–16]. The deep learning approach is a machine learning process in which deep neural networks such as convolutional neural networks can work directly on the input images and generate the required output without demanding the execution of intermediate steps such as segmentation and feature extraction. However, efforts to design and train deep neural networks are complex, time-consuming, and costly. Therefore, instead of designing and

training deep networks from scratch, several approaches were developed that use pre-trained deep networks to execute the necessary tasks. Transfer learning is the common term for these methods [17]. Additionally, this study adopted the transfer learning approach in the deep learning model proposed for forensic age estimation. The mentioned model aims to accurately predict an individual's forensic age from DPR images in a precise range. The proposed approach consists of three steps: image-preprocessing, feature extraction, and linear regression. Initially, DPR images are preprocessed and prepared for feature extraction in the image-preprocessing step. The data augmentation method is employed on the data set for this function. In the feature extraction step, however, the feature extraction capability of the pretrained InceptionV3, ResNet50V2, DenseNet201, MobileNetV2, VGG16, and EfficientNetB2 deep-learning models were applied. Finally, age estimation was made using statistical methods (machine learning algorithms) in the stage of linear regression. The remainder of this article is structured as follows: Section 2 includes similar and related works. Section 3 describes materials and methods, and section 4 presents experimental analysis results. The last section, concludes the final evaluations and future perspectives.

## II. RELATED WORK

Cular et al. [7] proposed a dental age estimation method based on DPR images. Accordingly, they combined active appearance and active shape models to determine the outer contour of the teeth. They employed statistical models for feature extraction and a neural network model for age estimation. They generated a dataset including DPR images of 203 individuals to test the effectiveness of the suggested strategy. The data set revealed that the active shape and appearance models had mean absolute errors (MAE) of 2.481 and 2.483, respectively. Stepanovsk, et al. [18] compared the performance of 22 different age estimation approaches in terms of their accuracy and complexity in estimating dental age. They used a dataset containing 976 DPR images. Experimental results suggested that the best practices were multiple linear regression models, tree models, and support vector machine (SVM) models. Hemalatha et al. [19] proposed a classification model grounding on the Demirjian approach to estimate the dental age of Indian children. They used a dataset of images from 100 healthy individuals aged between 4 and 18 to assess the viability of the proposed model. After beginning with preprocessing for noise removal and smoothing from the input images, they segmented the teeth and extracted various features from them. Finally, they applied a fuzzy neural network to perform the classification and concluded that the proposed approach had 89% accuracy. Sironi et al. [1] suggested an age estimation approach based on measuring the amount of the odonblast (pulpal tissue), capturing the images of the pulpal tissue through 3D cone beam computed tomography (CBCT). To estimate dental age in their proposed approach, they generated a dataset containing data on 286 healthy individuals and analyzed it using a Bayesian network. Their findings emphasized that the proposed method was promising for accuracy, bias, and sensitivity from the error matrix (confusion matrix) criteria.

Tao et al. [20] proposed a method of dental age estimation using the multilayer sensor algorithm, one of the machine learning algorithms. They used the k-fold cross-check in the training process to address the overfitting problem. In their experiments, they employed a dataset including 1,636 images. Their findings demonstrated that, in terms of MAE, MSE, and RMSE, the proposed method outperformed the traditional methods such as Demirjian and Willem. Back et al. [21] presented a dental age estimation technique based on DPR images. They employed a Bayesian convolutional neural network for uncertainty and age estimation, conducting experiments on the DPR image dataset of 12,000 individuals. According to the outcomes, they reported that the proposed approach had a correlation coefficient of 0:91. Kim et al. [22] defined a deep learning-based automatic dental age estimation approach, using a convolutional neural network (CNN) for age estimation. Their dataset contained 9,435 DPR images from three different age categories. Based on their result, they reported that the proposed approach revealed excellent performance. Asif et al. [23] studied a statistical model of dental age estimation Based on a volumetric examination of the dental pulp/tooth ratio. They applied simple linear regression and Pearson correlation analysis to the proposed model. Their dataset included 3D CBCT images of 300 individuals separated into five age groups. Their results indicated that the proposed approach figured a 6.48 MAE. Farhadian et al. [24] proposed a dental age estimation approach using a neural network model to perform the age estimation task. They experimentally verified the superiority of the neural network over a regression model using a dataset of 300 CBCT scans in their experiment. They identified the performance metrics as 4.12 MAE and 4.4 RMSE. Mualla et al. [25] described a transfer learning-based automated method for estimating dental age. Two deep neural networks, AlexNet and ResNet, were used in their strategy to extract features. They used several classifiers to execute the classification task, including a decision tree, k-nearest neighbor, linear discriminant, and support vector machine. They constructed a dataset containing 1,429 DPR images and tested various suitable performance metrics. Guo et al. [10] compared the performance of manual and machine learning methods for age estimation from DPR images. They executed analyzes of a dataset including 10,257 individuals' images. As a result, they reported that the end-to-end CNN models performed better, according to the comparison of the manual method and machine learning.

## III. MATERIALS AND METHODS

An automated dental age estimation approach based on DPR images was proposed in this section. This approach sought to identify the age group using DPR images and included three primary steps: image-preprocessing, feature extraction, and regression.

### A. Image pre-processing

A data set that included 1,332 DPR images between the ages of 8-68 and acquired from the Periodontology clinic of Diyarbakır Oral and Dental Health Hospital was analyzed. Fig. 1 depicts the distribution of DPR images by age. Before

processing the feature extraction, the DPR images in the dataset underwent two preprocessing stages. Initially, the images were scaled to  $256 \times 256$  resolution to suit the deep neural network requirements. Then, the data augmentation technique was applied to improve the model's success on diverse images and minimize the overfitting effect. This technique is used in expanding the training data set by entraining additional images that are variants of the existing images in the dataset. In an effort to establish the optimal image values and investigate how the data diversity method's various functions affect images, the fairest values were designated.

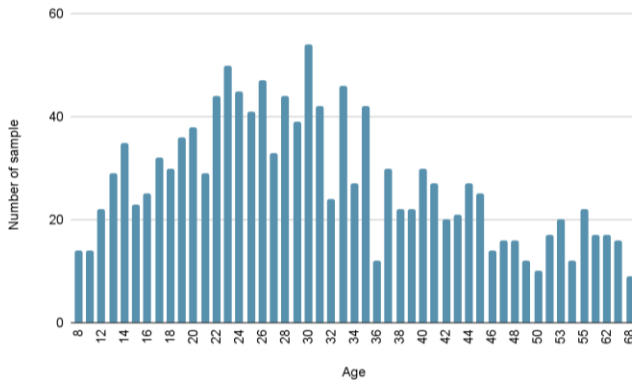


Fig. 1 Distribution of DPR images by age range.

The DPR images were subjected to the data augmentation technique with the following preferences: rotation between -5 and +5, 15% augmentation, 10% horizontal and vertical scrolling, and 70% - 110% brightness. Fig. 2 demonstrates the data augmentation for DPR images. Experimental analyzes were performed on a computer with an Intel® Core™ i7 2.4Ghz CPU @ 1.60 GHz processor with 16GB primary memory, 4-GB NVIDIA GeForce GTX 1080 graphics card.

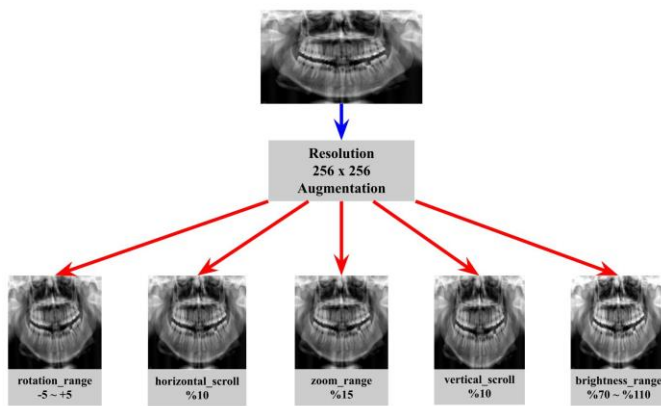


Fig. 2 Data augmentation process for DPR images.

**B. Feature extraction**

The dimensionality reduction procedure known as feature extraction divides the dataset into more logical groups for processing. Convolutional Neural Networks (CNN) is a deep learning algorithm used in image processing and working with images as input. This algorithm utilizes various techniques to

capture the images' features and consists of several layers. A CNN typically contains three layers: convolutional, pooling, and fully connected [26]. The proposed study carried out the feature extraction step by applying transfer learning technique to InceptionV3 [27], MobileNetV2 [28], ResNet50V2 [29], EfficientNetB4 [30], VGG16 [31] and DenseNet201 [32] deep learning models. As a deep learning strategy, transfer learning uses pre-trained model parameters on a sizable dataset (ImageNet, COCO, etc.). In addition, transfer learning is used for many reasons. For instance, it is a challenging issue to train a CNN from scratch using random initial values in case of an insufficient dataset. Therefore, using the weights of a pre-trained network as initial values may address many existing problems effectively. Transfer learning models are used to accelerate the speed of this process and generate the best practicing model. Fig. 3 displays the transfer learning procedure. As taken reference in this study, the analyses made with the deep learning models revealed that the InceptionV3 model outperformed other models in terms of performance. As a result, the InceptionV3 model underwent a modification process.

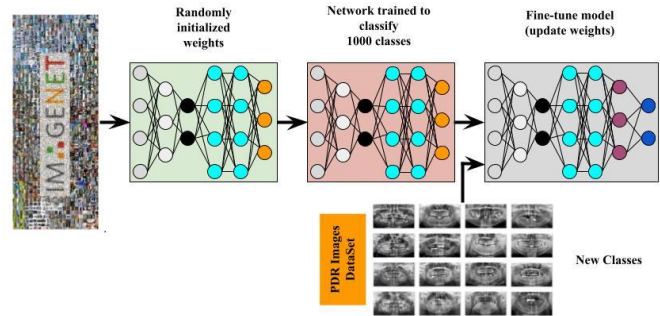


Fig. 3 Transfer learning procedure.

**1) Feature extraction using InceptionV3**

InceptionV3 is a convolutional neural network model used in image analysis and recognition problems and consists of multiple-convolution and maxpooling layers. The bottom layer contains a fully connected neural network. The prominent component of the model is that it replaces small kernels with large kernels by learning multi-scale representations to minimize computational complexity and the total number of parameters [27]. Fig. 4 depicts the typical functional architecture of the InceptionV3 model.

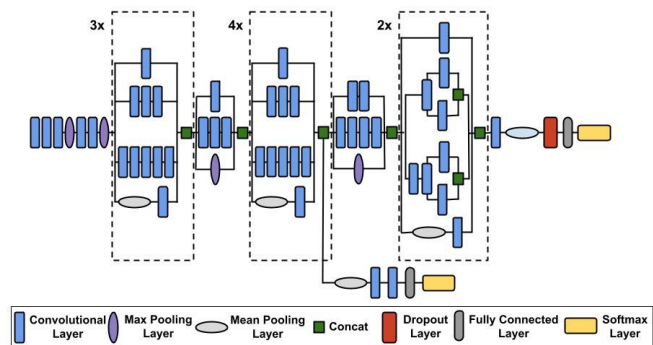


Fig. 4 The architecture of InceptionV3 model.



The InceptionV3 model contains 11 different blocks. The model’s architecture underwent special modifications/ adjustments to achieve higher model success. The blocks labeled ‘mixed’ were omitted from the model to simplify the algorithm, lower the number of parameters, and obtain the most optimal results from the available data set. As a result, the InceptionV3 model generated a large number of sub-patterns. Subsequently, the generated sub-patterns were labeled with the block number. For instance, the sub-pattern consisting of the first seven blocks was called InceptionV3Mixed7. The InceptionV3Mixed\_04 model performed the best among the sub-patterns. Fig. 5 displays the sub-pattern extraction derived from the InceptionV3 model.

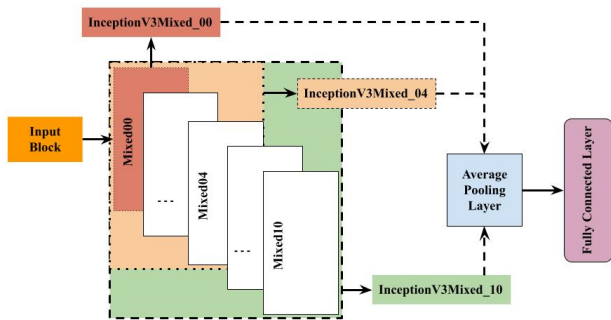


Fig. 5 Inter-block exchange in InceptionV3 architecture.

C. Linear Regression

Machine learning refers to the algorithms that train to predict an outcome after learning from a set of user-generated data. The model training proceeds until the performance hits up to the optimal level. An outcome-predicting algorithm from an available input data set is called a supervised learning algorithm, and it involves classification and regression methods to develop prediction models. In a supervised learning algorithm, if the predicted responses are labels, the classification method is used; however, if they are within a particular value range, then the regression method is used. The term regression denotes the analysis of estimating the independent variable from the dependent variables. The utmost preferred two regression types in machine learning are logistic and linear regressions. If the target (independent variable) is categorical, it is typically binary (logistics) regression, whereas the reference is the linear regression if it is continuous. Mean square error (*MSE*), root mean square error (*RMSE*), mean absolute error (*MAE*), and coefficient of determination ( $R^2$ ) are the metrics used for assessing the regression model in convolutional neural network and machine learning architectures [33].

Mean Square Error (*MSE*): It measures the performance of the imputed values in the machine learning model, and the results are positive. Metric values close to zero refer to better performance [13].

$$MSE = \frac{1}{N} \sum_{i=1}^n (\hat{A}_i - A_i)^2 \tag{1}$$

Root Mean Square Error (*RMSE*): It denotes the square root of the *MSE*. When the *MSE* value is significantly larger than the other metrics, it is typically preferred over the *MSE* metric to facilitate interpretation [13].

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^n (\hat{A}_i - A_i)^2} \tag{2}$$

Mean Absolute Error (*MAE*): It is a metric measuring two continuous variables. It takes the sum of the absolute error values because it accurately represents the sum of the error terms. The *MAE* value is frequently used in regression and time series problems because it is interpretable [13].

$$MAE = \frac{1}{N} \sum_{i=1}^n |(\hat{A}_i - A_i)| \tag{3}$$

R square ( $R^2$ ): It is a statistic used to estimate the performance of regression models. The frequency with which the independent variable influences the dependent variable is indicated by this statistic. Additionally, it displays the 0 to 1% linkage power between the independent and dependent variables [13].

$$R^2 = \frac{(\frac{1}{N} \sum_{i=1}^n (A_i - \underline{A})^2) - (\frac{1}{N} \sum_{i=1}^n (\hat{A}_i - A_i)^2)}{\frac{1}{N} \sum_{i=1}^n (A_i - \underline{A})^2} \tag{4}$$

Here,  $A_i$  refers to the estimated value, while  $\hat{A}_i$  stands for actual age and  $\underline{A}$  the true mean age. *N* denotes the total number of samples. *RMSE* and *MAE* are positive values, and these statistical criteria should be smaller. Values close to zero indicate that the forensic age range predictions are reasonably accurate. For the models to be successful,  $R^2$  should be close to 1.

Fig. 6 illustrates the general deep learning architecture used for age determination from PDR images. Accordingly, following the loading procedure of the preprocessed DPR images in Fig. 6, the input feature is extracted in the CNN architecture. Finally, these features are integrated into the regression in the ML layer, and the model is trained for age estimation.

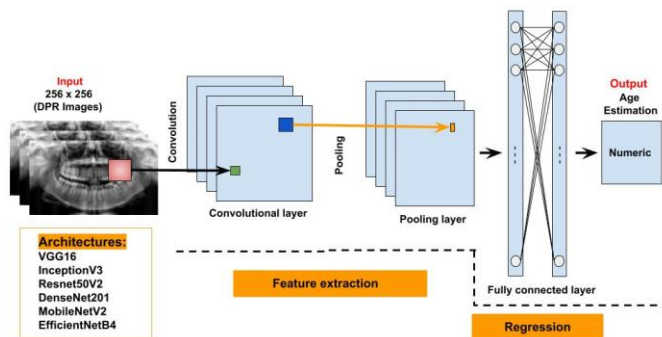


Fig. 6 Feature Extraction and Regression from DPR Images Based on Deep Learning.

#### IV. EXPERIMENTAL RESULTS

This study utilized DPR images for a fully automated forensic age estimation. Deep learning models were used to perform experimental analyses on the provided data sets, and subsequently, the data augmentation technique was applied to increase the model success on images. The data augmentation technique acquires batch images and uses a series of random transformations to each image in the batch. The best values from DPR photographs were determined by analyzing how each function affected the photos in the data triangulation process. Only the noising (normalizing) method of the transfer learning model was applied to the images reserved for the validation and test data sets. For all models, DPR image sizes were scaled to  $256 \times 256$ , subjecting the epoch preference to 500 for training. The Adam algorithm was used as the optimization function, taking the default values as reference. However, if the value for validation loss failed to reach the minimum level after every seven epochs, the step size was reduced by 0.8. Finally, the training process ended if the validation loss remained at epoch value 25 (validation loss). The models were not adjusted at all. We took a balanced approach to all models. Since the DPR images were untrained with ImageNet from extensive image databases, parameters were trained entirely for deep learning, and their weights were recalculated accordingly. Of the 1,332 DPR images in the dataset, 932 (70%), 200 (15%), and 200 (15%) were used for training of the selected models, validation, and testing processes, respectively. Table 1 lists the performance criteria for the test dataset of the deep learning models applied in the study.

TABLE I  
RESULTS ACCORDING TO TRANSFER LEARNING MODELS

Model	Total Parameter	MAE	RMSE	R <sup>2</sup>
EfficientNetB4	19,5M	3.47	5.11	0.85
ResNet50V2	25,6M	4.35	6.20	0.78
DenseNet201	20,2M	4.22	5.99	0.80
InceptionV3	23,9M	3.44	5.20	0.85
MobileNetV2	3,5M	3.68	5.17	0.85
VGG16	138,46M	10.46	12.60	0.11

Analysis of Table 1 revealed that the InceptionV3 model generated the best performance values in the InceptionV3 model. Parameter models with large quantities seemed to contribute less to the results in training the DPR images. Yet, it was also observed that the performance values of both InceptionV3 and MobileNetV2 models were comparable, as the MobileNetV2 model had the smallest parameter among the models selected. The InceptionV3 model was modified to acquire the most optimum values. As a result, several blocks were eliminated from the InceptionV3 model, resulting in the generation of new sub-patterns and analyzing their effects on the outcomes. Table 2 lists the performance criteria achieved by these processes. As deduced from Table 2, the InceptionV3Mixed\_04 model produced the best performance values after modifying the InceptionV3 model. Additionally,

the InceptionV3Mixed\_06 and InceptionV3Mixed\_07 models generated the most comparable values to these outcomes. The InceptionV3 model had about 24M trainable parameters, while the InceptionV3Mixed\_04 model had 3,68M after the elimination procedure. As a result of our experimental study, faster and more successful results were obtained with approximately 6 times less parameters.

TABLE II  
PERFORMANCE RESULTS OF MODIFIED INCEPTIONV3  
SUBMODELS

Model	Total Parameter	MAE	RMSE	R <sup>2</sup>
InceptionV3Mixed_03	2,38M	3.64	5.11	0.85
InceptionV3Mixed_04	3,68M	3.13	4.77	0.87
InceptionV3Mixed_05	5,37M	3.49	5.06	0.86
InceptionV3Mixed_06	7,06M	3.15	4.77	0.87
InceptionV3Mixed_07	9,20M	3.18	4.79	0.87
InceptionV3Mixed_08	11,04M	3.55	5.08	0.85
InceptionV3Mixed_09	16,29M	3.46	4.95	0.86

Fig. 7 displays the MAE and loss plots during the training and validation processes of the InceptionV3Mixed\_04 model.

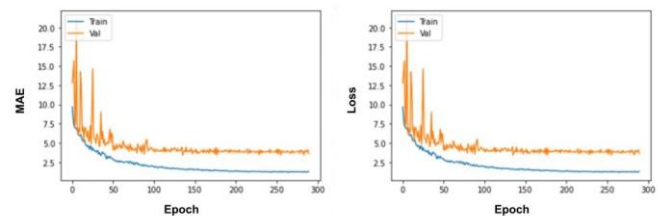


Fig. 7 MAE and loss graphs of InceptionV3Mixed\_04 model.

Fig. 8 displays the statistical distribution of the DPR image's actual age values and the model's predicted age values from testing with the InceptionV3Mixed\_04 model. Fig. 8 illustrates that the actual and estimated age values of the InceptionV3Mixed\_04 model are reasonably similar to each other.

However, the estimation rate was lower for individuals aged over 55. Consequently, the absence or limited image counts in the dataset for ages older than 55 years-old patients and the advanced age degradation in the maxilla and mandible areas could be two potential factors among the reasons. Therefore, it is conceivable to argue that the model's learning capacity for images over 55 years and older was insufficient.

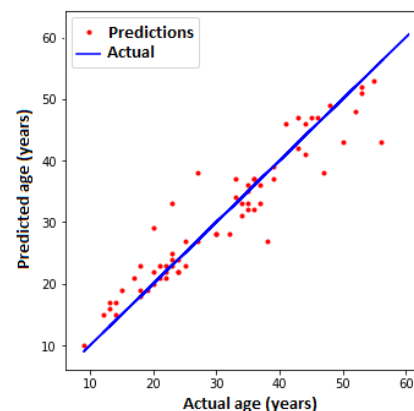


Fig. 8 Statistical distribution of actual age values and estimated age values.



A heat map (Grad-CAM) was created to resolve which regions the InceptionV3Mixed\_04 model concentrated. Grad-CAM shows the distinctive are in the images that a trained model is more likely to detect [34]. Fig. 9, illustrates Grad-CAM heatmaps for the compared female and male DPR tooth images. According to the heatmaps, the InceptionV3Mixed\_04 model focused on the teeth, gingival tissue, and upper jaw (maxilla).

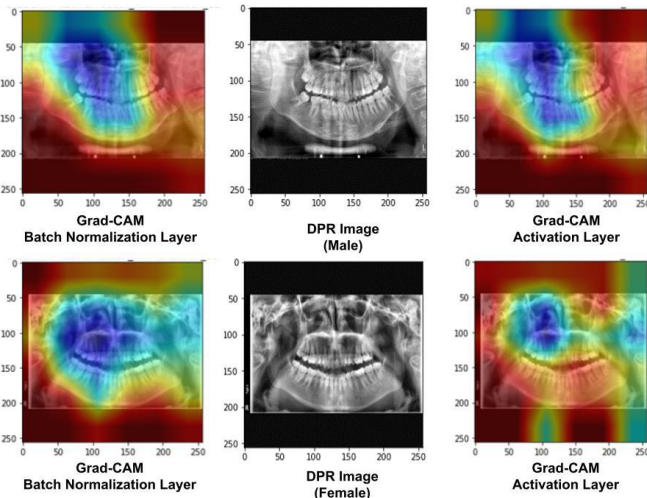


Fig. 9 Heatmaps of our proposed model on male and female DPR images.

Fig. 10 displays the actual and estimated age values in the various DPR images of the InceptionV3Mixed\_04 model. Table 3 summarizes the literature on Dental Age Estimation using DPR pictures. Since there was no code or DPR dataset for dental age estimation in the literature, we failed to make a rational comparison between our proposed study and the previous studies. Analysis of Table 3 revealed that studies in the literature typically focused on a limited age range. Kim et al. [22] demarcated the dental ages of individuals between the ages of 2 and 98 by dividing them into age categories. In their study, Asif et al. [23] reported (MAE=6.48 - 8.58 years) in the 16-65 age range (a wide age range). This study, however, found the following metric values in the 8-68 age range: (MAE=3.13 years, RMSE=4.77 years, and  $R^2=0.87$ ). These findings proved that the current study is one of the attempts that has produced the best outcomes in a wide age range on age determination from dental images to date. In addition, age group analysis was performed from DPR images to compare MAE and RMSE estimation errors according to age ranges, and the test performance results are given in Table 4.

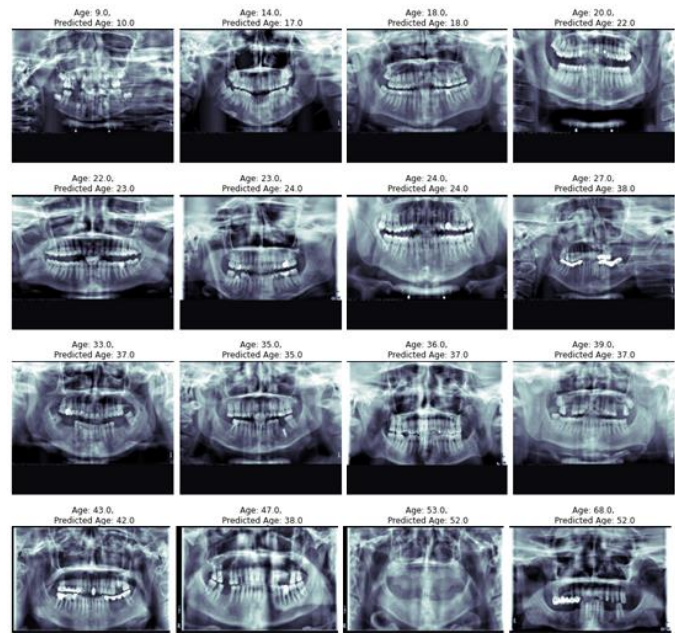


Fig. 10 Age estimation detection of our proposed model on different DPR images.

TABLE III  
PERFORMANCE RESULTS OF MODIFIED INCEPTIONV3 SUBMODELS

Model	Province	Age Range	Dataset	Performance
[7]	Croatia	10-25	283	MAE=2.3 - 2.5
[9]	Germany	5-24	10,257	ACC (14) = 95.9% ACC (16) = 95.4% ACC (18) = 92.3%
[19]	Malaysian	4-18	426	ACC=89% F1 SCORE=71% RECALL=92%
[21]	Germany	5-25	619	MAE=2.34 - 4.61 RMSE=5.58 - 7.49 $R^2=0.92 - 0.96$
[22]	Korea	2-98	9,485	MAE (2-11) = 0.8 MAE (12-18) = 1.2 MAE (19-) =4.4
[23]	Malaysian	16-65	300	MAE=6.48 - 8.58
[35]	Poland	4-15	619	MAE=2.34 - 4.61 RMSE=5.58 - 7.49 $R^2=0.92 - 0.96$
This Work	Turkey	8-68	1,332	MAE=3.13 RMSE=4.77 $R^2=0.87$

Since the tooth development stages of patients younger than 30 years of age are different in age-range experiments, the estimation ability is expected to be high [36]. Also, for those older than 30, which is when the permanent teeth are completely formed, it is difficult to assess age accurately and it has been confirmed by clinical analysis [37].

Low estimation errors of MAE and RMSE according to years were obtained in Table 4 for 10 - 19 and 20 - 29 age ranges. In addition, the age regression model performed poorly as it was unable to precisely determine the relevant age characteristics from images of older individuals.

TABLE IV  
PERFORMANCE COMPARISON OF AGE ESTIMATION ERROR METRICS FOR AGE GROUPS

Age Group (years)	MAE	RMSE	#Images
8 - 9	0	0	6

10 - 19	2.39	3.07	34
20 - 29	2.41	3.53	50
30 - 39	3.58	4.92	41
40 - 49	4.33	5.50	30
50 - 59	6.77	6.87	23
60 - 68	13.75	14.09	16

## V. CONCLUSION

Age determination of individuals is a critical and fundamental element in forensic identification and the work of physicians in administering medications. It is achievable to utilize the traditional age determination methods in forensic medicine after years of education and training procedure of the medical experts. Hitherto, age estimation studies have produced a sizable number of approaches and analyses, each of which touts a variety of uses, criteria for accuracy, and levels of reliability. Nonetheless, age estimations based on dental discoveries are regarded as a depiction of the closely estimated age compared to actual chronological age. Minimizing the average error rates in age estimation is also critical in making the most realistic age determination. This study proposed a deep transfer learning-based modified InceptionV3 model using DPR images. A total of 1,332 DPR images were used for experimental analyses among individuals whose ages ranged from 8 to 68. The modifications in the InceptionV3 model introduced the most optimum results with the InceptionV3Mixed\_04 model. The metrics of the proposed model were (MAE = 3.13 years, RMSE = 4.77 years) error rates, and ( $R^2 = 87.2$ ) for accuracy. The current study emphasized the criticalness of modifying transfer learning models to minimize costs and save time rather than developing a new deep learning architecture. As a result, this study developed the InceptionV3Mixed\_04 module with fewer parameters than the InceptionV3 model; consequently, it delivered results faster and estimated ages more precisely. This study also proved that even architectures with fewer parameters of DPR images absent in the ImageNet dataset could still outperform transfer learning models in terms of performance. Generating the heatmap of the InceptionV3Mixed\_04 model uncovered that the right upper jaw (maxilla) and upper teeth (right) were the primary focus of the images. Therefore, we opine that the current study, which aimed to focus on one of the broad age ranges in age estimation, will yield promising outputs for future studies.

## REFERENCES

- [1] E. Sironi *et al.* 'Age estimation by assessment of pulp chamber volume: a bayesian network for the evaluation of dental evidence', *International Journal of Legal Medicine*, 2018, vol. 132, no. 4, pp. 1125–1138.
- [2] J. Lu, V. E. Liong, and J. Zhou, 'Cost-sensitive local binary feature learning for facial age estimation', *IEEE Transactions on Image Processing*, 2015, vol. 24, no. 12, pp. 5356–5368.
- [3] A. Schmeling, and S. Black, 'An introduction to the history of age estimation in the living. Age Estimation in the Living', *Chichester, UK. John Wiley & Sons Ltd*, 2010, pp. 1–18.
- [4] A. Olze *et al.* 'Assessment of the radiographic visibility of the periodontal ligament in the lower third molars for the purpose of forensic age estimation in living individuals', *International journal of legal medicine*, 2010, vol. 124, no. 5, pp. 445–448.
- [5] R.B. Basset, C. Briggs, and O.H. Drummer 'Age estimation and the developing third molar tooth: an analysis of an australian population using computed tomography', *Journal of forensic sciences*, 2011, vol. 56, no. 5, pp. 1185–1191.
- [6] K.A. Kasper *et al.* 'Reliability of third molar development for age estimation in a texas hispanic population: a comparison study', *Journal of forensic sciences*, 2009, vol. 54, no. 3, pp. 651–657.
- [7] L. Cular *et al.* 'Dental age estimation from panoramic x-ray images using statistical models', *In: Proceedings of the 10th International Symposium on Image and Signal Processing and Analysis IEEE*, 2017, pp. 25–30.
- [8] D. Tabakcilar, R. Bundak, and K. Gencay, 'Dental age in precocious and delayed puberty periods', *European Journal of Dentistry*, 2021, vol. 15, no. 3, pp. 539–545.
- [9] A. Demirjian, H. Goldstein, and J.M. Tanner, 'A new system of dental age assessment', *Human biology*, 1973, pp. 211–227.
- [10] Y.C. Guo *et al.* 'Accurate age classification using manual method and deep convolutional neural network based on orthopantomogram images', *International Journal of Legal Medicine*, 2021, vol. 135, no. 4, pp. 1589–1597.
- [11] M.I. Razzak, and S. Naz, 'Microscopic blood smear segmentation and classification using deep contour aware cnn and extreme machine learning' *In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2017, pp. 49–55.
- [12] İ. Ataş, 'Performance Evaluation of Jaccard-Dice Coefficient on Building Segmentation from High Resolution Satellite Images' *Balkan Journal of Electrical and Computer Engineering*, 2023, vol. 11, no. 1, pp. 100-106. <https://doi.org/10.17694/bajece.1212563>
- [13] C. Özdemir, 'Classification of Brain Tumors from MR Images Using a New CNN Architecture', *Traitement du Signal*, 2023, vo. 40, no. 2, pp. 611-618. <https://doi.org/10.18280/ts.400219>
- [14] M. Ataş, M.İ. Yeşilnacar, and A. Demir Yetiş, 'Novel machine learning techniques based hybrid models (LR-KNN-ANN and SVM) in prediction of dental fluorosis in groundwater', *Environmental Geochemistry and Health*, 2022, vol. 44, no. 11, pp. 3891-3905.
- [15] C. Özdemir, M. Ataş, and A.B. Özer, 'Classification of Turkish spam e-mails with artificial immune system', *21st Signal Processing and Communications Applications Conference (SIU)*, 2013, pp. 1-4.
- [16] Y. Doğan, 'A new global pooling method for deep neural networks: Global average of top-k max-pooling', *Traitement du Signal*, 2023, vo. 40, no. 2, pp. 577-587. <https://doi.org/10.18280/ts.400216>
- [17] M. Castelluccio *et al.* 'Land use classification in remote sensing images by convolutional neural networks', 2015, *arXiv preprint arXiv:1508.00092*.
- [18] M. Stepanovsk'y *et al.* 'Novel age estimation model based on development of permanent teeth compared with classical approach and other modern data mining methods', *Forensic science international*, 2017, vol. 279, pp. 72–82.
- [19] B. Hemalatha, and N. Rajkumar, 'A versatile approach for dental age estimation using fuzzy neural network with teaching learning-based optimization classification', *Multimedia Tools and Applications*, 2020, vol. 79, no. (5-6), pp. 3645–3665.
- [20] J. Tao *et al.* 'Dental age estimation: a machine learning perspective', *In: International Conference on Advanced Machine Learning Technologies and Applications*, 2019, pp. 722–733.
- [21] D. Back *et al.* 'Forensic age estimation with bayesian convolutional neural networks based on panoramic dental x-ray imaging' *Proceedings of Machine Learning Research*, 2019, pp. 1-4.
- [22] J. Kim *et al.* 'Development and validation of deep learning-based algorithms for the estimation of chronological age using panoramic dental x-ray images', *Proc. Mach. Learn. Res.*, 2019.
- [23] M.K. Asif *et al.* 'Dental age estimation in malaysian adults based on volumetric analysis of pulp/tooth ratio using cbct data' *Legal Medicine*, 2019, vol. 36, pp. 50–58.
- [24] M. Farhadian, F. Salemi, S. Saati, and N. Nafisi, 'Dental age estimation using the pulp-to-tooth ratio in canines by neural networks', *Imaging science in dentistry*, 2019, vol. 49, no. 1, pp. 19–26.
- [25] E.H. Houssein, N. Mualla, and M. Hassan, 'Dental age estimation based on x-ray images' *Computers, Materials & Continua*, 2020, vol. 62, no. 2, pp. 591–605.

- [26] W. Yu *et al.* ‘Automatic classification of leukocytes using deep neural network’, *In: 2017 IEEE 12th International Conference on ASIC (ASICON)*, 2017, pp. 1041–1044.
- [27] C. Szegedy *et al.* ‘Rethinking the inception architecture for computer vision’, *In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 2818–2826.
- [28] M. Sandler *et al.* ‘Mobilenetv2: Inverted residuals and linear bottlenecks’, *In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 4510–4520.
- [29] I. Bello *et al.* ‘Revisiting resnets: Improved training and scaling strategies’, *Advances in Neural Information Processing Systems*, 2021, vol. 34, pp. 22614–22627.
- [30] M. Tan, and Q. Le, ‘Efficientnet: Rethinking model scaling for convolutional neural networks’, *In: International Conference on Machine Learning*, 2019, pp. 6105–6114.
- [31] K. Simonyan, and A. Zisserman, ‘Very deep convolutional networks for largescale image recognition’ *arXiv preprint arXiv:1409.1556*, 2014.
- [32] G. Huang *et al.* ‘Densely connected convolutional networks’ *In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 4700–4708.
- [33] O.J. Babajan, R. Bagherpour, ‘Estimating the wear rate of diamond cutting wire bead in building stone cutting using svr and ga mlp system’, *Springer*, 2022, pp. 1–13.
- [34] R.R. Selvaraju *et al.* ‘Grad-cam: Visual explanations from deep networks via gradient-based localization’ *In: Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 618–626.
- [35] M. Zaborowicz *et al.* ‘Deep learning neural modelling as a precise method in the assessment of the chronological age of children and adolescents using tooth and bone parameters’ *Sensors*, 2022, vol. 22, no.2, p. 637.
- [36] N. Vila-Blanco, M.J. Carreira, P. Varas-Quintana, C. Balsa-Castro, I. Tomas, ‘Deep neural networks for chronological age estimation from opg. images’, *IEEE transactions on medical imaging*, 2020, vol. 39, no. 7, p. 2374–2384.
- [37] J.L. Prieto, E. Barberia, R. Ortega, C. Magana, ‘Evaluation of chronological age based on third molar development in the spanish population’, *International journal of legal medicine*, *Springer*, 2005, vol. 119, no. 6, p. 349–354.

#### BIOGRAPHIES

**İsa ATAŞ** has been a Faculty Member with the Computer Technology Department, Diyarbakır Vocational School of Technical Sciences, Dicle University. The research area is deep learning, computer vision, biomedical image analysis, and wireless networks.

**Cüneyt ÖZDEMİR** has been a Faculty Member with the Computer Engineering Department, Siirt University. He studies on computer vision, biomedical image analysis, and data mining.

**Musa ATAŞ** has been a Faculty Member with the Computer Engineering Department, Siirt University, since 2012. He is the developer of Open Cezeri Library domain specific language framework known as Java-based matrix and computer vision library. His research interests include digital signal processing, artificial intelligence, machine learning, autonomous systems, computer and machine vision, robotics, virtual reality, and programming languages.

**Yahya DOĞAN** has been a Faculty Member with the Computer Engineering Department, Siirt University. He studies computer vision, and deep learning, particularly in generative models.



# Development of a Human-Robot Interaction System for Industrial Applications

Musataf Can Bingol and Omur Aydogmus


**Abstract**— The use of robots is on the rise, and this study focuses on developing manufacturing-assistant robot software for small production plants involved in non-mass production. The primary objective is to address the challenges of hiring expert robot operators by creating user-friendly software, thus enabling non-experts to operate robots effectively. The software comprises three main modules: the convolutional neural network (CNN), process selection-trajectory generation, and trajectory regulation. To initiate operations within these modules, operators record the desired process and its trajectory through hand gestures and index finger movements, captured in a video. The recorded video is then separated into images. These images undergo classification by the CNN module, which also calculates the positions of landmarks, such as joints and index finger's fingernail. Out of eight different pre-trained CNN structures tested, the Xception structure yielded the best result, with a test loss of 0.0051. Using the CNN's output data, the desired process is determined, and its trajectory is generated. The trajectory regulation module identifies the connection points between the generated trajectory and the object, subsequently eliminating unnecessary trajectory segments. The regulated trajectory, along with desired tasks like welding or sealing, is simulated using an industrial robot within a simulation environment. In conclusion, the developed software empowers non-expert operators to program industrial robots, particularly beneficial for companies with non-standardized production lines, where hiring expert robot operators might be challenging.

**Index Terms**—Classification and localization, Fingertip detection, Human-robot interaction, Welding process, Sealing process.


## I. INTRODUCTION

ROBOTS CAN be classified according to the location (mobile and fixed), power systems (pneumatic, hydraulic, and electric), locomotion methods (stable, wheeled, legged, and others), or application areas (industrial and non-industrial) [1]. An industrial robot, as defined by the Robotic Institute of America [2], is a programmable mechanical device used to perform dangerous or repetitive tasks with high accuracy,

**Mustafa Can BINGOL**, is with Department of Electrical-Electronic Engineering Burdur Mehmet Akif Ersoy University, Burdur, Turkey, (e-mail: mcbingol@mehmetakif.edu.tr).

 <https://orcid.org/0000-0001-5448-8281>

**Omur AYDOGMUS**, is with Department of Mechatronic Engineering Firat University, Elazig, Turkey, (e-mail: oaydogmus@firat.edu.tr).

 <https://orcid.org/0000-0001-8142-1146>

Manuscript received July 11, 2023; accepted Aug 6, 2023.

DOI: [10.17694/bajece.1326072](https://doi.org/10.17694/bajece.1326072)

replacing human labour. Collaborative robots accounted for 5.37%, 6.59%, and 7.54 % of the installed industrial robots between 2019 and 2021, respectively [3]. These installed industrial robots are used in various manufacturing processes such as machine tending, welding, and assembling, as shown in Figure 1. The aim of the current study is to transform a traditional industrial robot into a modern collaborative robot capable of performing welding and sealing processes.

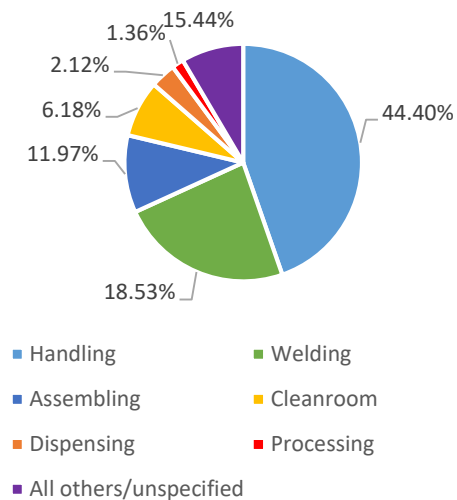


Fig. 1. Task distribution of industrial robots installed in 2022 [3]

Human-robot interaction (HRI) plays a crucial role in enabling human-robot collaboration (HRC). At least one communication channel, such as vision or speech should be used in order to occur HRI. Vision involves the interpretation of images captured by sensors such as cameras. Numerous studies in the literature have utilized vision to establish HRI in literature [4]–[6]. For instance, Hamabe et al. trained a lightweight robot for the assembly process using vision communication channels [7]. In another work, Ding et al. developed robot software to ensure safe manufacturing using vision [8]. In the current study, vision-based HRI software was developed.

Hand gestures recognition and fingertip position determination are commonly used methods for human-computer interaction and HRI [9]–[12]. Raheja et al. calculated to fingertip position by using skin colour of hand [9]. In another study, sequential mathematical operations were employed to obtain the fingertip position after subtracting the color-based hand image from the main image [13]. Other studies have also explored similar approaches using colour-based method [11],

[12]. Another method employed for in hand gesture identification and fingertip position detection involves obtaining information from depth images captured by RGB-D sensors [14], [15]. Shin and Kim achieved an air writing process utilizing an RGB-D sensor and fingertip detection [16]. Similarly, another study successfully detected with successful by using RGB-D sensor [17]. Huang et al. has achieved fingertip detection by using a cascaded convolutional neural network (CNN) and RGB images, employing a different approach than the aforementioned studies [18]. In another study, air writing has been performed using color separation and faster R-CNN structures together [10]. In the robotic field, one notable example involves determining the orientation of the robot using an image from a sensor worn on the operator's wrist, along with hand gesture recognition and fingertip detection [19]. Many other studies have also utilized specialized sensors for detecting hand gestures or fingertips [20], [21]. In the current study, hand gestures and fingertip position were determined by processing the RGB images obtained from the environment using a single CNN.

The designed CNN structure incorporates a pre-trained CNN, and a transfer learning method was employed to train this designed CNN. Transfer learning is a skill that people often unwittingly use to apply an acquired ability to another similar task. It has been widely used in various applications, ranging from fault detection [22] to time series forecasting [23]. Li et al. used transfer learning to classify text data [24]. In another study, transfer learning has been utilized in order to process hyperspectral image [25]. Moreover, a robot has been

successfully developed to detect damaged ropes on bridges using transfer learning [26]. Similarly, in another project, a robot employed transfer learning to distinguish objects from underground images [27]. In the current study, eight different pre-trained CNN architecture was trained for pre-defined task by using transfer learning and CNN structure that was obtained best result was chosen.

In the current study, a robot software was developed to assist in welding and sealing processes based on HRI. The developed software intended for use in small scale plants with non-mass production lines. Firstly, the desired task and a trajectory were defined according to the operator's hand gestures and positions of the fingertip. Next, the relationship between the defined trajectory and the metal object was searched. Finally, we implemented the obtained trajectory onto the robot in the simulation environment. As a result, the user could command the robot to perform the desired task without the need for manual programming. This study offers a user-friendly approach, allowing the robot to perform similar tasks without requiring any specific knowledge of robotics.

## II. MATERIAL AND METHODS

### A. Structure of Developed Software

Developed software consists of CNN, process selection-trajectory generation, and trajectory regulation modules. The modules of developed software and data traffic between modules are given in Figure 2.

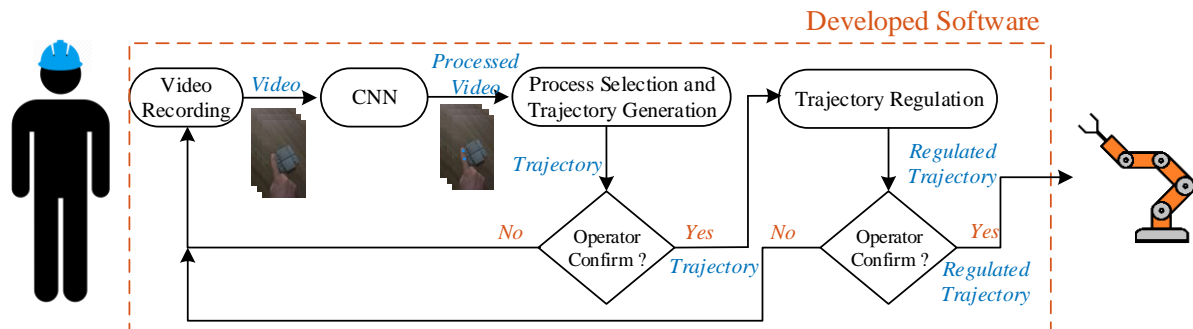


Fig. 2. Block diagram of developed software

A camera was used to be the robot aware of its environment and perceive the desired process of the robot. The camera's task was to record the hand movement of the operator in the robot's environment. The operator starts the video recording process before the operation, and the video recording is stopped by the operator when the defined process demonstration is finished. After the recorded video is separated into images, the images are sent to the CNN structure. CNN classifies the images according to hand gestures. If classifying result is calculated as index finger, middle joint, and fingernail positions of the index finger is produced by CNN. In the process selection and trajectory generation module, the type and start-end times of the operation such as welding or sealing are determined from hand motions. The index finger class between the start and end times determines the trajectory of the process. Then, the obtained process and trajectory are submitted for operator approval. If the operator does not confirm the process or trajectory, the

program returns to the video recording stage. If the operator confirms the action and the trajectory, the trajectory regulation is applied to the trajectory. After the regulated trajectory is confirmed by the operator again, the determined process is carried out by the KUKA KR Agilus KR6 R900 sixx robot located in the simulation environment along the arranged trajectory. After the process is complete, the program returns to the first step.

#### 1) CNN Module

Training of CNN, which is a part of developed software, consist of three steps as dataset forming, train dataset augmenting, and model training.

Fifty-three videos with a total size of 1.54GB were recorded in the experimental environment in order to create a dataset. Forty-five of these videos were used for the training and validation dataset. Rest of these videos were used for the test



dataset. Total 8000 images, that were 180×320px size, were obtained from training and validation videos. 512 images, that were according to homogeneous each class, were randomly separated from this dataset for the validation dataset. Rest of these images was used as training dataset. Total 128 images, that were 180×320px size, were obtained from test videos to generate test dataset. Formed datasets were contained of four class as *Zero*, *One*, *Two*, and *Three*. Each class label was typified finger count of hand gestures. The data contained in *One* labelled class has 4 location information ( $x_j, y_j$  and  $x_T, y_T$ ), as well as class labels.  $J$  and  $T$  letters were symbolized joint and fingernail of index finger, respectively. Class label and position data of datasets was given in Figure 3.

Data augmentation is an operation that artificial images, which are generated from training dataset images, are incorporated into the training dataset to increase the performance of CNN. The images were rotated 180° and the brightness of the rotated images were modified ratio of ±25%

in order to increase the variety of images in the training data set. After these processes, obtained artificial images was added in the training dataset.

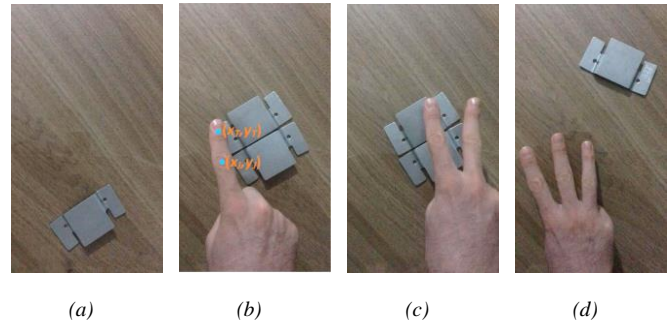


Fig. 3. Classes in the datasets; (a) Zero, (b) One, (c) Two, (d) Three

After the data sets were created, CNN structure was trained using the block diagram in Figure 4.

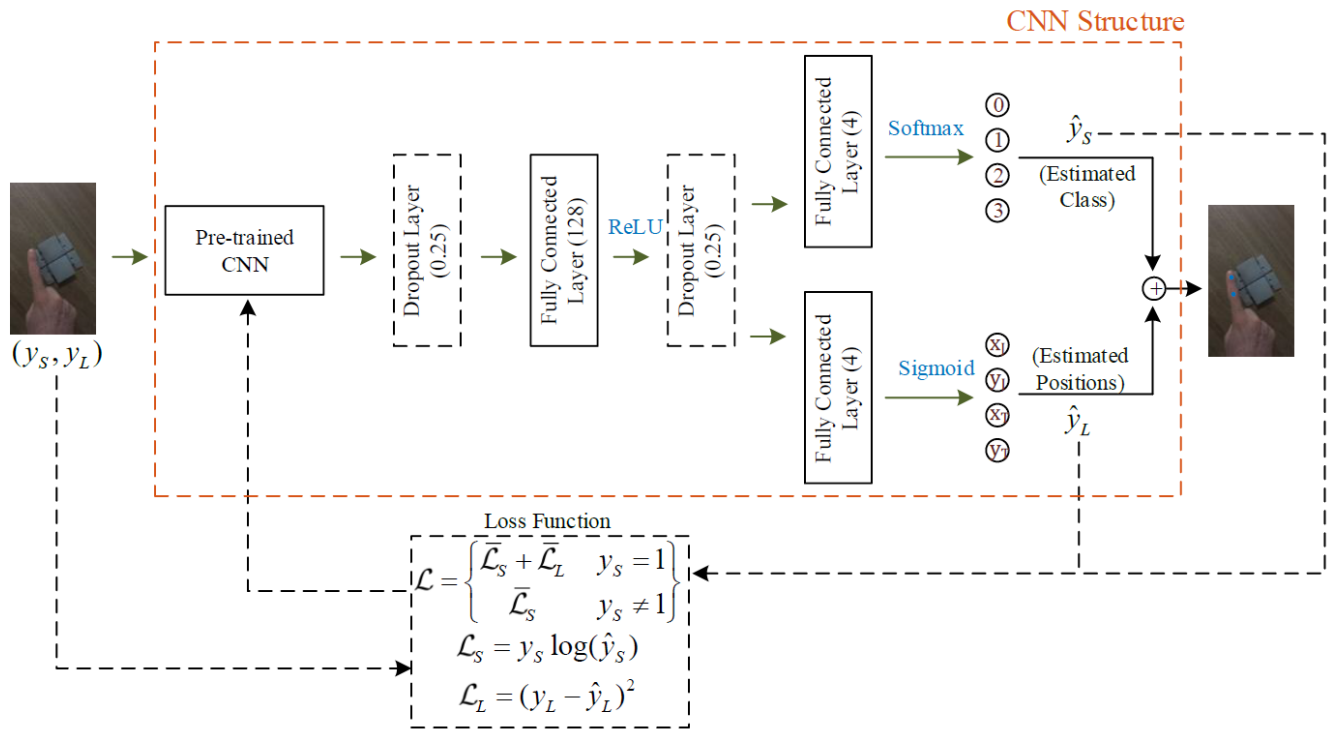


Fig. 4. Block diagram of CNN structure

$y_s$  and  $y_L$  were shown class label and positions data of image in Figure 4, respectively. Eight different CNN model was tried as pre-trained CNN.  $\hat{y}_s$  and  $\hat{y}_L$  were symbolized predicted class and position data of CNN structure, respectively. Total, classification, and localization loss functions were presented as  $L, L_s$ , and  $L_L$  symbols, respectively. Cross-Entropy loss function was used as classification loss function and squared error was utilized as localization loss function. If the class of the input image is One, the total loss function is calculated by using a formula that sums the mean of  $L_s$  and  $L_L$ . Otherwise, the total loss function equals the mean of the classification loss function. The black dashed lines in Figure 4 show the blocks used only during the training of the CNN architecture. The dropout layer is also a structure consisting of dashed lines. The dropout layer

ratio was chosen as 0.25. The black solid-lined blocks show the blocks used in both training and testing phases. In the fully connected layer, one of the black solid-lined blocks, there are 128, 4, and 4 neurons, respectively. ReLU, Softmax, and Sigmoid are activation functions found in network outputs. Detailed information about the functions of the blocks was mentioned in [28]–[30]. Also, the CNN structure was trained during 10 epochs by using the Adam optimization algorithm. The learning rate was chosen as  $10^{-3}$  in the first 5 epochs and the learning rate was adjusted as  $10^{-4}$  for the rest of the training process. Mini-batch size was chosen as 32. Weights of the pre-trained CNNs were set up as ImageNet and update of pre-trained CNNs weights were continued during the training

process. The training process was shortened by using transfer learning.

2) *Process Selection and Trajectory Generation Module*

Process selection and trajectory generation were realized in this part of developed software by depending on class and position data obtained from CNN structure. Process selection operation was performed before trajectory generation. Firstly, noised class labels need to remove from obtained class labels to choose the process. The noised labels are formed by blurred images that occur when the operator’s hand enters or quits on video. It is inspired by the exponential weight averages formula presented in Equation 1 to eliminate noised labels.

$$V_k = \beta V_{k-1} + (1 - \beta)Q_k, k = 2,3, \dots, N \quad (1)$$

In Equation 1,  $V, \beta, Q, N$ , and  $k$  refer to mean value, mean coefficient, current measure, total sample size, and discrete time index, respectively. How many samples will be averaged with  $\beta$  is determined by using as flows:

$$T_s = \frac{1}{1 - \beta} \quad (2)$$

Total sample size is shown as  $T_s$  in Equation 2. Equation 1 was not used because the class output of CNN wasn’t numerical value. Also, bias coefficient was not used owing to the same reason. Equation 3 that formed by inspiring Equation 1 was used to filter class labels.

$$fL_k = \begin{cases} Q_k & k \leq T_s \\ \text{mod}(fL_{k-T_s}, fL_{k-T_s+1}, \dots, fL_{k-1}, Q_k) & k > T_s \end{cases} \quad (3)$$

$fL$  is typified filtered label data in Equation 3. Raw and filtered label data that belong to the sealing process were given in Figure 5.

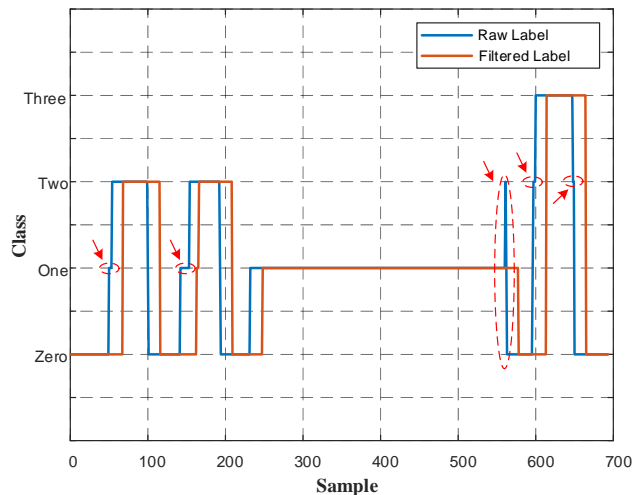


Fig. 5. Label data for sealing process

The red arrows in Figure 5 were shown noised label data. The detected noised labels were eliminated by using Equation 3. In the current study,  $\beta$  coefficient was chosen as 0.97.  $\beta$  coefficient is range of between 0 and 1 as can be understood from Equation 2. Also, while the noise increases as the  $\beta$  approaches 0, the inertia of the system increases as it approaches 1.

After filtering the label data, the process is determined by using the flowchart presented in Figure 6.

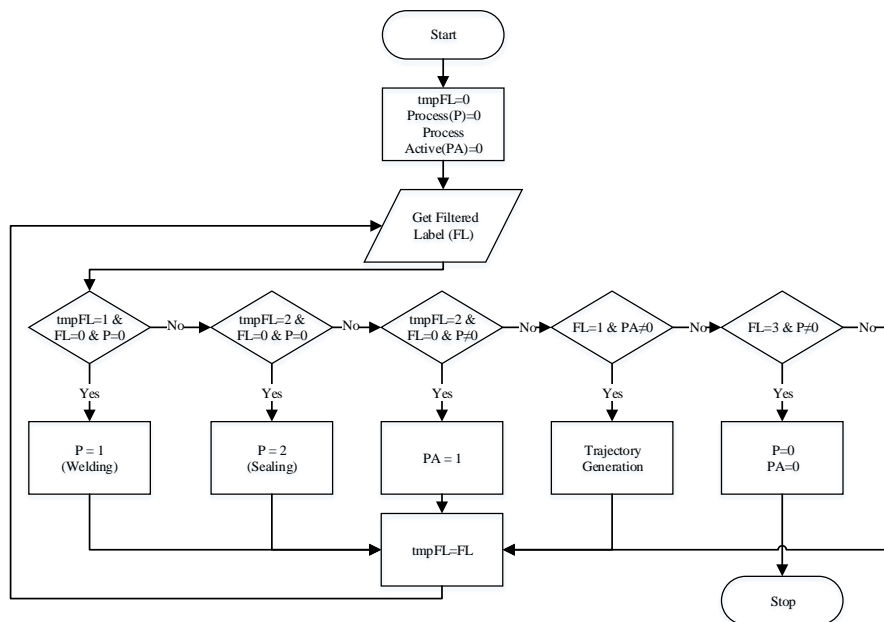


Fig. 6. Process determination flowchart

Firstly, the process is determined in Figure 6. After the defined process is activated, the trajectory generation is started when the filtered label is *One*. The trajectory generation process is continued until the filtered label is *Three*. The position and

orientation of the index finger are calculated by using Equation 4-7 and the joint and fingernail position of the index finger. The reason for calculating the positions of the joint and fingernail of the index finger is that the operator can also determine the

orientation of the manipulator in orientation-dependent operations such as welding.

$$\alpha = \tan^{-1}\left(\frac{y_T - y_J}{x_T - x_J}\right) \quad (4)$$

$$d = \sqrt{(x_T - x_J)^2 + (y_T - y_J)^2} \quad (5)$$

$$x_F = x_T + \frac{d}{4} \sin(\alpha) \quad (6)$$

$$y_F = y_T + \frac{d}{4} \cos(\alpha) \quad (7)$$

$\alpha$  refers to the orientation angle of the index finger in Equation 4.  $d$  represents the distance between the joint and fingernail of the index finger in Equation 5.  $(x_F, y_F)$  are typified fingertip positions of the index finger on X and Y axes, respectively. The hand quickly moves towards the metal object during the trajectory generation process. The distance between two points was calculated using Equation 5 with finger position information obtained from two sequential images in order not to create a trajectory during this orientation process. If the calculated distance is lower than 7px, obtained  $(x_F, y_F)$  was included trajectory. Otherwise,  $(x_F, y_F)$  was not incorporated trajectory.

$$tx_k = \delta x_{F_k} + \overline{\emptyset(x_{F_{k-1}}, x_{F_{k-2}}, \dots, x_{F_{k-N}})} \quad (8)$$

$$ty_k = \delta y_{F_k} + \overline{\emptyset(y_{F_{k-1}}, y_{F_{k-2}}, \dots, y_{F_{k-N}})} \quad (9)$$

Trajectory of  $trj_{2 \times k} = [tx, ty]$  was obtained utilizing Equation 8 and 9.  $tx$  and  $ty$  refer to trajectory position on X and Y axes, respectively.  $\delta, \emptyset$ , and  $N$  represent last measure coefficient, mean coefficient, and count of elements to be averaged, respectively. These value of the coefficient were chosen as 0.5, 0.5, and 5, respectively. Noises on the trajectory were partially cleared by using Equations 8 and 9.

### 3) Trajectory Regulation Module

Trajectory regulation module was formed to establish relationship between generated trajectory and metal object and decrease noise on trajectory. An image that is not consist operator's hand was taken from the video to regulate trajectory. Initially, an image without operator's hand was taken from the video to regulate trajectory. The image was converted greyscale image and Sobel filter was applied to the greyscale image. Edges of object that is in the image was roughly calculated with the method as can be seen in Figure 7b. After this step, section of object in image was cropped by help of object edges. The cropped image was converted to grayscale image and blurred, respectively. Lastly, Sobel filter was applied on the blurred image and Figure 7d, that shows more clearly edges of object, was obtained. Blur filter was not used in first step because undesired edges were occurred in image.

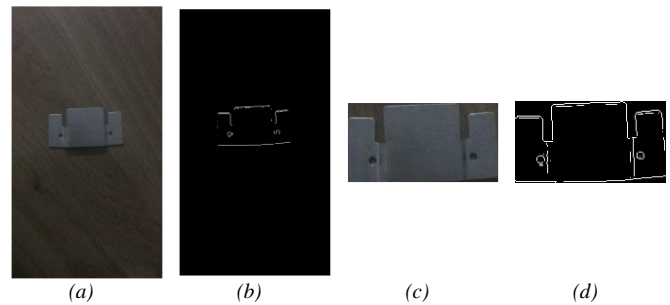


Fig. 7. Obtaining edge images process; (a) Original image, (b) Rough edge image, (c) Cropped image, (d) Edge image

Distance between each of the edge points in the obtained edge image and each of the  $trj$  points was measured by using Equation 5.  $trj$  points that were 15px away from edge points were removed from the trajectory after the measured distances.  $mtrj$  trajectory that was not contained in the 15px away points was formed. Finally, the trajectory regulating process was carried out by applying a  $20 \times 1$  dimensional median filter to the positions of the X and Y axis in  $mtrj$ . Figures 11 and 12 can be examined for a better understanding of the trajectory and regulated trajectory difference.

### B. Simulation of Developed Software

In this study, the KUKA KR Agilus KR6 R900 sixx robot with 6 axes and an Euler wrist was used in our laboratory. The maximum payload and reach of this robot are 6kg and 901mm respectively. Also, the position repeatability of this robot is 0.03mm. Developed software was simulated on CoppeliaSim program that is a simulation environment. Before the simulation scene was not formed, a 3D solid model of the robot was drawn on the SolidWorks program. The formed 3D solid model was converted to URDF (Unified Robotic Description Format) with URDF exporter [31]. Then the URDF file was added to the designed scene as presented in Figure 8.

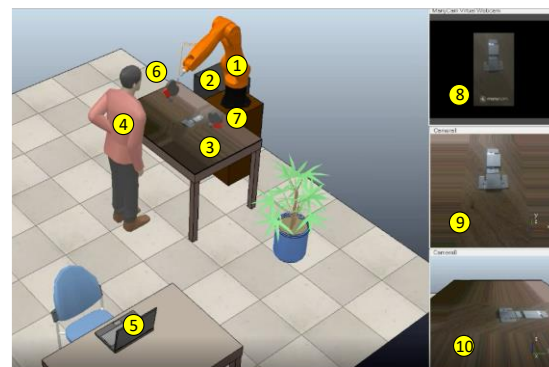


Fig. 8. Simulation environment

Manipulator (1) and robot PC (2) are components of the robot in Figure 9. Work plane, operator, and the PC that the software will run represents as (3), (4), and (5) respectively. Cameras (6-7) were added to the simulation environment to watch to the work plane from different angles. Video streams from the cameras were shown (9-10) windows, respectively. It is assumed that a camera is placed at the endpoint of the manipulator that records the hand movements of the operator and the video stream of this camera is presented on the screen (8). ManyCam program [32] was used to input video from

outside to the simulation environment while creating the screen. The manipulator in the simulation environment was moved using MATLAB package program and the inverse kinematic solution of the simulation program. The simulation program was run using the Newton dynamic engine with 50ms step size.

III. RESULTS

A part of the developed software is the CNN structure. The most important building block of this CNN architecture is pre-trained CNN structures. In the current study, CNN structure that using 8 different pre-trained CNN was trained. After the training process, training, validation, and test performance were presented in Table 1.

TABLE I  
LOSS FUNCTION VALUES

Algorithms	PS[33] (MB)	TT (s)	$L_{train}$	$L_{valid}$	$L_{test}$
ResNet50[34]	98	1030	0.712	0.429	2.396
VGG16[35]	528	2460	1.539	1.515	1.412
DenseNET121[36]	33	2120	0.961	0.886	0.709
InceptionResNetV2 [37]	215	3820	0.016	0.003	0.533
EfficientNetB0[38]	29	2570	0.031	0.092	0.407
MobileNetV2[39]	14	1930	0.019	<b>0.002</b>	0.155
InceptionV3[40]	92	1720	0.017	0.230	0.080
Xception[41]	88	2200	<b>0.011</b>	0.003	<b>0.005</b>

Bold numbers indicate the best results.

Parameter size and training time were shown as PS and TT in Table 1, respectively. The CNN training process was performed on the Google Colab platform. PC components used on this platform; GPU: Nvidia P100-16GB, CPU: Intel Xeon-2.30GHz, RAM: 25.51GB, Disk memory: 68.40GB. Since the best test result was obtained from the Xception algorithm, the loss values of Xception and other algorithms were compared by using multivariate Tukey comparison test and the Tukey test results are presented in Table 2.

TABLE II  
XCEPTION AND OTHER METHODS COMPARISON

Algorithms	$\bar{L} \pm SD$	p (According to Xception)
ResNet50	1.1795±0.8683	<0.05*
VGG16	1.4889±0.0552	<0.05*
DenseNET121	0.8522±0.1057	0.238
InceptionResNetV2	0.1846±0.2468	0.999
EfficientNetB0	0.1771±0.1649	0.999
MobileNetV2	0.0592±0.0685	1.000
InceptionV3	0.1096±0.0894	1.000
Xception	0.0067±0.0033	-

\* Statistically significant difference.  $\bar{L} \pm SD$  represents the mean of loss and standard deviation.

The Xception algorithm was found to have statistical differences with the ResNET50 and VGG16 algorithms as can be seen in Table 2. There is no statistical difference between other algorithms and the Xception algorithm. In addition, after the training process was completed, all data in the validation and test datasets were classified by the CNN architecture and the relevant classes were located. These classification and localization results are shown in Figure 9.

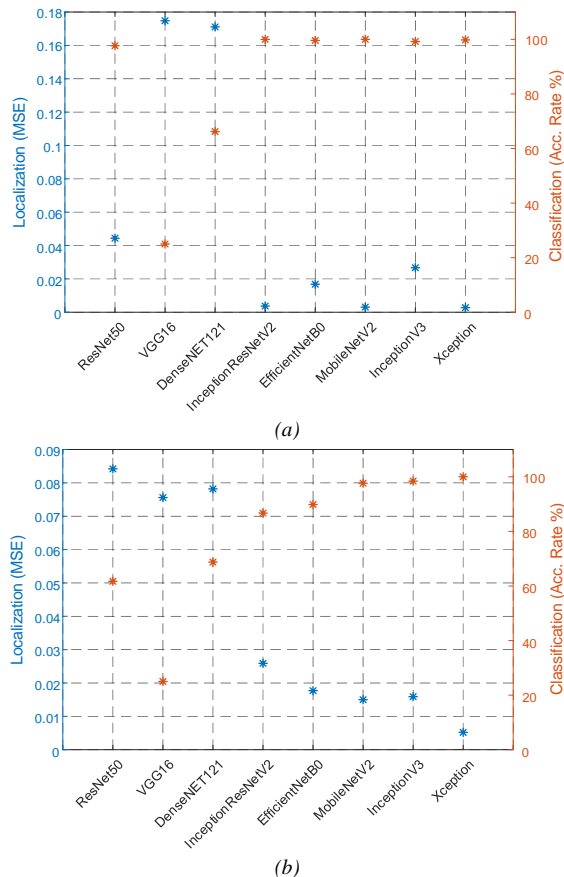


Fig. 9. Classification and localization results; (a) Results of the validation dataset. (b) Results of the test dataset

The Xception algorithm was used in the application as the best performance was obtained using the algorithm. Firstly, the operator recorded various sealing and welding process videos. These recorded videos were processed by CNN as seen in Appendix 1. Sealing process, welding process, and joint-fingernail of index finger were shown with black, red, and blue colour, respectively in Appendix 1. Then, generated trajectories were regulated and regulated trajectories were given in Figure 10 and 11.

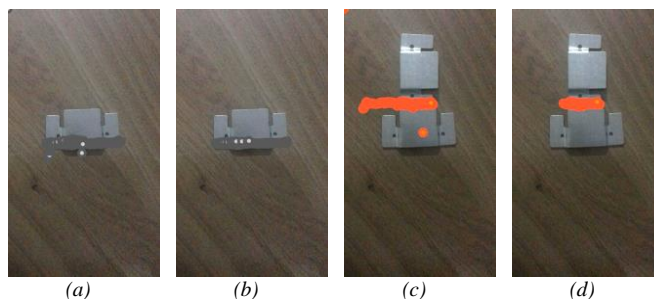


Fig. 10. Visual results of trajectory regulation; (a) Sealing process trajectory, (b) Regulated sealing process trajectory, (c) Welding process trajectory, (d) Regulated welding process trajectory



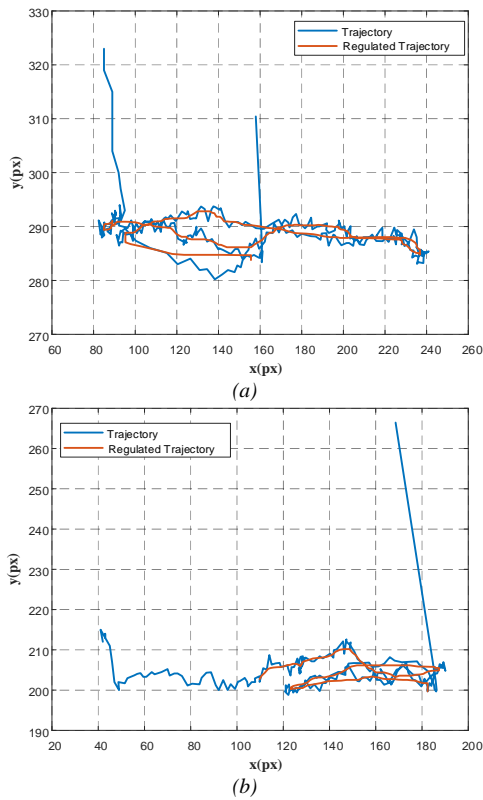


Fig. 11. Graphical results of trajectory regulation; (a) Sealing process trajectory, (b) Welding process trajectory

The regulated trajectories were sent to the robot that in the simulation environment. The desired tasks were simulated as seen in Figure 12.

Axis angles, axis moments, tool centre point (TCP) position and trajectory tracking error occurring during the sealing and welding process are presented in Figures 13 and 14.

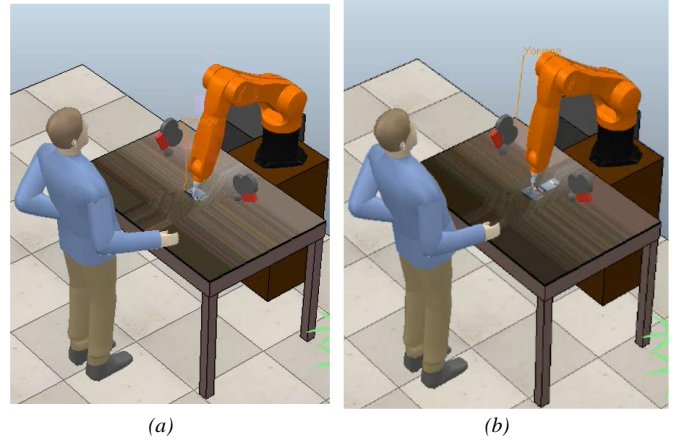


Fig. 12. Simulated desired tasks; (a) Sealing process, (b) Welding process

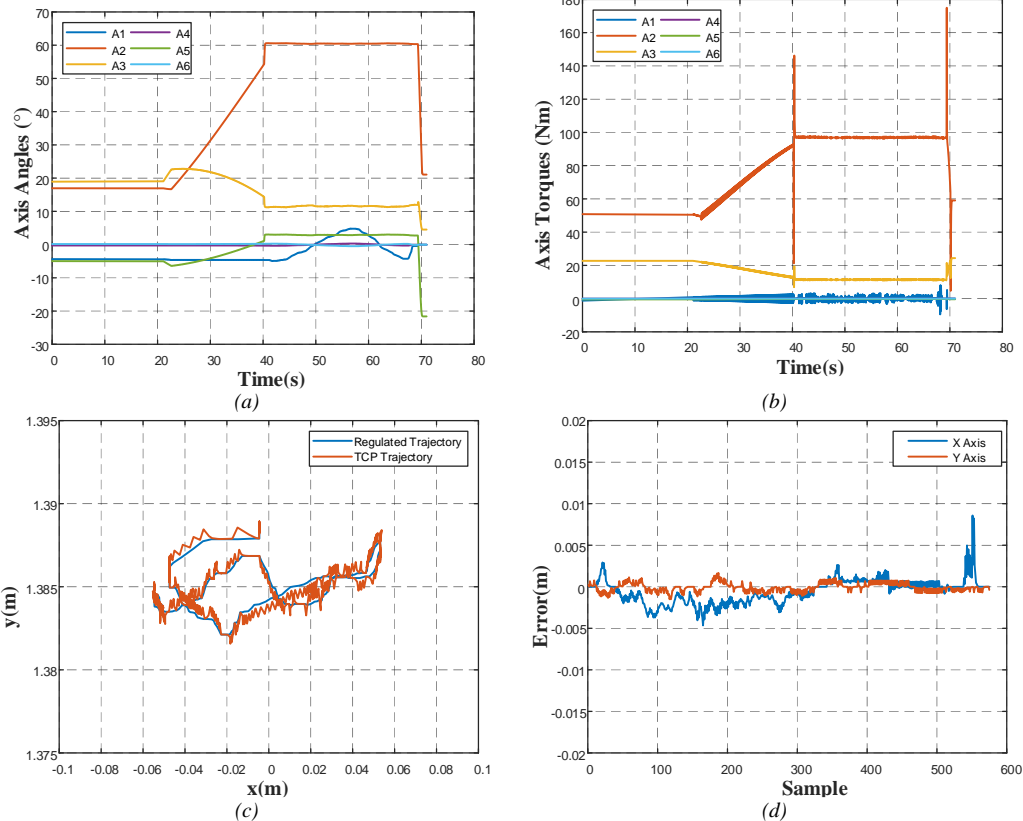


Fig. 13. Values occurring during the sealing process; (a) Axis angles, (b) Axis Moments, (c) TCP trajectory, (d) Error values during trajectory tracking



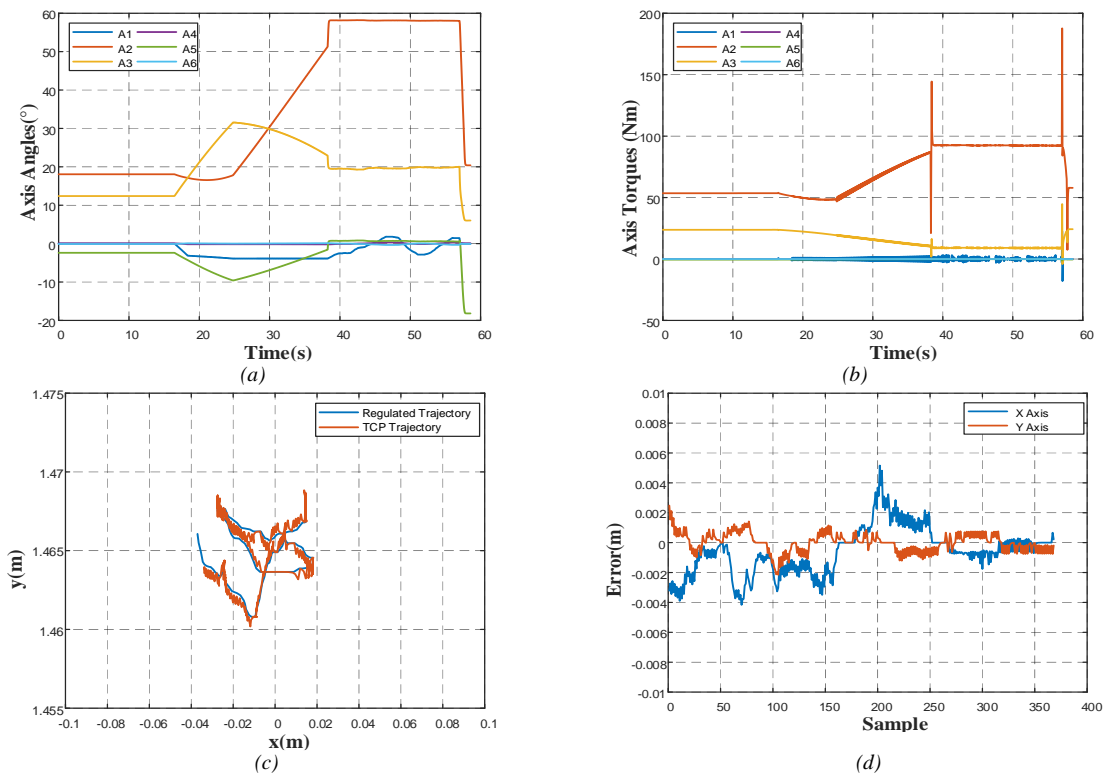


Fig. 14. Values occurring during the welding process; (a) Axis angles, (b) Axis Moments, (c) TCP trajectory, (d) Error values during trajectory tracking

After sealing and welding processes, metal object in work plane was given in Figure 15.

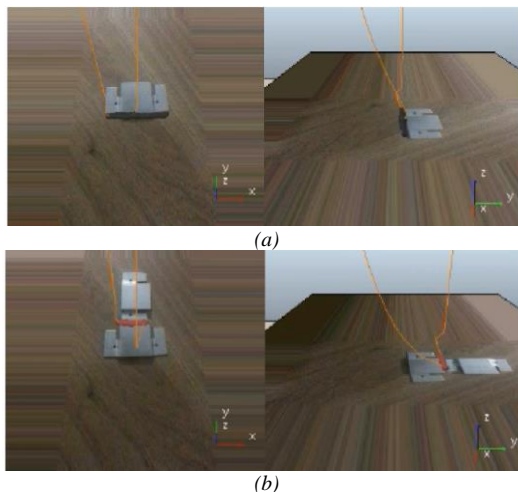


Fig. 15. Processed metal object; (a) Sealing process, (b) Welding process

The desired task by the operator was performed by the robot as it can be seen in Figure 15. The sealing process, welding process, and robot movements were given as videos in the [42] and [43], respectively.

#### IV. DISCUSSION

Fingertip location was calculated by using skin colour in some studies when studies in the literature that fingertip detection were examined [11]–[13]. Depth images that were obtained from RGB-D sensors were used to detect fingertip position in other fingertip detection studies [14]–[16]. Skin

colour and depth images were not used in this study. Also, when fingertip detection studies that were based on CNN structures were investigated, using cascade CNN structure was seen [18]. A single CNN was used as different from the study. In addition, when fingertip detection studies that were in the robotic field were researched, special sensors were developed to perceive hand gestures [19], [20]. In the current study, a standard camera was used to sense hand gestures.

In the current study, solving of classification and localization problem was implemented to hand gestures recognition and fingertip position detection. In this way, two different problems were solved with a single structure. This study has some limitations. In this study, the most important restriction of the current study is that the thicknesses of the parts to be machined were predefined and a standard depth was worked on. Another limitation is the CNN architectures used. Pre-trained CNN architectures were used to increase the accuracy performance by reducing the training time with the transfer learning method.

#### V. CONCLUSION

In this study, a robot software capable of performing processes such as sealing and welding was developed for small-scale plants without mass production capabilities. Operators without any prior robot education/knowledge can program the robot using finger movements through the developed robot software. This programmability capability was achieved through the integration of the CNN, process selection-trajectory generation, and trajectory regulation modules. The CNN structure consisted of a pre-trained CNN, fully connected layers, and activation functions connected in series. Eight pre-trained CNNs were trained on formed datasets and subsequently tested, with the Xception algorithm yielding the best result ( $L_{test}=0.0051$ ). The CNN structure was used to

classify image data and determine the positions of the robot's joints and the index finger's fingernail. With the classification data, the process selection and trajectory generation module detected the desired task, and the same module created the trajectory based on the positions data. Furthermore, a special algorithm was developed within the process selection and trajectory generation module to reduce any noise that may occur during video processing. The generated trajectory was then regulated by the trajectory regulation module to ensure proper alignment with the objects. Following this step, the robot performed the desired process within the simulation environment. In future work, an additional module will be developed to predict trajectories based on the objects and will be incorporated into the software. Subsequently, the software will undergo testing on a real robot. In addition, the developed software will become more improved by using other deep learning architectures such as LSTM.

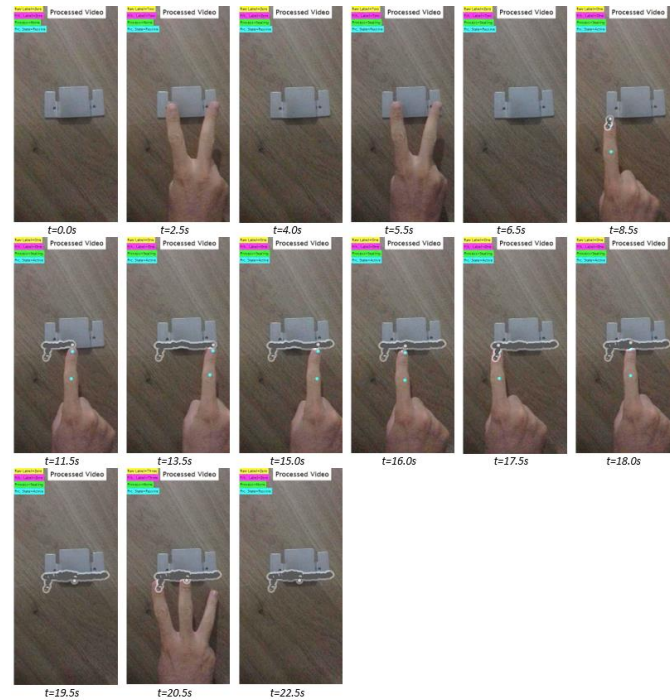
## REFERENCES

- [1] A. Dobra, "General classification of robots. Size criteria," in *2014 23rd International Conference on Robotics in Alpe-Adria-Danube Region (RAAD)*, IEEE, 2014, pp. 1–6.
- [2] "Defining The Industrial Robot Industry and All It Entails." <https://www.robotics.org/robotics/industrial-robot-industry-and-all-it-entails> (accessed Sep. 28, 2020).
- [3] IFR, *World Robotics 2022*. 2022. [Online]. Available: [https://ifr.org/downloads/press2018/2022\\_WR\\_extended\\_version.pdf](https://ifr.org/downloads/press2018/2022_WR_extended_version.pdf)
- [4] S. M. M. Rahman, Z. Liao, L. Jiang, and Y. Wang, "A regret-based autonomy allocation scheme for human-robot shared vision systems in collaborative assembly in manufacturing," in *IEEE International Conference on Automation Science and Engineering*, 2016, pp. 897–902. doi: 10.1109/COASE.2016.7743497.
- [5] H. Ding, M. Schipper, and B. Matthias, "Collaborative behavior design of industrial robots for multiple human-robot collaboration," in *2013 44th International Symposium on Robotics, ISR 2013*, 2013. doi: 10.1109/ISR.2013.6695707.
- [6] S. M. M. Rahman, Y. Wang, I. D. Walker, L. Mears, R. Pak, and S. Remy, "Trust-based compliant robot-human handovers of payloads in collaborative assembly in flexible manufacturing," in *IEEE International Conference on Automation Science and Engineering*, 2016, pp. 355–360. doi: 10.1109/COASE.2016.7743428.
- [7] T. Hamabe, H. Goto, and J. Miura, "A programming by demonstration system for human-robot collaborative assembly tasks," in *2015 IEEE International Conference on Robotics and Biomimetics, IEEE-ROBIO 2015*, 2015, pp. 1195–1201. doi: 10.1109/ROBIO.2015.7418934.
- [8] H. Ding, J. Heyn, B. Matthias, and H. Staab, "Structured collaborative behavior of industrial robots in mixed human-robot environments," in *IEEE International Conference on Automation Science and Engineering*, 2013, pp. 1101–1106. doi: 10.1109/CoASE.2013.6653962.
- [9] J. L. Raheja, K. Das, and A. Chaudhary, "Fingertip Detection: A Fast Method with Natural Hand," *Int. J. Embed. Syst. Comput. Eng. Local Copy*, vol. 3, no. 2, pp. 85–88, 2012, [Online]. Available: <http://arxiv.org/abs/1212.0134>
- [10] S. Mukherjee, S. A. Ahmed, D. P. Dogra, S. Kar, and P. P. Roy, "Fingertip detection and tracking for recognition of air-writing in videos," *Expert Syst. Appl.*, vol. 136, pp. 217–229, 2019, doi: 10.1016/j.eswa.2019.06.034.
- [11] S. K. Kang, M. Y. Nam, and P. K. Rhee, "Color based hand and finger detection technology for user interaction," in *2008 International Conference on Convergence and Hybrid Information Technology, ICHIT 2008*, 2008, pp. 229–236. doi: 10.1109/ICHIT.2008.292.
- [12] G. Wu and W. Kang, "Vision-Based Fingertip Tracking Utilizing Curvature Points Clustering and Hash Model Representation," *IEEE Trans. Multimed.*, vol. 19, no. 8, pp. 1730–1741, 2017, doi: 10.1109/TMM.2017.2691538.
- [13] G. Wu and W. Kang, "Robust Fingertip Detection in a Complex Environment," *IEEE Trans. Multimed.*, vol. 18, no. 6, pp. 978–987, 2016, doi: 10.1109/TMM.2016.2545401.
- [14] J. Yang, X. Ma, Y. Sun, and X. Lin, "LPPM-Net: Local-aware point processing module based 3D hand pose estimation for point cloud," *Signal Processing: Image Communication*, vol. 90, p. 116036, 2021. doi: 10.1016/j.image.2020.116036.
- [15] C. Wang, Z. Liu, M. Zhu, J. Zhao, and S. C. Chan, "A hand gesture recognition system based on canonical superpixel-graph," *Signal Processing: Image Communication*, vol. 58, pp. 87–98, 2017. doi: 10.1016/j.image.2017.06.015.
- [16] J. Shin and C. M. Kim, "Non-Touch Character Input System Based on Hand Tapping Gestures Using Kinect Sensor," *IEEE Access*, vol. 5, pp. 10496–10505, 2017, doi: 10.1109/ACCESS.2017.2703783.
- [17] J. L. Raheja, A. Chaudhary, and K. Singal, "Tracking of fingertips and centers of palm using KINECT," in *Proceedings - CIMSIm 2011: 3rd International Conference on Computational Intelligence, Modelling and Simulation*, 2011, pp. 248–252. doi: 10.1109/CIMSIm.2011.51.
- [18] Y. Huang, X. Liu, L. Jin, and X. Zhang, "DeepFinger: A Cascade Convolutional Neuron Network Approach to Finger Key Point Detection in Egocentric Vision with Mobile Camera," in *2015 IEEE International Conference on Systems, Man, and Cybernetics, SMC 2015*, 2016, pp. 2944–2949. doi: 10.1109/SMC.2015.512.
- [19] F. Chen *et al.*, "WristCam: A Wearable Sensor for Hand Trajectory Gesture Recognition and Intelligent Human-Robot Interaction," *IEEE Sens. J.*, vol. 19, no. 19, pp. 8441–8451, 2019, doi: 10.1109/JSEN.2018.2877978.
- [20] G. Shi, C. S. Chan, W. J. Li, K. S. Leung, Y. Zou, and Y. Jin, "Mobile human airbag system for fall protection using mems sensors and embedded SVM classifier," *IEEE Sens. J.*, vol. 9, no. 5, pp. 495–503, 2009, doi: 10.1109/JSEN.2008.2012212.
- [21] L. Peternel, N. Tsagarakis, and A. Ajoudani, "A human-robot co-manipulation approach based on human sensorimotor information," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 7, pp. 811–822, 2017, doi: 10.1109/TNSRE.2017.2694553.
- [22] C. Li, S. Zhang, Y. Qin, and E. Estupinan, "A systematic review of deep transfer learning for machinery fault diagnosis," *Neurocomputing*, vol. 407, pp. 121–135, 2020, doi: 10.1016/j.neucom.2020.04.045.
- [23] R. Ye and Q. Dai, "Implementing transfer learning across different datasets for time series forecasting," *Pattern Recognit.*, vol. 109, 2021, doi: 10.1016/j.patcog.2020.107617.
- [24] Z. Li, B. Liu, and Y. Xiao, "Cluster and dynamic-TrAdaBoost-based transfer learning for text classification," in *ICNC-FSKD 2017 - 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery*, 2018, pp. 2291–2295. doi: 10.1109/FSKD.2017.8393128.
- [25] S. Mei, X. Liu, G. Zhang, and Q. Du, "Sensor-specific Transfer Learning for Hyperspectral Image Processing," in *2019 10th International Workshop on the Analysis of Multitemporal Remote Sensing Images, MultiTemp 2019*, 2019. doi: 10.1109/MultiTemp.2019.8866896.
- [26] S. Hou, B. Dong, H. Wang, and G. Wu, "Inspection of surface defects on stay cables using a robot and transfer learning," *Autom. Constr.*, vol. 119, 2020, doi: 10.1016/j.autcon.2020.103382.
- [27] G. A. Atkinson, W. Zhang, M. F. Hansen, M. L. Holloway, and A. A. Napier, "Image segmentation of underfloor scenes using a mask regions convolutional neural network with two-stage transfer learning," *Autom. Constr.*, vol. 113, 2020, doi: 10.1016/j.autcon.2020.103118.
- [28] M. C. Bingol and O. Aydogmus, "Practical application of a safe human-robot interaction software," *Ind. Rob.*, vol. 47, no. 3, pp. 359–368, 2020, doi: 10.1108/IR-09-2019-0180.
- [29] M. C. Bingol and O. Aydogmus, "Performing predefined tasks using the human-robot interaction on speech recognition for an industrial robot," *Eng. Appl. Artif. Intell.*, vol. 95, 2020, doi: 10.1016/j.engappai.2020.103903.
- [30] M. C. Bingol and Ö. Aydoğmuş, "İnsan-Robot Etkileşiminde İnsan Güvenliği için Çok Kanallı İletişim Kullanarak Evrişimli Sinir Ağı Tabanlı Bir Yazılımının Geliştirilmesi ve Uygulaması," *Fırat Üniversitesi Müh. Bil. Derg.*, vol. 31, no. 2, pp. 489–495, 2019, doi: 10.35234/fumbd.557590.
- [31] OSRF, "SolidWorks to URDF Exporter," 2020. [http://wiki.ros.org/sw\\_urdf\\_exporter](http://wiki.ros.org/sw_urdf_exporter) (accessed Oct. 17, 2020).
- [32] "ManyCam Main Page." <https://manycam.com/> (accessed Oct. 17,

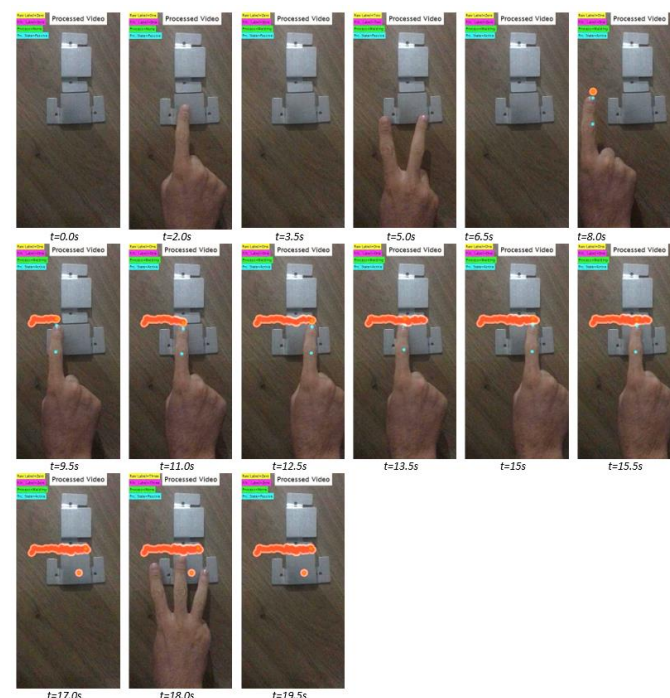
- 2020).
- [33] TensorFlow, "Keras Applications," 2020. <https://keras.io/api/applications/%0A> (accessed Oct. 15, 2020).
- [34] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," 2015.
- [35] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," 2014.
- [36] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks," 2016.
- [37] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning," 2016.
- [38] M. Tan and Q. V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," 2019.
- [39] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," 2018.
- [40] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," 2015.
- [41] F. Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions," 2016.
- [42] M. C. Bingol and O. Aydogmus, "Sealing Process," 2020. [https://drive.google.com/file/d/1BbK\\_wv2Q76ltVLSbViC1Vb7YpxIN-tEa/view?usp=sharing](https://drive.google.com/file/d/1BbK_wv2Q76ltVLSbViC1Vb7YpxIN-tEa/view?usp=sharing) (accessed Nov. 16, 2020).
- [43] M. C. Bingol and O. Aydogmus, "Welding Process," 2020. <https://drive.google.com/file/d/1Y5jE8uPoiJMKLEHeG-wj1OqVRs1bO48-/view?usp=sharing> (accessed Nov. 16, 2020).

## APPENDICES

Appendix 1. Processing video images, (a) Sealing process, (b) Welding process



(a)



(b)

## BIOGRAPHIES



Burdur, Turkey. He works in the field of robotics.

**MUSTAFA CAN BINGOL** received the B.S. (2014), M.S. (2016), and Ph.D. (2021) degree in mechatronic engineering from the University of Firat. Since 2023, He has been an Assistant Professor in Department of Electrical-Electronic Engineering, Faculty of Engineering-Architecture, Burdur Mehmet Akif Ersoy University, Burdur, Turkey. He works in the field of robotics.



**OMUR AYDOGMUS** received the B.S. (2000), M.S. (2004), and Ph.D. (2011) degree in electric-electronics engineering from the University of Firat. Since 2019, He has been a Professor in Department of Mechatronics Engineering, Faculty of Technology, Firat University, Elazig, Turkey. His research interests include electric machines and robotics.



# Vision Transformer Based Photo Capturing System

Abdulkadir Albayrak

**Abstract**—Portrait photo is one of the most crucial documents that many people need for official transactions in many public and private organizations. Despite the developing technologies and high resolution imaging devices, people need such photographer offices to fulfil their needs to take photos. In this study, a Photo Capturing System has been developed to provide infrastructure for web and mobile applications. After the system detects the person's face, facial orientation and facial expression, it automatically takes a photo and sends it to a graphical user interface developed for this purpose. Then, with the help of the user interface of the photo taken by the system, it is automatically printed out. The proposed study is a unique study that uses imaging technologies, deep learning and vision transformer algorithms, which are very popular image processing techniques in several years. Within the scope of the study, face detection and facial expression recognition are performed with a success rate of close to 100% and 95.52%, respectively. In the study, the performances of Vision Transformer algorithm is compared with the state of art algorithms in facial expression recognition.

**Index Terms**—Deep learning, facial expression recognition, photo capturing system, single shot detection, vision transformer


## 1. INTRODUCTION

AT THE present time, with the rapid development of mobile technologies, most of the persons have digital photo camera and can capture high-quality photos. However there is still a lack of applications that are useful for capturing portraits for passports and other legal documents. In this study, a sophisticated real time portrait capturing system that combines hand crafted image processing techniques with state of art deep learning approaches is proposed.

The proposed system involves automatic detection of frontal face, determining the face orientation through detected landmark points and facial expression analysis. Face detection and determination of face orientation are two basic steps that should be performed for various computer vision applications. These tasks are also critical for the proposed portrait capturing system since they constitute a precondition for the subsequent facial expression analysis step.

**ABDULKADİR ALBAYRAK**, is with Department of Computer Engineering,

Dicle University, Diyarbakır, Turkey, (e-mail: [kadir.albayrak@dicle.edu.tr](mailto:kadir.albayrak@dicle.edu.tr)).

 <https://orcid.org/0000-0002-0738-871X>

Manuscript received August 18, 2023; accepted Oct 7, 2023.

DOI: [10.17694/bajece.1345993](https://doi.org/10.17694/bajece.1345993)

In order to develop a robust face detection system, Single-Shot-Multibox detector with ResNet-10 is used as a backbone architecture [1]. In the literature, face detection is generally focused on finding 68 points using the distinctive textural features of the face[2-3] Facial landmarks are found in order to localize eyes, nose, contour of the face and mouth. Landmark points are exploited for determining whether the eyes and mouth are open or closed. Histogram of Oriented Gradients (HOG) and Support Vector Machine (SVM) are employed for selecting the images with opened eyes, while the proportion of the width and height of the mouth is calculated for determining whether the mouth is opened or closed. Frontal face images with open eyes and closed mouth are than processed for facial expression analysis.

Facial expression is one of the most effective channels of human communication, and therefore, automatic facial expression analysis systems can take place in various applications related to human-computer interaction. The success of the deep learning methods in modeling complex systems lead researchers to apply various deep learning approaches in this challenging task. Convolutional Neural Networks (CNN) based systems have proven their success in facial expression recognition problem [4]. However, facial expressions have high variability due to the nature of the human. This reality has demanded improvements in CNN-based systems, which require large amounts of training data to build a reliable model. Another drawback of the CNN based models is their relatively fragile structure against variant backgrounds and head poses [5]. Lozoya et.al aimed to improve generalization of their CNN based system by learning from mixed instances taken from different databases [6]. In [7], researchers employed CNN with Rectified Adam Optimizer in order to improve generalization.

Success of the Transformers in natural language processing pave the way for attempts to adapt transformers to computer vision problems. One of the key ideas of Transformer models is being pre-trained on a large corpus and fine-tuning on the target task with a smaller dataset [8]. Naseer et.al. compared CNN and Vision Transformer networks and stated that Vision Transformers are robust to occlusion and pose variant [9]. For all that reasons, in this study, a Vision Transformer based system is used for facial expression recognition. Proposed automatic portrait capturing system selects the neutral faces for further processing.

Selected proper portrait photographs (Frontal and neutral face images with opened eyes and closed mouth) are than post-processed for cleaning small speckles on the face and then the photographs are sent to the printer according to the preferences of the users.

The main contributions of the proposed system are given below:

- At the first time in the literature, a real time portrait capturing system that combines handcrafted image processing techniques with deep learning approaches is proposed.
- Performance of Vision Transformers in facial expression recognition task is evaluated.
- A comparison between the performance of Vision Transformer and state of art methods is done in facial expression recognition.

The rest of the paper is structured as follows: The proposed system is presented in Section 2. The experimental results are discussed in Section 3. The performance of the algorithms presented in this study is discussed in Section 4. Finally, the paper is concluded with Section 5.

## 2. MATERIALS AND METHODS

### 2.1 Performing Face Detection and Finding Face Orientation

First step of the this proposed system is to focus on face region in an image. Because one of the most basic conditions that must be met in passport photos or official documents is that the face should be detected and centered symmetrically. Most of the methods suggested in the literature try to find a total of 68 points belonging to the face including eyes, chin, nose and eyebrows. In this study, Single Shot Detection (SSD) method has been applied to detect face regions. The orientation of the face was tried to be calculated by using the positions of these points according to a line that passes through the points of nose vertically. Figure 1 shows a representation of the line dividing the face exactly in half from the vertical. The distance of this line to points 0 and 16 separately is calculated. The ratio of these distances to each other should be approximately 1. The acceptable range of face orientation is set to 0.9 and 1.1, but this range can be narrowed if this value is desired to be more precise. Equation 1 expresses the distance from point 0 to point 27 shown in Figure 1 at the nose level.

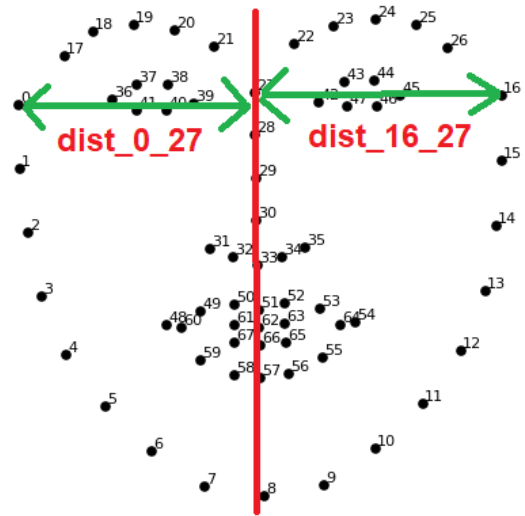
Equation 2 expresses the distance from point 16 to point 27 shown in Figure 1 at the nose level.

$$dist_{0\_27} = \sqrt{(x_0 - x_{27})^2 + (y_0 - y_{27})^2} \quad (1)$$

Here  $x_0$  and  $y_0$  represent the x and y coordinates of the point 0.  $x_{27}$  and  $y_{27}$  represent the x and y coordinate information of point 27.  $dist_{0\_27}$  shows the distance between these two points.

Equation 2 expresses the distance between point 16 in Figure 1 and point 27 at nose level.

$$dist_{16\_27} = \sqrt{(x_{16} - x_{27})^2 + (y_{16} - y_{27})^2} \quad (2)$$

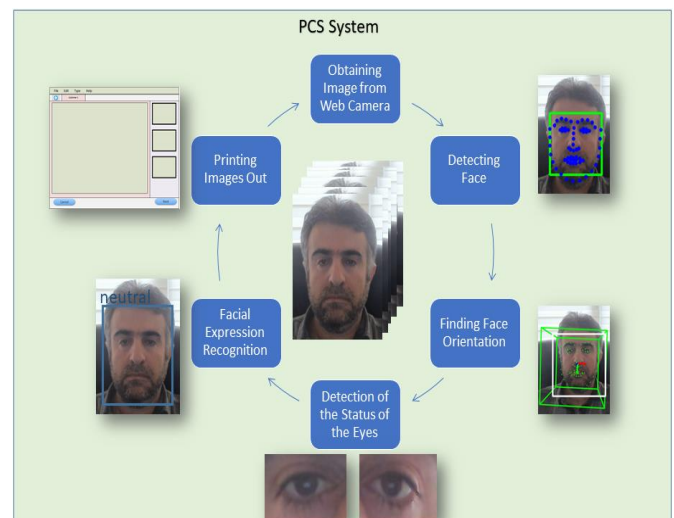


**Figure 1.** Sample image with 68 points used in face detection

Here,  $x_{16}$  and  $y_{16}$  represent the x and y coordinates of the point 16 on the front.  $x_{27}$  and  $y_{27}$  represent the x and y coordinates of the point 27 on the front.  $dist_{16\_27}$  shows the distance between these two points.

$$dist = \frac{dist_{0\_27}}{dist_{16\_27}}, \quad 0.9 \leq dist \leq 1.1 \quad (3)$$

The ideal value of the dist value obtained in Equation 3 should be 1. In this study, the value range was chosen between 0.9 and 1.1. Thus, the system is able to take pictures in small value ranges without being bound by a very strict rule. Figure 2 shows the whole flowchart of the proposed system from turning on the camera to getting output from the system.



**Figure 2.** Process steps followed in the Photo Capturing System (PCS) developed within the scope of the study

#### 2.1.1 Single Shot Detection (SSD) Deep Learning Algorithm

SSD is a feed-forward convolutional network-based deep learning method used to detect objects in images. The SSD approach makes a score estimation of the proportions of the boxes surrounding the objects that are intended to be detected.



This approach does not make an estimation about the whole image like the methods in the early studies of deep learning, but it is used to determine in which part of the image the object to be classified is located. SSD architecture consists of 3 parts (parts):

**Base Convolutions:** Networks such as VGG, ResNet, which are suggested for image classification, are the name given to the part as the base.

**Auxiliary convolutions:** It is the part that is placed as the backbone to obtain higher level features.

**Prediction Convolutions:** The attributes of the object to be detected are classified in this section. It is the part that makes predictions about the location and score of the object in the image.

The deep learning model applied for face detection within the scope of this study is the SSD model, which is trained with 300x300 image sizes and 140000 iterations. This SSD model in OpenCV's DNN library uses ResNET-10 architecture as backbone.

## 2.2 Detection of the Status of Eyes

One of the conditions that must be met in passport photos (or passport photos) is to have eyes open. No matter how accurately the eye lines are determined with facial landmark detection, there is no control such as whether the eyes are open or not. In order to carry out this process, the Histogram of Oriented Gradients (HOG) algorithm, which is one of the traditional image processing methods, was used. A total of 200 eye picture systems, 100 closed and 100 open, were trained with HOG.

### 2.2.1 Histogram of Oriented Gradients (HOG)

Histogram of Oriented Gradients was first proposed by Dalal and Triggs for pedestrian detection [10]. HOG is used to obtain shape-based features of the regions whose attributes are desired to be extracted. It has been applied in many different areas of computer vision, particularly pedestrian detection [11]. In the HOG algorithm, the orientation of all the pixels in the image is calculated and the focus is on determining the silhouette of the object (or region) that is desired to be distinguished. The basic processing steps applied in the HOG algorithm are as follows:

First, edge information is obtained by applying a sobel filter on the horizontal and vertical axis of the image. For this, the following formula referred in Equation 4 is used:

$$I_x = I * S_x, \quad I_y = I * S_y \quad (4)$$

Here,  $I$  denotes the input image,  $S_x$  the vertically applied sobel filter and  $S_y$  the horizontally applied sobel filter.  $I_x$  and  $I_y$  show the output images obtained after applying sobel vertically and horizontally, respectively. The  $I_x$  and  $I_y$  images are then used in the formula below to calculate the magnitude values.

$$|G| = \sqrt{I_x^2 + I_y^2} \quad (5)$$

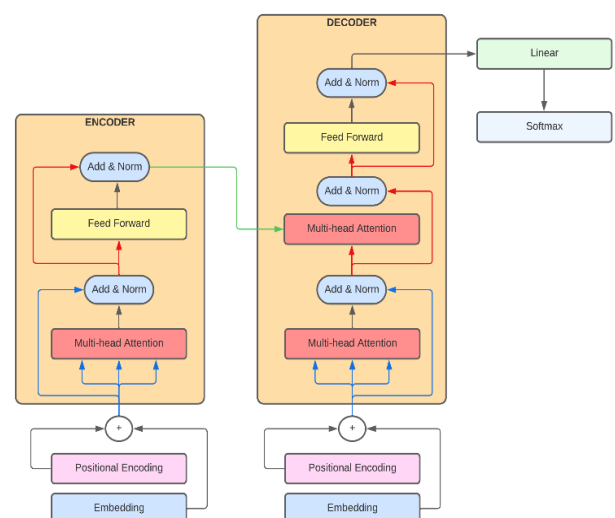
Here  $|G|$  value is expressed as a gradient and is calculated using the square root value of the sum of the squares of the  $I_x$  and  $I_y$  values specified in the previous formula. Finally, the magnitude value is  $|G|$  are obtained by calculating the arctan of the values. As a result of these operations, the orientation of the object or region is calculated and the process of distinguishing it from other objects or regions is performed.

## 2.3 Facial Expression Recognition with Vision Transformer

Neutral facial expression is one of the conditions that must be handled in portrait or passport photos. The vision transformer algorithm, which is inspired by the transformer algorithm, which has been very popular in the field of natural language processing in recent years, has been applied in automatic facial expression detection.

### 2.3.1 Vision transformer (ViT)

Transformer is a deep learning model that adopts self-attention mechanism. It can be expressed as the calculation of the relationship of each word of the data given as input in the attention mechanism with all the other words. It is primarily used in the fields of natural language processing (NLP) tasks such as machine translation and text summarization which has sequential input data. However, Transformers do not necessarily process data sequentially like Long Short Term Memory(LSTM) and Recurrent Neural Network(RNN). Instead, the attention mechanism provides context for any position in the input data. Inspired by the Transformer scaling achievements in NLP, we attempted to apply a standard Transformer directly to images with the least possible modification. To do this, we split an image into patches and provide the linear embedding order of those patches as an input to a Transformer. Image patches are treated in the same way as tokens (words) in an NLP application. We train the model on image classification in a supervised manner. Figure 3 shows the processing steps of the vision transformer method used in the proposed system for facial expression recognition.

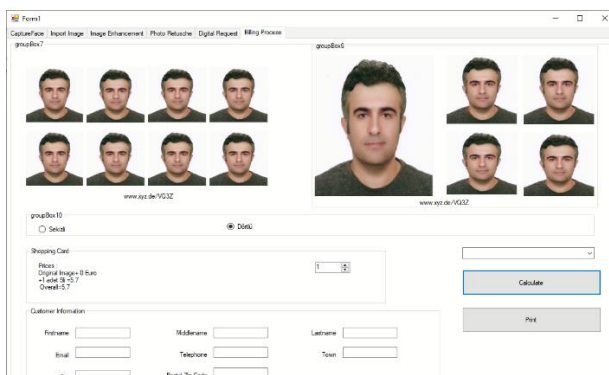


### Figure 3. Vision Transformer

The classification header is implemented by an MLP with a hidden layer and a single linear layer of fine tuning at pre-training time. The MLP consists of two fully connected layers with an occasional GeLU nonlinear activation. The most effective part to highlight in the Transformer model is the attention mechanism. The attention mechanism looks at the input sequence and selects parts of the sequence that remain important at each step, preserving knowledge of which parts of the sequence are important. The attention mechanism instantly takes into account several other input data and assigns different weights to these inputs, providing prioritization. All encoder layers use an attention mechanism for each input that measures the fitness of all other inputs and retrieves the appropriate information to produce the output. Then, as a result of the attention mechanism, it takes the weights sent as output and the encoded string as input. These networks consist of repeated multi-headed attention blocks and feedforward layers. Multi-headed attention runs the processes in the attention mechanism in parallel and combines the results. Thus, it is ensured that different relationships are learned.

### 2.4 Graphical User Interface (GUI)

After the image processing steps were completed, a design that could be output to the printer was realized with the help of the user interface developed for the users. Face detection, face orientation, eye and mouth opening, facial expression recognition operations are performed in "CaptureFace" option which located in the top menu in GUI. When all the determined rules are fulfilled, 10 pictures are saved to a folder named for a specific user/customer. One of the captured image is given as input to the system using the GUI. Then, possible noises are cleaned by applying 3x3 median filter. Finally, after the user is asked whether he wants a digital copy, the interface page where the information is entered comes. On this page, the images is sent to the printer after the user fills all the contact information. Photos are printed from the printer by clicking the Confirm button. Figure 4 shows the graphical user interface (GUI) developed to proceed further processes and to print out the final images after billing.



**Figure 4.** Graphical user interface (GUI) developed to proceed further processes and to print out the final image.

### 3. EXPERIMENTAL RESULTS

In this part of the study, face detection, recognition of facial expression, openness of the eyes and openness of the mouth were analyzed. While the publicly available facial expression recognition data set available in the literature was used within the scope of the study, the system performance was tried to be increased with a data set created within the scope of the study for eye opening/closed status.

CK+ data set for facial expression recognition: The CK+(Cohn and Kanade) is a publicly available data set for facial expression recognition [12]. There are 7 classes in total in the data set: neutral, happy, disgusted, surprised, sad, angry and afraid. The data set includes 593 sequences from 123 individuals. These sequences begin with a neutral facial expression and end with the expression belonging to each class. Figure 5 shows sample facial expressions from the data set.



**Figure 5.** Sample images obtained from the CK+ dataset used for facial expression recognition within the scope of the study. The system is required to take pictures with neutral facial expressions.

Eye Status (ES) Data set: The points around the eyes obtained in the detection of facial regions are not sufficient to understand whether the eye is open or closed alone. In order to solve this problem, another data set was created within the scope of the study to evaluate the open/closed status of the eyes in the frames obtained from the camera (See Figure 6). A data set consisting of a total of 280 images, including 140 open-eye image sections and 140 closed-eye image sections, was created.



**Figure 6.** (a), (b) represent the close eyes and (c), (d) represents the opened eye image crops obtained from the data sets.

**Evaluation:** In order for the proposed system to be used successfully in real time, face detection must be performed with high success in the first stage. For this, the SSD network, which is frequently used in object detection in the literature, is used. At this stage of the study, the facial regions in all of the randomly shot sample videos were detected with SSD. Since the ambient lighting must be very good in passport photos, it is seen that the facial regions are easily detected. In all the trials conducted within the scope of the study, it was observed that face detection was detected in all images. However, face detection may not be performed when the face area is rotated up to a certain angle. This situation is not directly related to the ambient effect.

After the face detection was performed, it was tried to decide whether the image to be obtained was appropriate by taking into account the orientation of the head. An image taken with a certain angle or tilt is not valid. Therefore, in the proposed system, it is necessary to guide the user in images that do not comply with Equations 1,2 and 3. Provided that the face area is at the level of the camera, the user needs to adjust the angle and tilt of the person. Otherwise, the system will not automatically take a picture. If the user abides by the determined rules, the orientation phase will be successfully passed for the system to take a picture along with other conditions.

After the orientation phase is passed without any problems, the stage of determining the appropriate image according to the opening/closing status of the eyes and mouth is started. 140 images belonging to each class in the dataset were separated as training and test sets with 5-fold cross-validation method. A total of 28 test images belonging to each class were classified with HOG method and SVM (rbf kernel) with a success of 93%. While the open eye success rate was 96%, the detection success of the closed eye was 85%. When the picture was taken, the condition of having an eye opening ratio above 95% was accepted as successful. Since the success rate was high enough at this stage, it was not necessary to try alternative methods in order to determine the openness / closure of the eye. It has been seen that the HOG algorithm performs well in detecting the eye opening/closure state. A model was not trained for the aperture and closure of the mouth, but the points obtained by facial landmark detection were used. Since these points for the eyes did not show a significant difference, the data set was created, but the landmark points were sufficient

within the scope of the study, since the points changed more significantly according to the shape of the mouth.

Finally, the facial expression recognition phase was carried out. Since the image accepted by the official authorities must be neutral, if the facial expression is not neutral, the picture is not taken. The method used for facial expression detection is the ViT method. The achievements after applying the ViT model to the CK+ dataset are shown in TableXX. As seen in the table, the expression recognition success of the ViT network is 95.52%. In the literature, the successes obtained in the CK+ dataset using traditional image processing techniques are relatively low. It is seen that deep learning networks, which have been performing successfully in many fields recently, have carried this success over 90%. It has been observed that the ViT model applied within the scope of the study gives as successful results as other deep learning methods. When all

**Table 1.** The classification success obtained with the ViT model applied to the CK dataset and its comparison with the results obtained in the literature

Method	Accuracy(%)
3D SIFT [13]	81.35
LBP-TOP [14]	88.99
ITBN [15]	86.3
CERT [16]	87.21
MCF [17]	89.4
MSR [18]	91.4
TMS [19]	91.89
STM [20]	91.13
AUDN [21]	93.70
BDBN [22]	<b>96.7</b>
CNN [23]	92.73
<b>Proposed ViT</b>	<b>95.52</b>

these conditions are fulfilled, the system automatically takes 10 images and these images are sent to the printer with the help of the user interface prepared within the scope of the study.

#### 4. DISCUSSION

For the necessary processes in public or private institutions and organizations, taking a passport photo has an important place in real life. People often go to places specialized for this process and passport photos are taken with high resolution cameras. Today, many technological devices have cameras capable of taking high-resolution pictures. By processing the images obtained with these devices, it is possible to obtain passport images with the same quality and similar standards. In this study, an autonomous system that takes passport photos with people's own devices or systems that can be installed in public environments is designed. The system focuses on major aspects such as face detection, detection of facial regions, eye and mouth condition, and facial expression. Since deep learning methods perform quite successfully in face detection processes, they can be used in such systems. The process of finding the orientation of the face is similarly possible. It may not be possible to determine whether the eye is open or not by using the points located around the eye. By using the coordinates of the points around the eyes, the image sections with the eyes were cropped and the status of the eyes was determined more easily with the help of the model trained with shape based HOG algorithm. A similar situation could be considered for the mouth, but since it is not as solid as the eye,



the opening and closure of the mouth could be determined. Facial expression is very important in the process of taking passport photos. The CK+ dataset, which is one of the frequently used facial expression recognition datasets in the literature, was used for model training in this study. Since the expression recognition performance with traditional image processing methods is somewhat limited compared to deep learning methods, deep learning methods can be preferred. In this study, as an alternative to deep learning methods, the ViT model, which is inspired by the transformer method, which has been very successful in natural language processing, has been used. The use of the convolution layer in training the data in deep learning methods considerably extends the training time. The absence of a training phase in ViT models enables data to be modeled in a relatively faster training period. The biggest handicap of ViT models is that they need larger resources to achieve greater success. Since the GPU used in this study has 8 GB of ram, the success is limited to 95.22%. It is thought that the success can be increased by increasing the RAM resource. Because after the image is divided into sections, the calculation of the relationship of each section with each other and with all other sections directly contributes to the success of classification.

## 5. CONCLUSION

In this study, a system that takes passport photos is proposed for use in official documents or in public and private institutions. The system is designed to combine traditional image processing methods and deep learning methods and have the capacity to perform real-time processing. In the study, the transformer method, which has recently shown a very high performance in natural language processing, has been applied for facial expression recognition. When the results obtained were evaluated, the success rates of facial detection, eye opening rate and facial expression recognition were obtained as 100%, 96% and 95.22%, respectively. The system works in real time and can take dozens of pictures depending on certain rules. Users can then select one of the saved images and take a printout with the help of a developed user interface. The system has a very important place in that it combines traditional image processing methods, deep learning methods and ViT models and works in real time. In future studies, it can be developed as a system where people can perform similar operations using their own mobile devices and the resulting image can be sent to users via mail. For this, designing it with a logic that will work on users' mobile devices will improve this system a little more.

## REFERENCES

- [1] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "Ssd: Single shot multibox detector," in European conference on computer vision. Springer, 2016, pp. 21–37.
- [2] S. Ren, X. Cao, Y. Wei, and J. Sun, "Face alignment at 3000 fps via regressing local binary features," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 1685–1692.
- [3] O. Déniz, G. Bueno, J. Salido, and F. De la Torre, "Face recognition using histograms of oriented gradients," Pattern recognition letters, vol. 32, no. 12, pp. 1598–1603, 2011.

- [4] I. M. Revina and W. S. Emmanuel, "A survey on human face expression recognition techniques," Journal of King Saud University-Computer and Information Sciences, vol. 33, no. 6, pp. 619–628, 2021.
- [5] H. Li, M. Sui, F. Zhao, Z. Zha, and F. Wu, "Mvt: Mask vision transformer for facial expression recognition in the wild," arXiv preprint arXiv:2106.04520, 2021.
- [6] S. M. González-Lozoya, J. de la Calleja, L. Pellegrin, H. J. Escalante, M. Medina, A. Benitez-Ruiz et al., "Recognition of facial expressions based on cnn features," Multimedia Tools and Applications, vol. 79, no. 19, pp. 13 987–14 007, 2020.
- [7] D. O. Melinte and L. Vladareanu, "Facial expressions recognition for human-robot interaction using deep convolutional neural networks with rectified adam optimizer," Sensors, vol. 20, no. 8, p. 2393, 2020.
- [8] S. Khan, M. Naseer, M. Hayat, S. W. Zamir, F. S. Khan, and M. Shah, "Transformers in vision: A survey," ACM Computing Surveys (CSUR), 2021.
- [9] M. M. Naseer, K. Ranasinghe, S. H. Khan, M. Hayat, F. Shahbaz Khan, and M.-H. Yang, "Intriguing properties of vision transformers," Advances in Neural Information Processing Systems, vol. 34, 2021.
- [10] P. Scovanner, S. Ali, and M. Shah, "A 3-dimensional sift descriptor and its application to action recognition," in Proceedings of the 15th ACM international conference on Multimedia, 2007, pp. 357–360.
- [11] G. Zhao and M. Pietikainen, "Dynamic texture recognition using local binary patterns with an application to facial expressions," IEEE transactions on pattern analysis and machine intelligence, vol. 29, no. 6, pp. 915–928, 2007.
- [12] Z. Wang, S. Wang, and Q. Ji, "Capturing complex spatio-temporal relations among facial muscles for facial expression recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2013, pp. 3422–3429.
- [13] G. Littlewort, J. Whitehill, T. Wu, I. Fasel, M. Frank, J. Movellan, and M. Bartlett, "The computer expression recognition toolbox (cert)," in 2011 IEEE International Conference on Automatic Face & Gesture Recognition (FG). IEEE, 2011, pp. 298–305.
- [14] S. W. Chew, S. Lucey, P. Lucey, S. Sridharan, and J. F. Conn, "Improved facial expression recognition via uni-hyperplane classification," in 2012 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2012, pp. 2554–2561.
- [15] R. Ptucha, G. Tsagkatakis, and A. Savakis, "Manifold based sparse representation for robust expression recognition without neutral subtraction," in 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops). IEEE, 2011, pp. 2136–2143.
- [16] S. Jain, C. Hu, and J. K. Aggarwal, "Facial expression recognition with temporal modeling of shapes," in 2011 IEEE international conference on computer vision workshops (ICCV workshops). IEEE, 2011, pp. 1642–1649.
- [17] M. Liu, S. Shan, R. Wang, and X. Chen, "Learning expressionlets on spatio-temporal manifold for dynamic facial expression recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2014, pp. 1749–1756.
- [18] M. Liu, S. Li, S. Shan, and X. Chen, "Au-inspired deep networks for facial expression feature learning," Neurocomputing, vol. 159, pp. 126–136, 2015.
- [19] P. Liu, S. Han, Z. Meng, and Y. Tong, "Facial expression recognition via a boosted deep belief network," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2014, pp. 1805–1812.
- [20] X. Sun, M. Lv, C. Quan, and F. Ren, "Improved facial expression recognition method based on roi deep convolutional neural network," in 2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII). IEEE, 2017, pp. 256–261.



**ABDULKADIR ALBAYRAK** currently works at the Department of Computer Engineering at Dicle University, Diyarbakir, Turkey. He received his BS, MS and PhD degrees all in Computer Engineering. He worked as a research assistant at the Department of Computer Engineering at Yildiz Technical University, Istanbul, Turkey between 2011-2019. Since January 2019, he has been working at the Department of Computer Engineering at Dicle University as a research assistant position. He mainly works on computer vision, pattern recognition and biomedical image processing.




# Finite Element Analysis of SMC Core Magnetic Gear for Vehicle Powertrain Systems

Kadir Yilmaz , Taner Dindar, Murat Ayaz, Serkan Aktas and Serkan Sezen


**Abstract**— The utilization of gears as intermediary components for power transmission in electric drive systems addresses the insufficiency of electric machines in handling torque loads effectively. Gears, commonly employed in the industry, can be either mechanical or magnetic, allowing for the balanced transfer of torque and speed at specified ratios. The mechanical and electrical actuation of in-vehicle accessories persists both in traditional and next-generation vehicles. Particularly concerning safety and the sustainability of spare part production and supply, various electrical accessories continue to operate at the 12 V. level in modern vehicles. In this context, the use of the Lundell alternator (claw pole) also continues in next-generation vehicles. While the mechanical accessories are driven by a belt-pulley system connected to an internal combustion engine in conventional vehicles, in next-generation vehicles, both belt-pulley systems and x-drive by wire are present. The low efficiency and operational costs of belt-pulley power transmission systems necessitate the adoption of more efficient transmission systems. This study focuses on the development of a soft magnetic composite (SMC) core magnetic gear power transmission system that can serve as an alternative to belt-pulley systems in both traditional and next-generation vehicles. In the proposed system, mechanical power transfer to the Lundell alternator is realized through the intended magnetic gear. The gear ratio is determined to ensure a speed of 1800 rpm at the input of the Lundell alternator while the drive system operates at 6000 rpm. To achieve low volume and high efficiency for the proposed magnetic gear, the SMC material is considered, and a comprehensive analysis using the Ansys Maxwell finite element software is conducted. As a result of the analyses, in the magnetic gear designed with a transmission ratio of 3:1, when a torque of 11.7547 Nm is applied to the input shaft, a torque transmission of 35.9806 Nm has been achieved with an efficiency of 84.73% through the output shaft.

**Index Terms**— Magnetic gear, soft magnetic composite (SMC), power transmission systems, vehicle powertrain systems.


**KADIR YILMAZ**, is with Department of Energy Systems Engineering, Kocaeli University, Kocaeli, Turkey, (e-mail: [kayel@kocaeli.edu.tr](mailto:kayel@kocaeli.edu.tr)).

 <https://orcid.org/0000-0002-0819-3420>


**TANER DİNDAR**, is with Department of Electronic and Automation, Ankara University, Ankara, Turkey, (e-mail: [tdindar@ankara.edu.tr](mailto:tdindar@ankara.edu.tr)).

 <https://orcid.org/0000-0002-4112-2114>


**MURAT AYAZ**, is with Department of Electric and Energy, Kocaeli University, Kocaeli, Turkey, (e-mail: [murat.ayaz@kocaeli.edu.tr](mailto:murat.ayaz@kocaeli.edu.tr))

 <https://orcid.org/0000-0001-9839-3970>

**SERKAN AKTAŞ**, is with Department of Machinery and Metal Technologies, Kocaeli University, Kocaeli, Turkey, (e-mail: [serkan.aktas@kocaeli.edu.tr](mailto:serkan.aktas@kocaeli.edu.tr))

 <https://orcid.org/0000-0001-6135-7597>

**SERKAN SEZEN**, is with Department of Electric and Energy, Kocaeli University, Kocaeli, Turkey, (e-mail: [serkan.sezen@kocaeli.edu.tr](mailto:serkan.sezen@kocaeli.edu.tr))

 <https://orcid.org/0000-0001-7273-7376>

Manuscript received August 28, 2023; accepted Sep 19, 2023.

DOI: [10.17694/bajece.1352245](https://doi.org/10.17694/bajece.1352245)

## I. INTRODUCTION

Magnetic gear systems have emerged as a groundbreaking technological solution, harnessing the power of permanent magnets and magnetic forces to offer a distinct departure from conventional gear systems. These systems excel in transmitting torque between input and output shafts without the need for physical contact, thereby eliminating energy losses caused by friction and enabling high-efficiency operations. Unlike traditional gearboxes that are susceptible to friction-related heating issues due to gear contact, magnetic gear systems operate without encountering such challenges. This is due to their inherent ability to facilitate motion and torque transmission through magnetic fields, thus reducing maintenance costs and ensuring dependable power transmission [1].

In contemporary industrial applications, the demand for power transmission systems with high torque density has prompted extensive research. The conventional approach involves using mechanical gearboxes to enhance torque output from electric motors, obviating the need for larger motors and achieving high torque density [2-4]. The literature reflects a burgeoning interest in various aspects of magnetic gear technology [4,5-10]. Research endeavors span a spectrum of objectives, encompassing the reduction of magnetic gear costs, enhancement of torque density and efficiency, and exploration of diverse application domains including wind energy and automotive systems. One avenue of exploration delves into the distinctive structures of magnetic gears, such as radial-axial combinations and dual-stator designs [7]. Additionally, efforts are concentrated on increasing torque density and optimizing efficiency [2-3,11-13]. Magnetic gears are proposed as alternatives to mechanical gears in marine propulsion systems, addressing the need for compact, efficient torque solutions [14]. In the burgeoning wind turbine sector, magnetic gears are positioned to tackle challenges posed by mechanical gearboxes in direct drive systems, thereby contributing to enhanced efficiency and durability [15]. Specific studies target the fundamental elements of magnetic gear design. Analytical methods are devised to characterize the magnetic field distribution in axial magnetic gears, allowing for accurate predictions [5]. Novel topologies that incorporate modulation rings are explored to tailor torque characteristics for varying speed ranges, exemplifying the versatility of magnetic gear technology [7]. Linear magnetic gears, driven by rare-earth

magnets, emerge as a potential solution for converting rotary motion into linear motion with high force density [3].

The optimization of magnetic gear designs is another active research strand. Studies delve into diverse aspects, from losses due to eddy currents and iron, to the selection of magnet materials for optimal performance [16]. Efforts are also directed towards improving efficiency and minimizing losses, with an emphasis on analyzing rotor losses and exploring innovative rotor designs [17,18]. These research directions collectively contribute to the holistic understanding and advancement of magnetic gear systems.

In view of this dynamic landscape, this study aims to contribute to the field by proposing a novel magnetic gear system design that incorporates a soft magnetic composite (SMC) core. This system integrates a permanent magnet outer rotor, an inner rotor, and a modulating carrier to achieve efficient and compact torque transmission for vehicle propulsion systems. The absence of comprehensive studies on SMC core magnetic gear systems in vehicle propulsion applications underscores the novelty and significance of this research, addressing a critical research gap. The paper is structured as follows: Section 2 details the vehicle propulsion and accessory drive system. Section 3 presents the fundamental dimension configuration and the proposed magnetic gear design. Finally, Section 4 presents the analysis results of the created models.

## II. VEHICLE ACCESSORY DRIVE SYSTEMS

The power requirement for mechanical loads within a vehicle (such as power steering pumps, air conditioning compressors, water pumps, air compressors, Lundell alternators) is approximately 6 kW, while electrical loads (such as heated steering systems, power windows, headlights, interior lighting, battery, etc.) require around 2-3 kW of power [19]. The drive block diagram of vehicle interior mechanical and electrical accessories, as depicted in Figure 1, remains applicable to both traditional and new-generation vehicles. Particularly for safety and sustainability in terms of spare part production and supply, various electrical accessories at the 12 V level continue to be used in new-generation vehicles. In this context, Low voltage battery and Lundell alternators (claw pole) are still employed in new-generation vehicles. The drive of mechanical accessories is achieved through a belt-pulley system via an internal combustion engine (ICE) in traditional vehicles, while new-generation vehicles may feature both belt- pulley systems and x-drive by wire systems. The low efficiency and operating costs of belt-pulley power transmission systems necessitate the use of more efficient transmission systems.

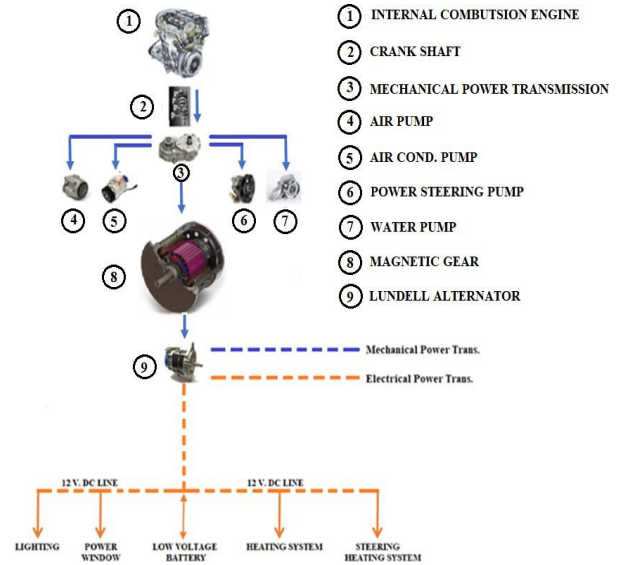


Fig.1. Vehicle accessory drive block diagram

In this study, the development of a low-volume, high-efficiency magnetic gear power transmission system is addressed as an alternative to belt-pulley power transmission systems, applicable to both traditional and new-generation vehicles. In the proposed system, mechanical power transfer to the Lundell alternator is achieved through a magnetic gear. The gear ratio is determined such that the drive system speed is 6000 rpm, while the input speed of the Lundell alternator is 1800 rpm. To achieve a low-volume, lightweight, and high-efficiency magnetic gear, SMC materials are considered, and a detailed analysis is conducted.

## III. MAGNETIC GEAR DESIGN

Magnetic power transmission devices have made significant progress with the advancement of technology. Magnetic gears have a similar structure to traditional mechanical gearboxes, but the transmission process occurs through magnetic forces instead of mechanical contact. Figure 2 illustrates the representation of a magnetic gear and its mechanical counterpart, the straight-tooth gear.

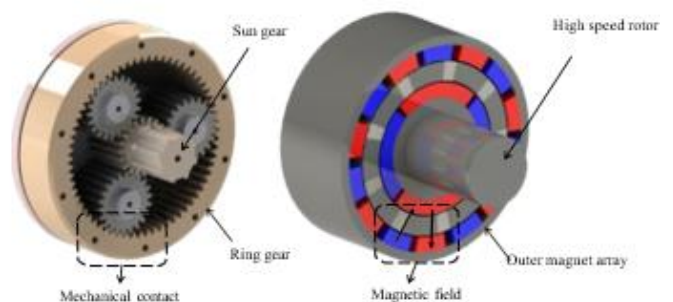


Fig.2. Mechanical gear and magnetic gear structures

Axial magnetic gears are the most common type of magnetic gear, enabling power transmission at a specific gear ratio. An axial magnetic gear consists of three main parts: the outer rotor, inner rotor, and carrier. The inner and outer rotors feature magnet pairs based on specific calculations, while the carrier is a specially designed component made from ferromagnetic and non-magnetic materials. This design allows the outer rotor to remain fixed while permitting the rotational motion of the carrier and inner rotor. The power transmission can be configured as inner rotor input and carrier output, and in some cases, the carrier can be fixed to allow rotation of both the inner and outer rotors, facilitating power transmission.

Additionally, the carrier can be made from SMC material to ensure magnetic flux transmission and high magnetic permeability. A special Hallbach array arrangement is utilized in the magnets to prevent magnetic flux leakage. In the design of magnetic gear systems, the gear ratio is determined by the number of fixed magnets. In traditional gear systems, the number of teeth and the gear ratio are interdependent, making modifications or adaptations challenging. However, in magnetic gear systems, the gear ratio can be easily adjusted. Thus, if a different gear ratio is needed, changing the number of fixed magnets is sufficient. The main components of an axial magnetic gear are shown in Figure 3, and its geometric model is presented in Figure 4.

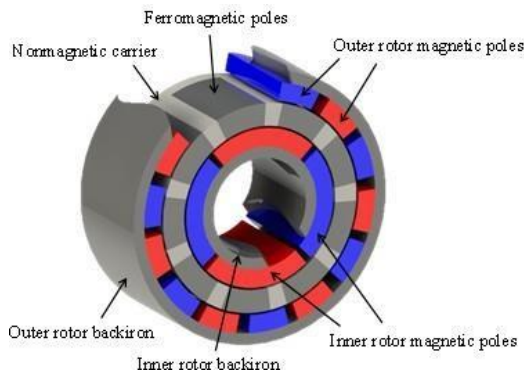


Fig.3. Axial magnetic gear model

The axial magnetic gear shown in Figure 3 is created by combining a movable inner rotor with a low number of magnets, a fixed outer rotor with a high number of magnets, and a movable carrier made from non-ferromagnetic material. In the magnetic gear, the outer rotor remains stationary, the inner rotor serves as the high-speed input shaft with low torque, and the carrier acts as the low-speed output shaft with high torque. The number of fixed magnets on the inner rotor of the axial magnetic gear is defined by using equation 1 as  $n_{PM-in}$ , while the number of magnets on the outer rotor is defined as  $n_{PM-out}$ .

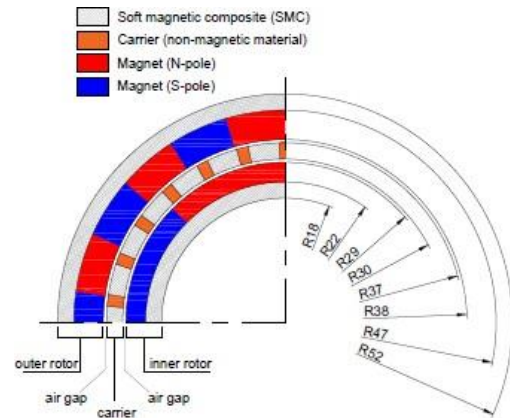


Fig.4. Geometric dimensions of the axial magnetic gear

In the designed magnetic gear, 6 fixed magnets are used on the inner rotor and 14 fixed magnets are used on the outer rotor. The Halbach array structure is preferred for magnet arrangement. The sum of the pole pairs in the inner and outer rotors determines the number of carriers. With the developments in magnet technology, studies on magnetic gear and their use have increased [22].

$$n_{sp} = \frac{n_{PM-in}}{2} + \frac{n_{PM-out}}{2} \tag{1}$$

The selection of the number of fixed magnets in magnetic gear design is also a significant aspect. In magnetic gears, the rotational ratio can be determined similarly to mechanical gear systems. In other words, gear ratios in this context are replaced by the ratio of magnet counts. In the magnetic gear depicted in Figure 3, the number of magnets on the outer rotor defines the stationary condition of the outer rotor, with the output shaft being driven by the poles in the carrier and the input shaft being driven by the poles in the inner rotor. For a magnetic gear where the outer rotor remains fixed, the gear ratio is provided by Equation 2.

$$g_r = \frac{\frac{n_{PM-in}}{2} + \frac{n_{PM-out}}{2}}{\frac{n_{PM-in}}{2}} \tag{2}$$

Here,  $g_r$  represents the gear ratio of the magnetic gear. When magnet counts are substituted into Equation 2, the gear ratio of the magnetic gear is obtained as 3.33:1. The nominal angular velocity relationship between the input shaft and the output shaft can be expressed using Equation 3.

$$w_{in} = \frac{1}{g_r} w_{out} \quad (3)$$

For this configuration, the nominal torque relationship can be calculated considering constant power with the assumption that the magnetic gear configuration operates at a constant angular velocity and there are no losses. The total mechanical power of the system should be equal to zero. This design presents a structure where there are no losses when the magnetic gear system operates, ensuring high energy efficiency. In the utilized topology, the outer rotor is fixed, resulting in zero angular velocity. With the angular velocities provided in Equation (3), the torque relationship can be formulated as follows:

$$\tau_{in} w_{in} + \tau_{out} w_{out} = 0 \quad (4)$$

$$\tau_{in} = -g_r \tau_{out} \quad (5)$$

The dimensions of the used magnets and the geometric information of the magnetic gear are presented in Figure 4, while the dimensional parameters of the magnetic gear are summarized in Table 1.

TABLE I  
BASIC MODEL PARAMETERS

Components	Parameters	Values	Unit
Inner Rotor	Inner radius	22	mm
	Outer radius	29	mm
	Pole pairs	6	-
	Angle, $\theta_1$ , $180/(2p_1)$	15	degree
Carrier	Inner radius	30	mm
	Outer radius	37	mm
	Number of poles, $n_2$	10	-
	Angle, $\theta_2$ , $180/(n_2)$	18	degree
Outer Rotor	Inner radius	38	mm
	Outer radius	47	mm
	Pole pairs	14	-
	Angle	12.85	degree
Overall	Inner radius	18	mm
	Outer radius	52	mm
	Axial length	40	mm

### III. PROPOSED MAGNETIC GEAR ANALYSIS AND SIMULATION RESULTS

In this study, a magnetic gear system with a coaxial configuration was designed, and electromagnetic analysis was conducted. Electromagnetic analysis is crucial for determining factors that influence the performance and torque transmission of the magnetic gear system. Axial torque values at specific speeds were obtained for the input and output shafts, showing variations over time. These values are essential for understanding how the magnetic gear system would perform under different load and speed conditions. The obtained data indicates the potential wide-ranging applications of magnetic gear systems.

For the analysis of a coaxial magnetic gear with SMC core, a 2-dimensional model was created using the Ansys Maxwell software. The magnetic gear was defined with the outer rotor fixed, and the carrier and inner rotor movable. SMC material, specifically SMC 130 1P, was chosen as the material for these three fundamental parts. Neodymium magnets with a Halbach arrangement were selected for the inner and outer rotors. Figure 5 illustrates the B-H characteristic of the SMC material used in the core of the magnetic gear.

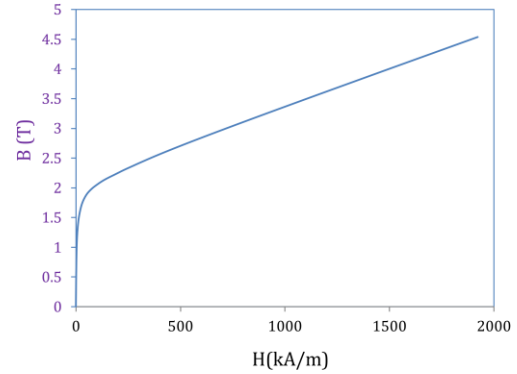


Fig.5. B-H characteristic of the SMC material.

As shown in Figure 5, the saturation point of the Somaloy series SMC material is higher compared to that of hardened steel materials. The physical, electrical, and thermal properties of SMC and hardened steel (Steel 1010) materials are provided in Table 2.

When comparing the densities of SMC material and hardened steel, it can be observed that the lower density of SMC material will lead to lower volume and weight gains compared to a core made of hardened steel of the same dimensions. In an SMC core magnetic gear, due to the material's higher resistivity, lower eddy current losses will occur compared to silicon steel. SMC materials can provide low volume-high efficiency, especially at high speeds [20]. The preference for SMC-type materials has two main purposes. The first is the ability to achieve complex geometries due to production methods, and the second is enabling three-dimensional flux flow [21].

For the parts defining the moving input and output shafts of the magnetic gear, time-dependent analyses were conducted using a moving mesh structure. The meshing process involves dividing the model into element networks to achieve realistic results. During the meshing process, it is necessary to predefine how each edge or region can be subdivided. If the mesh is too dense, it can lead to excessive computational time for solving the model.



TABLE II  
PARAMETERS OF SMC AND HARDENED STEEL MATERIALS

Materials	Density (kg/m <sup>3</sup> )	Resistivity (μΩcm)	Loss Power (50 Hz-1 T.) W/kg	Thermal Resistance (W/(m-K))
Somoloy 130 IP	7350	8000	6	22
Steel 1010	7872	5	0,5	49,8

Conversely, if the mesh is too sparse, it can deviate from accurate results. Figure 6 illustrates the distribution of magnetic flux density for the magnetic gear. It is deemed more suitable to perform the analysis with meshes of 3 degrees angle and a spacing of 0.01 mm. Within the gap between the inner rotor and the carrier, 242 elements are created, while 377 elements constitute the mesh between the inner rotor and the iron, and 40 elements form the mesh within the inner rotor's magnets.

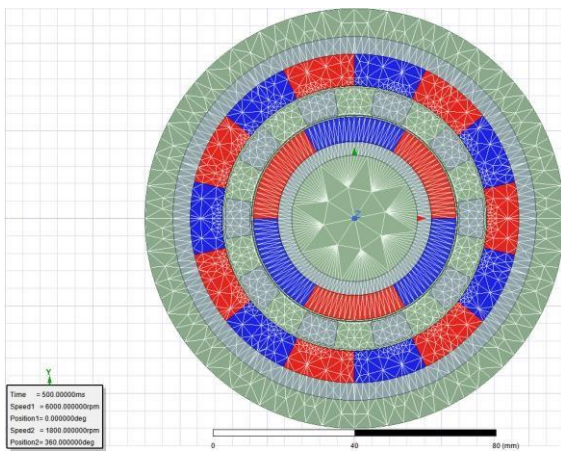


Fig.6. Mesh structure for the magnetic gear

An important aspect here is to generate denser meshes in critical areas, especially within the air gap. To enable rotational movement of the magnetic gear's moving sections, a rotating rotor model was established. To create a sliding surface for the air gap, the air gap needs to be divided into at least two parts. In Maxwell software, a band was formed to define the rotating motion, with the parts inside the band considered as moving and the parts outside as stationary. It is the division of the model into the element mesh in order to obtain a true-to-life mesh operation. When meshing, it is necessary to determine in advance which edge or how much the region can be divided. If the pattern network is large, more time can be spent in analyzing the model. If it is less, it may be possible to get away from realistic results. Figure 7 shows the magnetic flux density distribution for the magnetic gear. It is considered more appropriate to process with meshes at an angle of 3 degrees and in the range of 0.01 mm. A network of elements 242 is made in the space between the inner rotor and the carrier, 377 between the inner rotor and the iron, and 40 in the magnets in the inner rotor.

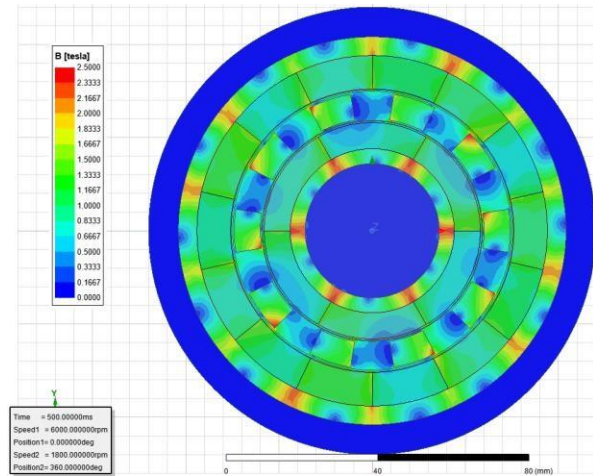


Fig.7. Distribution of magnetic flux density for the magnetic gear

Considering the moving inner rotor representing the input shaft and the moving carrier representing the output shaft, torque values were calculated for these components. Figure 8 presents the input and output torque curves of the magnetic gear model for 500 ms. The torque ripple in the input and output torques of the magnetic gear can be calculated using Equation (6):

$$T_{ripple} = \frac{T_{max} - T_{min}}{T_{avg}} \tag{6}$$

In Equation (6), T min, T max, and T avg respectively denote the minimum, maximum, and average torque values. The torque ripple percentage is 17.52% for the input torque and approximately 5% for the output torque. As shown from Figure 8, when observing the torque values between the input and output shafts, the output torque is 35.98 Nm, while the input torque is 11.75 Nm, resulting in a transmission ratio of 3:1. To minimize fluctuations in torque profiles, the numbers of poles in the inner and outer rotors should be prime. Torque ripple in magnetic gears is a result of modulation between magnetic pole rotors, primarily originating from magnetic flux harmonics. It is largely dependent on the chosen pole numbers. Significant torque ripple could render the magnetic gear practically unusable.

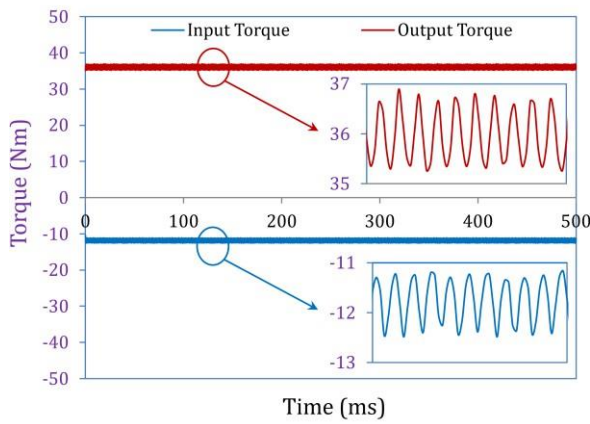


Fig.8. Torque variation for the magnetic gear inner rotor and carrier sections

Hence, studies in the literature underscore that magnetic gears with whole-number pole pairs exhibit high torque ripple.

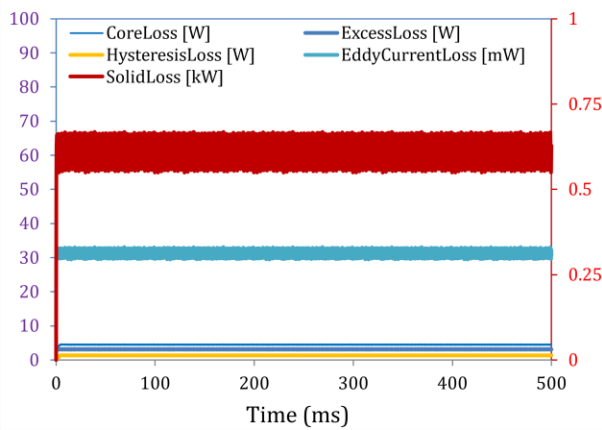


Fig.9. Core and hysteresis-eddy losses for the magnetic gear

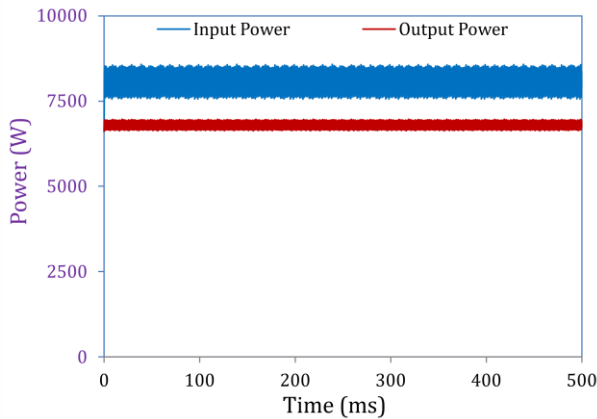


Fig.10. Power values for the input and output shafts of the magnetic gear

Magnetic gears are known to have high efficiencies due to limited mechanical losses, primarily attributed to bearing friction. However, apart from mechanical losses, there will be additional losses induced by the time-varying magnetic field. These additional losses encompass iron losses, hysteresis, and eddy current losses. Figure 9 shows the losses related to the core and hysteresis-eddy effects in the magnetic gear. When the eddy

current losses are analyzed, it becomes evident that the SMC material contributes to a relatively low value, around 31.4 mW. Additionally, the solid loss, which signifies the resistive loss in solid conductors, accounts for 615.817 W.

Using the axial torque values and angular velocities of the magnetic gear, the power values for the input and output shafts were plotted in Figure 10. Analyzing these power values reveals that torque transmission occurs with an efficiency of 84.71%, as demonstrated in Figure 11.

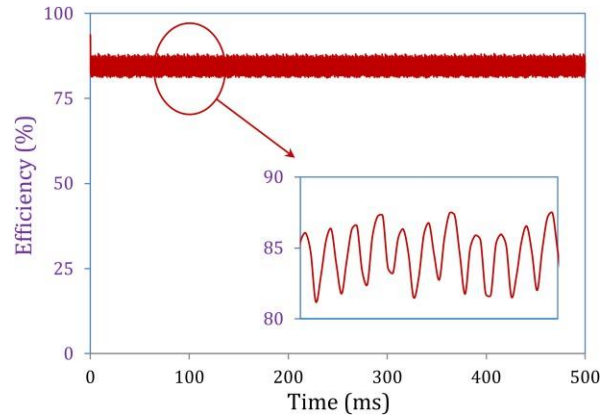


Fig.11. Efficiency of the magnetic gear

#### IV. CONCLUSION

Within the scope of this study, the development of a low-volume, high-efficiency magnetic gear power transmission system has been addressed as an alternative to the traditional belt and pulley power transmission system for both conventional and next-generation vehicles. In the proposed magnetic gear structure, the outer rotor remains fixed, while the carrier and inner rotor are designated as the output and input shafts, respectively. By constructing the magnetic gear core with SMC material, three-dimensional flux flow has been enabled, thus ensuring low eddy current losses. As a result of the analyses, power transmission has been achieved between the drive system and the Lundell alternator with a power value of 6.7 kW, at a ratio of 3:1, and with an efficiency of 84.73%. In conclusion, magnetic gear systems offer a valuable solution for increasing transmission efficiency, eliminating mechanical contact, and minimizing energy losses in torque transmission. With the advancement of technologies, the utilization of magnetic gear systems is becoming more widespread in various industrial applications, and it is anticipated to play an even more significant role in the future.

#### REFERENCES

- [1] K. Atallah, D. Howe. "A novel high performance magnetic gear." K. Atallah, D. Howe. "A novel high performance magnetic gear." IEEE transactions on magnetics, vol.37. 4, 2001, pp 2844-2846.
- [2] P. O. Rasmussen, T. O. Andersen, F. T. Jorgensen, O. Nielsen, "Development of a high-performance magnetic gear," IEEE Transactions on Industry Applications, vol. 41. 3, 2005, pp. 764-770.
- [3] J. W. K. Atallah, D. Howe, "A high-performance linear magnetic gear," Journal of applied physics 97, vol. 10. 516, 2005.
- [4] S. Nielsen, R. K. Holm, N. Berg, P. O. Rasmussen, "Magnetically geared conveyor drive unit - an updated version," IEEE Energy

- Conversion Congress and Exposition (ECCE), 2020.
- [5] L. Jian, K. T. Chau, "Analytical calculation of magnetic field distribution in coaxial magnetic gears," *Progress in Electromagnetics Research*, vol. 92, 2009, pp. 1-16.
  - [6] A. B. W. Hafila, W. Rucker, "Efficient design analysis of a novel magnetic gear on a high performance computer," *The International Journal for Computation and Mathematics in Electrical and Electronic Engineering*, vol. 26. 3, 2007, pp. 712-726.
  - [7] K. T. Chau, Z. Zhang, "Novel design of double-stator single-rotor magnetic-g geared machines," *IEEE Transactions on Magnetics*, vol. 48. 11, 2012, pp. 4180-4183.
  - [8] P. Alotto, E. Armando, "Design, Analysis and Realisation of a Magnetic Gear Prototype with Experimental Validation," 23rd International Conference on the Computation of Electromagnetic Fields, 2022.
  - [9] L. Dimauro, E. Bonisolia, M. Velardocchia, M. Repetto, P. Alotto, M. Filippini, R. Torchio "Magnetic gearbox for automotive power transmissions: an innovative industrial technology," *Engineering Science and Technology an International Journal*, 2023.
  - [10] D. Wong, H. Baninajar, B. Dechant, J. Bird, "Designing a Magnetic Gear for an Electric Aircraft Drivetrain," *IEEE Energy Conversion Congress & Expo*, 2020.
  - [11] S. K. K. Atallah and D. Howe, "High-performance magnetic gears," *Journal of Magnetism and Magnetic Materials*, 2004, pp. 272-276.
  - [12] G. Duan, T. Gupta, E. Sutton, M. Wang; M. C. Gardner, "Cycloidal Magnetic Gear Combining Axial and Radial Topologies," *IEEE Transactions on Energy Conversion* Vol.37, 3, 2022, pp.1-1.
  - [13] M. Filippini, P. Alotto, V. Cirimele, M. Repetto, C. Ragusa, L. Dimauro, E. Bonisoli, "Magnetic Loss Analysis in Coaxial Magnetic Gears," *Journals Electronics* Vol. 8, 11, 2019.
  - [14] N. Frank, H. Toliyat, "Gearing ratios of a magnetic gear for marine applications," in *IEEE Electric Ship Technologies Symposium, ESTS.*, 2009.
  - [15] N. Frank, H. Toliyat, "Gearing ratios of a magnetic gear for wind turbines," in *IEEE International Electrical Machines and Drives Conference, IEMDC'09.*, 2009.
  - [16] M. F. A. Halim, E. Sulaiman, M. Jenal, R. N. F. K. R. Othman, S. M. N. S. Othman, "Preliminary analysis of eddy current and iron loss in magnetic gear in electric vehicle," *International Journal of Electrical and Computer Engineering (IJECE)* 12(2):1161, 2022.
  - [17] Y. Tian, G. Liu, W. Zhao, J. Ji, "Design and Analysis of Coaxial Magnetic Gear Considering Rotor Losses," *IEEE Transactions on Magnetics* 51(11):1-1, 2015.
  - [18] E.J. Park, S.Y. Jung, Y.J. Kim, "Comparison of Magnetic Gear Characteristics Using Different Permanent Magnet Materials," *IEEE Transactions on Applied Superconductivity* Vol. 30, 4, 2020.
  - [19] E. Mese, M. Ayaz, M. Tezcan, K. Yilmaz, "Design of dual winding permanent magnet synchronous machines for hybrid electric vehicle accessory drives." *International Journal of Vehicle Design*, 69(1-4), 185-207, 2015.
  - [20] Y. Öner, "Rüzgar türbinleri için; düşük hızlı, sürekli mıknatıslı, yumuşak manyetik kompozit malzemeli, senkron generator tasarımı, optimizasyonu ve üretimi," *Doktora Tez*, 2013.
  - [21] O. Ocak, "Sürekli mıknatıslı senkron motorlarda moment kalitesini artırmaya yönelik melez mıknatıslı yerleşimli yeni bir rotor yapısı tasarımı ve deneysel doğrulaması," *Doktora Tez*, 2021.
  - [22] Kocaer, T. & Öner, Y. (2022). Axial Flux Motor Design for Ventilation Fans Used in The Automotive Industry . *Balkan Journal of Electrical and Computer Engineering* , 10 (3) , 295-299.

### BIOGRAPHIES



**KADİR YILMAZ** received the B.S. and M.S. degrees from the Department of Electrical Education, Gazi University, Ankara, Turkey, in 1996, 1999. He received PhD degree in 2004 from the Department of Electrical Education, Kocaeli University. Since 1999, he has been working with the Energy Systems Engineering Department, Kocaeli University. His research interests include design of electrical

machine, electromechanical energy conversion, simulation and modeling of electrical machine.



and energy systems.

**TANER DİNDAR** received the B.S. and M.S. degrees from the Department of Electrical Education, Marmara University, Istanbul, Turkey, in 2009, 2012. Since 2012, he has been working with the Department of Electronic and Automation, Ankara University. His research interests include design of electrical machine, magnetic gear



Associate Professor with the Electric and Energy Department. He has several studies on electric machine design and industrial automation system design. His research interests include electric machines, electromechanical systems, hybrid electric vehicles, and industrial automation systems.

**MURAT AYAZ** received the B.S., M.S., and Ph.D. degrees from the Department of Electrical Education, Kocaeli University, Kocaeli, Turkey, in 2005, 2008, and 2015, respectively. From 2009 to 2015, he was worked as a Research Assistant with the Electrical Education Department, Kocaeli University, where he is currently an



High School at Kocaeli University. His research fields are CAD-CAM-CNC technologies, manufacturing technologies, machine design, biomechanics.

**SERKAN AKTAŞ** received his M.Sc. and Ph.D. degrees in the Department of Machinery Technology and Biomedical Engineering, in 2009 and 2021, respectively, at Kocaeli University, Kocaeli, Turkey. He has been working as an Assistant Professor at the Uzunciftlik Nuh Cimento Vocational



Turkey, between 2007-2019. He is currently Assistant Professor at Department of Electric&Energy, Uzunciftlik Nuh Çimento Vocational School, Kocaeli University. His research interests are renewable energy, power electronics, control theories and electrical machines.

**SERKAN SEZEN** received the B.S. , M.Sc. and Ph.D. degrees in Electrical Education from Kocaeli University, Kocaeli, Turkey, in 2003, 2006 and 2015 respectively. He worked at the Department of Electronics & Automation, Edremit Vocational School, Balıkesir University, Balıkesir,



Research Article

# African Vultures Optimization Algorithm-Based Selective Harmonic Elimination for Multi-level Inverter

Yasin Bektas


**Abstract**—In recent years, multilevel inverters have gained significant attention due to their advantages, such as improved output waveform quality and reduced harmonic distortion. However, harmonics in multilevel inverter systems continue to be a persistent issue. Researchers have addressed this problem using the Selective Harmonic Elimination-Pulse Width Modulation (SHE-PWM) technique. However, the SHE equations formulated for eliminating or reducing the selected harmonics involve complex and computationally intensive calculations, encompassing nonlinear and transcendental equations. Various optimization techniques have been developed to tackle these intricate and demanding calculations. This article presents a new approach that utilizes the relatively newly developed African Vultures Optimization (AVO) algorithm to solve the SHE equations in multilevel inverters. The AVO-based SHE-PWM technique is tested on a three-phase cascaded multilevel inverter (CHB-MLI) with 7, and 11 levels. The proposed algorithm demonstrates the ability to locate adequate solutions within the modulation index range of 0.1 to 1.0. It is shown that the modulation index range of 0.5 to 1.0 allows for the successful elimination of the selected harmonics and precise control of the fundamental voltage with an error of less than 0.5%.

**Index Terms**—African Vultures Optimization Algorithm, Multi-level inverter, Selective Harmonic Elimination, Optimization.

## I. INTRODUCTION

MULTILEVEL INVERTERS find applications in various fields, including renewable energy systems (such as solar and wind energy integration), electric vehicles, high-voltage direct current (HVDC) transmission systems, and medium-voltage motor drives [1-6]. Their ability to generate high-quality output waveforms makes them suitable for applications that require solid power quality. Alongside the circuit structures of multilevel inverters, control techniques play a crucial role. Among the most popular modulation techniques used for multilevel inverters are pulse width modulation (PWM),

Yasin Bektas, is with Department of Electrical-Electronic Engineering Aksaray University, Aksaray, Turkey.(e-mail: [yasinbektas@aksaray.edu.tr](mailto:yasinbektas@aksaray.edu.tr))

 <https://orcid.org/0000-0002-3681-0123>

Manuscript received Jun 10, 2023; accepted Aug 13, 2023.

DOI: [10.17694/bajece.1312668](https://doi.org/10.17694/bajece.1312668)

selective Harmonic elimination (SHE-PWM), and space vector PWM (SV-PWM) [7-9]. These techniques enhance the efficiency of the inverter while also lowering the amount of harmonic distortion that is present in the waveform that is produced. SHE-PWM stands out as the most effective due to its direct harmonic elimination capability. SHE-PWM is a modulation technique that selectively eliminates or reduces lower-order harmonics in the output waveform. When SHE-PWM is used, the output waveform is improved with smoother characteristics, and the total harmonic distortion (THD) value is reduced. This minimizes the effect of unwanted harmonic components and yields a cleaner output waveform.

Various optimization methods have been developed to address the computational complexity of SHE. These optimization techniques aim to find optimum switching angles that satisfy the desired harmonic elimination conditions while minimizing the computational effort. Genetic Algorithms (GA) [10], Particle Swarm Optimization (PSO) [11], Differential Evolution (DE) [12], Artificial Bee Colony (ABC) [13], Red Deer Algorithm (RDA) [14], and other metaheuristic algorithms have been applied in SHE optimization. These algorithms explore the solution space to find the best solutions, considering factors such as convergence speed, computational efficiency, robustness, and the ability to handle different modulation indices. Utilizing optimization methods in selective harmonic elimination enables the efficient implementation of multilevel inverters, achieves high-quality output waveforms, and reduces harmonic distortion. Researchers and engineers can adapt the performance of multilevel inverters to meet specific application requirements, optimize power conversion efficiency, and adhere to power quality standards by selecting an appropriate optimization algorithm. Ongoing research in this field aims to enhance the efficiency and accuracy of optimization algorithms and their application in real-time control systems, thus paving the way for the widespread adoption of multilevel inverters in various power electronics applications.

This study presents a new approach that utilizes the African Vultures Optimization (AVO) algorithm [15] to solve the SHE equations in multilevel inverters. The results show that the proposed algorithm works well at finding reasonable solutions within a specific range of modulation indices, allowing the elimination of certain harmonics and precise control of the fundamental voltage. This research contributes to the



advancement of optimization techniques in multilevel inverters and offers promising possibilities for improving power quality in various applications.

II. MATERIAL AND METHOD

A. Cascaded Multilevel Inverters

Multilevel inverters come in three main topologies: diode-clamped, capacitor-clamped, and cascaded. Among these structures, cascade multi-level inverter has different advantages compared to other inverter types. The cascaded H-bridge inverter (CHB) structure with distinct DC sources is the topology that needs the fewest number of components overall. Because of its modular design and flexible circuit configuration, it is well suited for use in applications requiring high voltage and high power. A cascaded H-bridge multilevel inverter, also known as a CHB-MLI, is created by joining a number of single-phase H-bridge inverters in a series configuration, as seen in Fig. 1(a) for an N-level inverter.  $N=2m+1$  describes how many different levels of output phase voltage the CHB-MLI can produce, where m refers to the number of H-bridges used in each phase. Each H-bridge is fed by its own individual DC source. The three distinct voltage levels of  $+V_{DC}$ ,  $0 V_{DC}$ , and  $-V_{DC}$  may be achieved by choosing various combinations of the switches labelled  $S_1, S_2, S_3,$  and  $S_4$  located inside each H-bridge. In order to make the synthesized AC voltage waveform identical to the sum of all the voltages produced by the cascaded H-bridge cells, the outputs of the H-bridge switches are connected in series with one another.

Depending on the application, a three-phase circuit structure is obtained by either star or delta connection. This study uses three-phase circuit structures with 7, and 11 levels. For seven levels, three H-bridge structures are required for each phase, while four H-bridge structures are needed for nine levels, and five H-bridge structures are required for eleven levels.

B. Mathematical Model of SHE-PWM Technique

Selective Harmonic Elimination (SHE) is a PWM (pulse width modulation) technique commonly used at low frequencies. Its purpose is to control the output voltage while adjusting the fundamental harmonic to the desired value and eliminating selected harmonics to achieve a sinusoidal AC output voltage waveform. After applying SHE-PWM, the remaining insignificant higher-order harmonics can be eliminated using a small passive filter. In the SHE-PWM technique, a set of nonlinear equations, usually denoted by m, is solved to find the optimum switching angles for different modulation indices.

One of the equations equals the fundamental harmonic value, while the others are set to zero to eliminate the selected harmonics. Subsequently, the switching angles that satisfy these equations are calculated using appropriate methods. The nonlinear harmonic equations required to obtain the optimal switching angles can be expressed using the Fourier series expansion of the output voltage. The expression for the output voltage, which includes all harmonic components, can be defined as shown in Eq. (1).

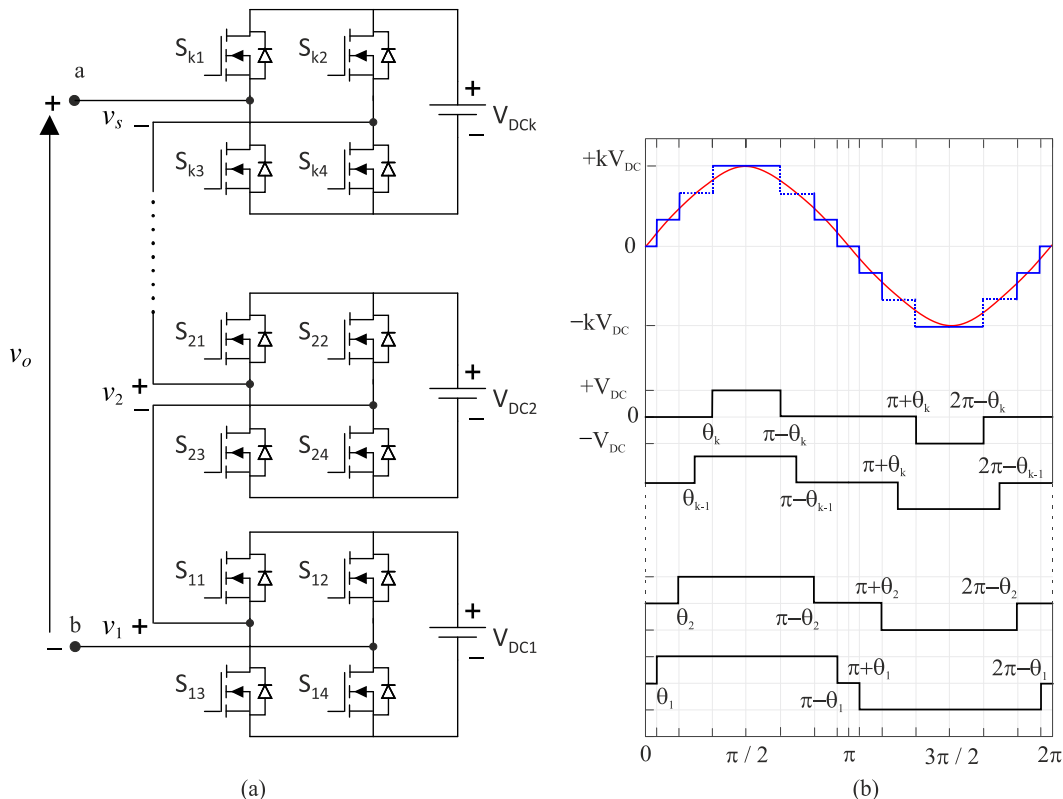


Fig.1. Single-phase N-level CHB-MLI (a) circuit structure (b) output voltage waveform

$$V_{ab}(\omega t) = \sum_{n=1,3,5,7,\dots}^{\infty} \frac{4V_{DC}}{n\pi} * [\cos(n\theta_1) + \cos(n\theta_2) + \dots + \cos(n\theta_{k-1}) + \cos(n\theta_k)] * [\sin(n\omega t)] \quad (1)$$

where,  $V_{DC}$  is the input voltage for each H-bridge inverter.  $\theta_1, \theta_2, \dots,$  and  $\theta_k$  are the switching angles, and due to quarter-wave symmetry, the switching angles should satisfy the condition in Eq.(2).

$$0 \leq \theta_1 < \theta_2 < \dots < \theta_{k-1} < \theta_k \leq \frac{\pi}{2} \quad (2)$$

$k$  represents the number of switching angles, and  $n$  represents the degree of the harmonic. For the 7-level, and 11-level inverters, the required number of discrete DC sources is three, four, and five, respectively. In a balanced three-phase system, there are no harmonics that are multiples of three, so the harmonics that are multiples of three are not considered. Typically, in the context of  $k$  switching angles, a single switching angle is employed to regulate the intended fundamental output voltage  $V_1$ , while the remaining  $(k-1)$  switching angles are utilised to elimination or diminish the low-order harmonics  $(k-1)$ . The Eq.(1) provides the formula for the fundamental output voltage  $V_1$ , expressed in relation to the switching angles.

$$V_1 = \frac{4V_{DA}}{\pi} \left( \cos(\theta_1) + \cos(\theta_2) + \dots + \cos(\theta_{k-1}) + \cos(\theta_k) \right) \quad (3)$$

The modulation index controls the fundamental voltage in the SHE technique.  $M$ , can be defined as the ratio of the peak value ( $V_{1p}$ ) of the desired base voltage given in (4) to the total DC input voltage [16].

$$M = \frac{V_{1p}}{kV_{DA}} \quad (4)$$

The determination of the modulation index and switching angles that result in the generation of an AC waveform with minimal total harmonic distortion (THD) necessitates the resolution of  $k$  transcendental nonlinear equations, commonly referred to as SHE equations, that describe the chosen harmonics.

The seven-level inverter utilizes three h-bridge structures. The SHE equations for the 7-level inverter are as follows:

$$\begin{aligned} V_1 &= \cos(\theta_1) + \cos(\theta_2) + \cos(\theta_3) = Mk \pi/4 \\ V_5 &= \cos(5\theta_1) + \cos(5\theta_2) + \cos(5\theta_3) = 0 \\ V_7 &= \cos(7\theta_1) + \cos(7\theta_2) + \cos(7\theta_3) = 0 \end{aligned} \quad (5)$$

The eleven-level inverter utilizes five h-bridge structures. The SHE equations for the 11-level inverter are as follows:

$$\begin{aligned} V_1 &= \cos(\theta_1) + \cos(\theta_2) + \dots + \cos(\theta_5) = Mk \pi/4 \\ V_5 &= \cos(5\theta_1) + \cos(5\theta_2) + \dots + \cos(5\theta_5) = 0 \\ V_7 &= \cos(7\theta_1) + \cos(7\theta_2) + \dots + \cos(7\theta_5) = 0 \\ V_{11} &= \cos(11\theta_1) + \cos(11\theta_2) + \dots + \cos(11\theta_5) = 0 \\ V_{13} &= \cos(13\theta_1) + \cos(13\theta_2) + \dots + \cos(13\theta_5) = 0 \end{aligned} \quad (6)$$

For the correct solution, the switching angles for all inverter states must satisfy the condition  $0 \leq \theta_1 < \theta_2 < \dots \leq \theta_k$ . In this study, two different total harmonic distortion values will be calculated. The first is %THD, and the other is % THDe. The %THD limit value is usually infinite but will be considered up to the 50<sup>th</sup> harmonic. THDe represents the total value of the harmonics to be eliminated. When calculating the THDe [16] value, the maximum harmonic value to be eliminated is considered. The maximum harmonic value for the seven-level is 7<sup>th</sup>, 11<sup>th</sup>, and 13<sup>th</sup> for the eleven-level.

$$\%THD = \frac{\sqrt{V_5^2 + V_7^2 + V_{11}^2 + \dots + V_{49}^2}}{|V_1|} \quad (7)$$

$$\%THD_e = \frac{\sqrt{V_5^2 + V_7^2 + \dots}}{|V_1|} \quad (8)$$

### C. African Vultures Optimization Algorithm

African Vulture Optimization (AVO) is an optimization algorithm developed by drawing inspiration from the feeding behavior of vultures in natural habitats. Vultures efficiently locate their prey over a wide area through their social interactions and observational abilities. Mimicking these natural strategies, the AVO algorithm is used to solve complex optimization problems. It employs a population-based approach and utilizes a set of candidate solutions (individuals) to solve the problem. The algorithm moves each individual through steps that imitate the vultures' prey-finding behavior. Individuals search to explore the potential solution space and move towards better solutions. This process is combined with a competitive selection mechanism to determine the best solutions among individuals.

The AVO algorithm has demonstrated its effectiveness in solving various optimization problems, as evidenced by several studies [15], [17-25]. It has been successfully applied to complex real-world problems inspired by nature, as well as function optimization, data mining, machine learning, artificial neural networks, and other domains [15]. AVO offers several key advantages, including computational efficiency, the ability to perform global searches, scalability, and high solution accuracy. This section provides a summary of the general steps of the African Vultures Optimization (AVO) algorithm. For more comprehensive details and in-depth explanations, readers are encouraged to refer to [15]. This optimization is capable of resolving major optimization issues and has a low computational complexity of  $O(P \times X (M + Mb))$ . Here, 'P' stands

for population size, 'M' for the maximum number of iterations, and 'b' for the dimensions of the problem.

The behavior of vultures in nature while foraging is depicted in Fig. 2. When many vultures congregate around one food source, it can cause severe conflicts over food acquisition. At such times, physically powerful vultures prefer not to share food with other vultures, as shown in Fig. 2(a). Vulture rotational motion is modeled using spiral motion in Fig. 2(b). The distance between the vulture and one of the two best vultures is first calculated in this method. Moreover, a spiral between the vulture and one of the best vultures is created. As seen in Fig. 2(c), when vultures are hungry and have low energy, vulture gathers together in search of food. Vultures compete in this pursuit. Vultures are aggressive in this search.

In general, the steps of the algorithm are as follows.

**Step 1.**

Create the initial population by considering the fitness function.

**Step 2.**

Choose the two best solutions as the best vultures of the first group and the second-best solution as the best vultures of the second group.

**Step 3.**

Move the other vultures to the best vultures using equations 9 and 10.

In Eq. (9),  $Vulture_{Best1}$  represents the position vector of the best vulture in the first group in the current iteration, and  $Vulture_{Best2}$  represents the position vector of the best vulture in the second group. One of the top vultures in the current iteration is represented by the position vector  $E(i)$ . This vector is selected based on the given selection operator in Eq. (10).  $\beta_1$  and  $\beta_2$  are parameters measured before the search process. The values of these parameters range between zero and one, and the sum of both parameters is equal to one.

$$k_i = \frac{F_i}{\sum_{i=1}^n F_i} \tag{10}$$

**Step 4.**

Use equations (11) and (12) to search for food.

$$F = (2 \times rand_1 + 1) \times m_1 \times \left( 1 - \frac{it_i}{max_{it}} \right) + t \tag{11}$$

$$t = m_2 \times \left( \sin^{m_3} \left( \frac{\pi}{2} \times \frac{it_i}{max_{it}} \right) + \cos \left( \frac{\pi}{2} \times \frac{it_i}{max_{it}} \right) - 1 \right) \tag{12}$$

$$E(i) = \begin{cases} Vulture_{Best1} & \text{if } k_i = \beta_1 \\ Vulture_{Best2} & \text{if } k_i = \beta_2 \end{cases} \tag{9}$$

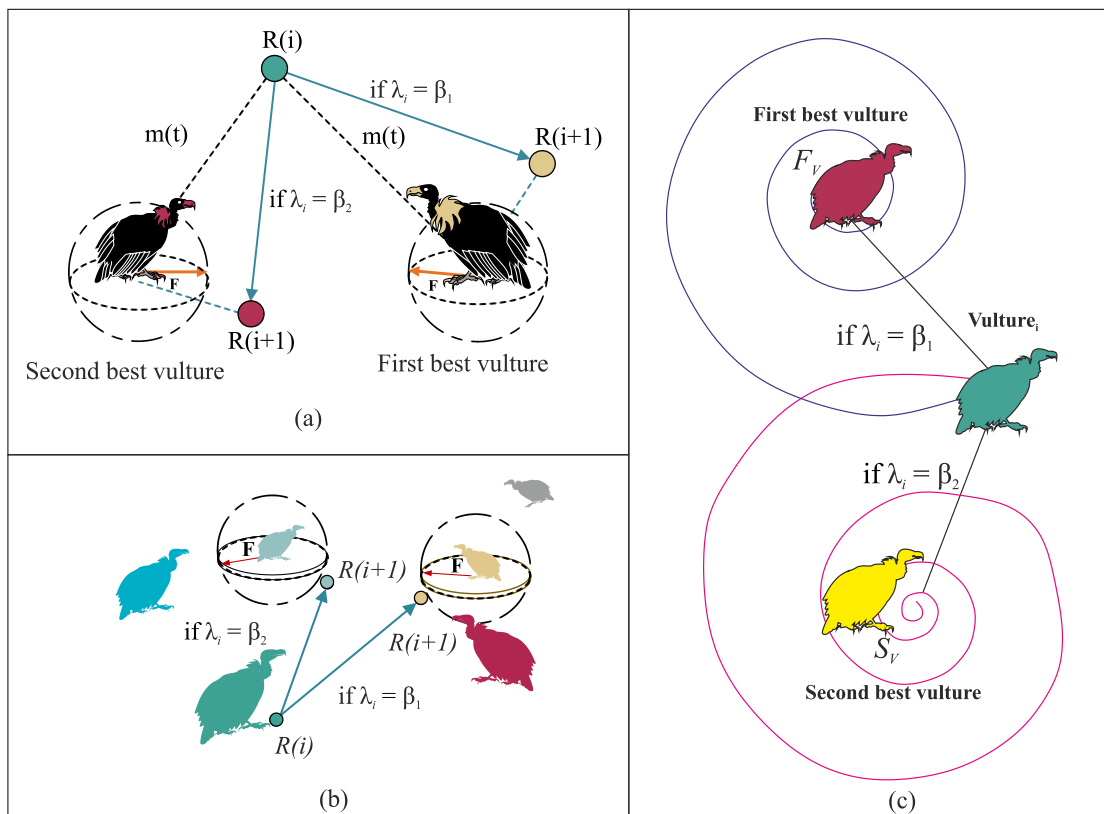


Fig.2. Natural behavior of vultures while foraging a) competition for food, b) rotating flight of vultures, c) aggressive competition for Food.

The F value given in Eq. (11) shows the saturation rate of vultures.  $it_i$  represents the valid number of iterations,  $max_{it}$  represents the maximum number of iterations. The  $rand_1$ ,  $m_1$ ,  $m_2$  parameters take random values in the ranges [0,1], [-1,1] and [2,-2], respectively. The  $m_3$  parameter is a constant number. Increasing this value increases the probability of entering the exploration phase towards the maximum iteration.

[F] If the value is greater than or equal to 1, the exploration phase begins, and the vultures forage for the best vulture location at random locations. In order to benefit from different exploration strategies, the  $k_1$ , and  $rand_{k_1}$  parameters, which take random values in the [0,1] range, are compared. Eqs. 13 and 14 are used if  $k_1$  is greater than or equal to  $rand_{k_1}$ . If  $k_1$  is less than  $rand_{k_1}$ , Eq.15 is used.

$$K(i+1) = E(i) - D(i) \times F \tag{13}$$

$$D(i) = |2 \times rand_2 \times E(i) - K(i)| \tag{14}$$

$$K(i+1) = E(i) - F + rand_3 \times ((s_2 - s_1) \times rand_d + s_1) \tag{15}$$

In equation, K(i) refers to the vulture's current position vector, and K(i+1) is the position vector in the following iteration. The lower and upper limits of the search space are represented by  $s_1$  and  $s_2$ , respectively. The parameters  $rand_2$ ,  $rand_3$ , and  $rand_4$  are assigned a random value between 0 and 1.

$$K(i+1) = D(i) \times (F + rand_4) - m(t) \tag{16}$$

$$m(t) = E(i) - K(i) \tag{17}$$

$$K(i+1) = E(i) - (S_1 + S_2) \tag{18}$$

$$S_1 = E(i) \times \left( \frac{rand_5 \times K(i)}{2\pi} \right) \times \cos(K(i)) \tag{19}$$

$$S_2 = E(i) \times \left( \frac{rand_6 \times K(i)}{2\pi} \right) \times \sin(K(i))$$

[F] when the value is less than 0.5, the  $k_3$  and  $rand_{k_3}$  parameters are compared. If  $k_3$  is equal to or greater than  $rand_{k_3}$ , when using Eq.20 and Eq.21; If  $k_3$  is less than  $rand_{k_3}$ , 21 and 22 are used in the equation. Thus, the accumulation and aggressive bickering of vultures around the food source is modeled.

$$K(i+1) = \frac{B_1 + B_2}{2} \tag{20}$$

$$B_1 = \text{Vulture}_{Best1}(i) - \frac{\text{Vulture}_{Best1}(i) \times K(i)}{\text{Vulture}_{Best1}(i) - K(i)^2} \times F \tag{21}$$

$$B_2 = \text{Vulture}_{Best2}(i) - \frac{\text{Vulture}_{Best2}(i) \times K(i)}{\text{Vulture}_{Best2}(i) - K(i)^2} \times F$$

$$K(i+1) = E(i) - |d(t)| \times F \times Levy(d) \tag{22}$$

$$L(z) = 0.01 \times \frac{m_4 \times \sigma}{|m_5|^\lambda}, \tag{23}$$

$$\sigma = \left( \frac{\Gamma(1+\lambda) \times \sin\left(\frac{\pi\lambda}{2}\right)}{\Gamma(1+\lambda) \times \lambda \times 2^{\left(\frac{\lambda-1}{2}\right)}} \right)^{\frac{1}{\lambda}}$$

In Eq. (23)  $\lambda$  is a constant number, parameters  $m_4$  and  $m_5$  take random values between 0 and 1.  $\Gamma(z) = (z-1)!$ . The flow chart of AVOA is given in Fig. 3.

#### D. Application of AVO Algorithm to SHE-PMW equations

The AVO algorithm implemented in MATLAB was utilized to solve the SHE equations (5), and (6) for the targeted harmonics in the 7, and 11-level inverters, respectively. Three scenarios were considered, with a population size of 100 and 100 iterations. The solutions were obtained by incrementing the modulation index M from 0 to 1 in steps of 0.01. The calculations were performed on a personal computer equipped with an Intel(R) Core (TM) i7-10870H CPU @ 2.20GHz, 16.0 GB RAM, and a GeForce RTX 2060 NVIDIA graphics card. At each step, the obtained solution was evaluated using a fitness function. The objective was to determine the switching angles that either eliminate or reduce the selected low-order harmonics to an acceptable level while achieving the desired voltage value. Eq. (24) defines the fitness function used for each set of solutions. The goal of this optimization process was to find the optimal set of switching angles that minimizes the distortion caused by the harmonics, ensuring the output voltage closely matches the desired waveform. The AVO algorithm, with its population-based approach and imitation of vulture behavior, efficiently searched the solution space to find the best configuration of switching angles for each modulation index.

$$f = \min_{\theta_i} \left\{ |V_{ref} - V_{1p}| + \left( \frac{4V_{DA}}{n\pi} \cdot \left( \sum_{i=2}^k \cos(n\theta_i) \right) \right)^2 \right\} = 0 \tag{24}$$

In Eq. (24)  $V_{ref}$  represents the maximum value of the desired base voltage, while  $V_{1p}$  represents the maximum value of the base voltage obtained at the inverter output applying the calculated switching angles to the inverter. The fundamental frequency of the fundamental voltage is 50Hz. Total source voltage values are selected to be a maximum of 311 volts. For the seven-level inverter, all sources are 311/3 volts, and all sources for the eleven-level inverter are 311/5 volts. An RL load was used as the load for both seven-level, and eleven-level CHB-MLI. The value of the resistance R is 10 ohms, and the value of the inductance L is 5 mH. The load connected to the inverter output affects the harmonic formation [26]. Here, harmonic analysis has been made only for RL load. The simulation results are explained in the next section.



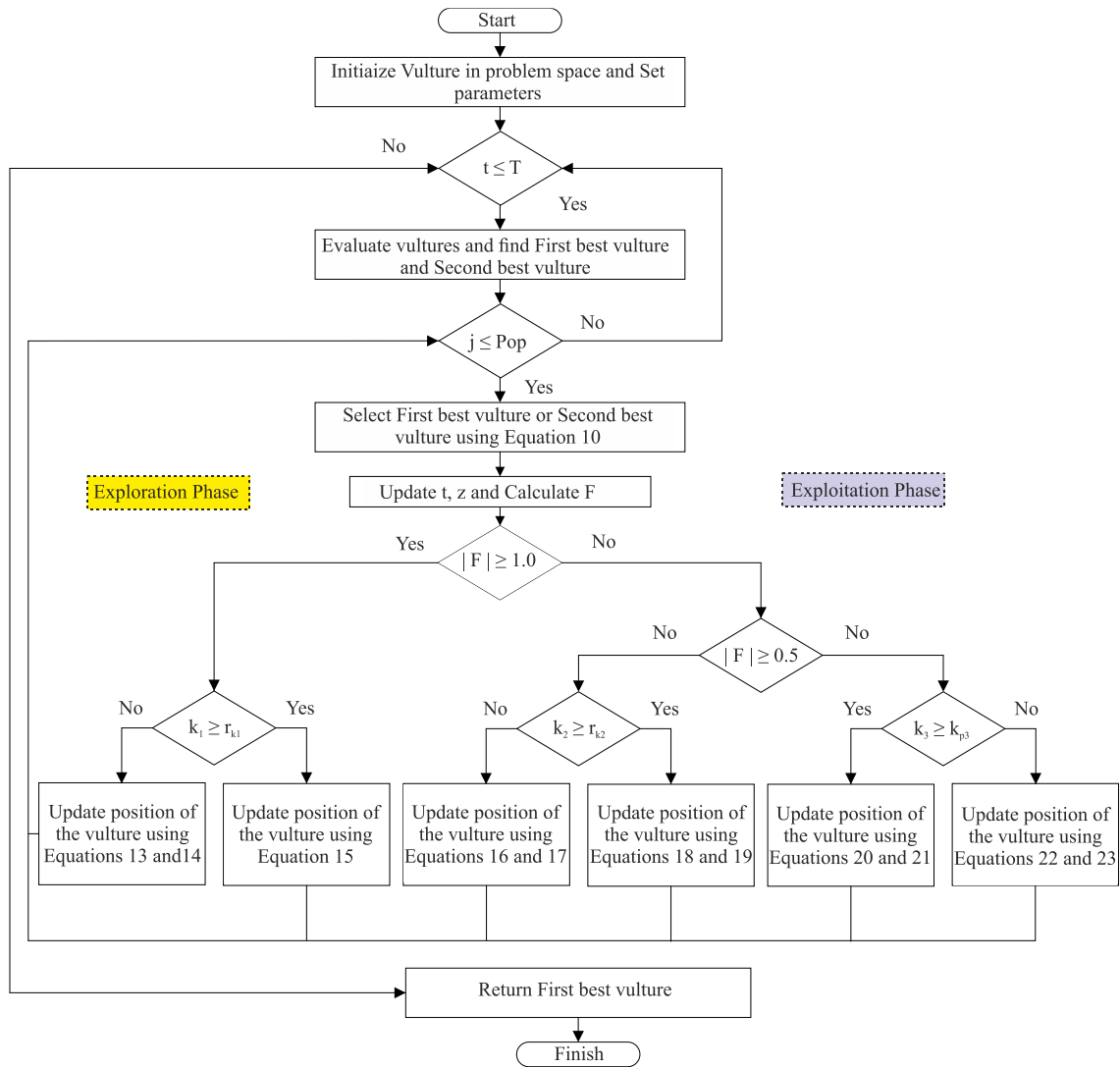


Fig.3. AVOA [15]

III. RESULTS AND DISCUSSION

To assess the effectiveness of the AVOA optimization in solving the SHE-PWM equations, the results obtained from applying the AVOA algorithm in the MATLAB Simulink environment for the 7-, and 11-level inverters are discussed below.

TABLE I  
PARAMETER SETTING OF AVOA

Parameter	Value
$\beta_1$	0.8
$\beta_2$	0.2
w	2.5
$k_1$	0.6
$k_2$	0.4
$k_3$	0.6

Table I displays the parameters utilized for AVOA at levels 7 and 11. The w operator mentioned in Table II is a parameter that influences the extent of disruption during the exploration and exploitation phases.

A. Simulation results for seven-level MLI

The graph in Fig. 4(a) depicts the fitness function values for each set of switching angles across the range of modulation index values Fig. 4(b) displays the switching angles determined for the corresponding modulation index values, providing insight into the optimal configuration for the 7-level MLI. In Fig. 4(c), the plots showcase the THD, THDe, and magnitudes of the 5<sup>th</sup> and 7<sup>th</sup> harmonics. These metrics are evaluated for modulation index values, highlighting the system's harmonic performance. Notably, the graph indicates that solutions within the M range of 0.1 to 0.4 fail to meet the IEEE-519 standard [27], while acceptable solutions are found in the M range of 0.5 to 1.0.

The switching angles calculated by AVO versus different modulation indices for 7-level MLI are shown in Table II. The switching angles calculated in Table II were applied to the 7-level inverter in the Matlab environment, and the simulation output values are shown in Table III.

As shown in Table III, the AVO algorithm has successfully found suitable solutions for the 7-level inverter within the modulation index range of 0.1 to 1.0. The desired fundamental voltage is controlled with an error of less than 0.5% compared to the reference value. The 0.5 to 1.0 index values modulation range effectively suppresses the selected harmonics. Examining this range, both the THDe value and the magnitudes of the 5th and 7th harmonics are below 0.09%. Figure 5 illustrates the load voltage waveforms for modulation indices of 0.4, 0.6, and 1.0. The THD and THDe analysis of these waveforms is presented in Figure 6 and Figure 7, respectively.

For a modulation index of  $M=0.4$ , the voltage waveform of the load exhibits a THD value of 16.94%. The maximum value of the fundamental voltage,  $V_{an(max)}$ , is measured as 124.2V, and the rms value,  $V_{an(rms)}$ , is calculated as 87.83V (Figure 6(a)). For a modulation index of  $M=0.6$ , the THD of the load voltage waveform is determined as 12.41%. The maximum value of the fundamental voltage,  $V_{an(max)}$ , is found to be 186.2V, and the rms value,  $V_{an(rms)}$ , is calculated as 131.7V (Figure 6(b)). Finally, for a modulation index of  $M=1.0$ , the THD of the load voltage waveform is reduced to 7.83%. The maximum value of the fundamental voltage,  $V_{an(max)}$ , is measured as 311.1V, and the rms value,  $V_{an(rms)}$ , is determined as 220.1V (Figure 6(c)).

Figure 7 illustrates the extent to which the selected harmonics are eliminated. For  $M=0.4$  modulation index, the calculated values are  $THDe=6.35\%$ ,  $h5=3.99$  ( $V5=4.95V$ ), and  $h7=3.99$  ( $V7=6.14V$ ). For  $M=0.6$  modulation index, the calculated values are  $THDe=0.09\%$ ,  $h5=0.02$  ( $V5=0.03V$ ), and  $h7=0.15$  ( $V7=0.08V$ ). Lastly, for  $M=1.0$  modulation index, the calculated values are  $THDe=0.04\%$ ,  $h5=0.03$  ( $V5=0.01V$ ), and  $h7=0.01$  ( $V7=0.03V$ ).

The analysis indicates that the AVO algorithm precisely controls the fundamental voltage and effectively suppresses the selected harmonics within the modulation index range of 0.5 to

1.0. The obtained results demonstrate the capability of the AVO algorithm to achieve desired voltage control and harmonic suppression for the 7-level inverter.

*B. Simulation results for eleven-level MLI*

The graph in Fig. 8(a) depicts the fitness function values for each set of switching angles across the range of modulation index values Fig. 8(b) displays the switching angles determined for the corresponding modulation index values, providing insight into the optimal configuration for the 11-level MLI. In Fig. 8(c), the plots showcase the THD, THDe, and magnitudes of the 5th, 7th, 11th and 13th harmonics. These metrics are evaluated for modulation index values, highlighting the system's harmonic performance. Notably, the graph indicates that solutions within the M range of 0.1 to 0.4 fail to meet the IEEE-519 standard, while acceptable solutions are found in the M range of 0.5 to 1.0.

The switching angles calculated by AVO versus different modulation indices for 11-level MLI are shown in Table IV. The switching angles calculated in Table IV were applied to the 11-level inverter in the Matlab environment, and the simulation output values are shown in Table V.

TABLE II  
SWITCHING ANGLES CALCULATED WITH AVO (FOR 7 LEVELS)

Modulation Index	Switching Angles (Radians)			
	M	$\theta_1$	$\theta_2$	$\theta_3$
0.1	0.1	1.3329	1.5708	1.5708
0.2	0.2	1.0801	1.5708	1.5708
0.3	0.3	0.8907	1.4928	1.5708
0.4	0.4	0.7787	1.3380	1.5708
0.5	0.5	0.7115	1.1488	1.5597
0.6	0.6	0.6882	1.0224	1.4504
0.7	0.7	0.6691	0.9414	1.2908
0.8	0.8	0.5102	0.9503	1.1254
0.9	0.9	0.3056	0.7512	1.1196
1.0	1.0	0.2029	0.5370	1.0181

TABLE III  
SIMULATION RESULTS (FOR 7 LEVELS)

M	Vref(peak)	Vref(rms)	V(rms)	Error (%)	THD (%)	THDe (%)	h5	h7
0.10	31.1	22	21.96	<b>0.18</b>	109.14	99.42	79.03	60.32
0.20	62.2	44	43.95	<b>0.11</b>	37.77	28.44	27.09	8.68
0.30	93.3	66	65.83	<b>0.26</b>	32.95	10.37	3.65	9.71
0.40	124.4	88	87.83	<b>0.19</b>	16.94	6.35	3.95	4.95
0.50	155.5	110	109.8	<b>0.18</b>	17.44	<b>0.06</b>	<b>0.01</b>	<b>0.06</b>
0.60	186.6	132	131.7	<b>0.23</b>	12.41	<b>0.09</b>	<b>0.02</b>	<b>0.08</b>
0.70	217.7	154	153.6	<b>0.26</b>	12.26	<b>0.07</b>	<b>0.02</b>	<b>0.06</b>
0.80	248.8	176	175.6	<b>0.23</b>	10.72	<b>0.04</b>	<b>0.01</b>	<b>0.00</b>
0.90	279.9	198	197.4	<b>0.30</b>	11.81	<b>0.03</b>	<b>0.00</b>	<b>0.01</b>
1.00	311.0	220	220.1	<b>0.05</b>	7.84	<b>0.04</b>	<b>0.01</b>	<b>0.03</b>

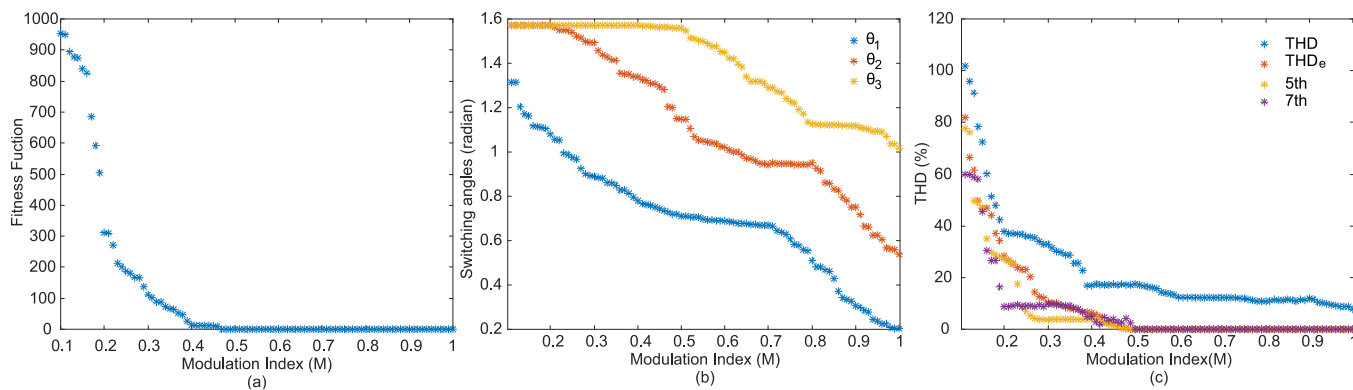


Fig.4. Analysis of the 7-level MLI inverter in terms of (a) fitness function, (b) switching angles, and (c) THD, THDe, 5<sup>th</sup>, and 7<sup>th</sup> harmonic characteristics as a function of M.

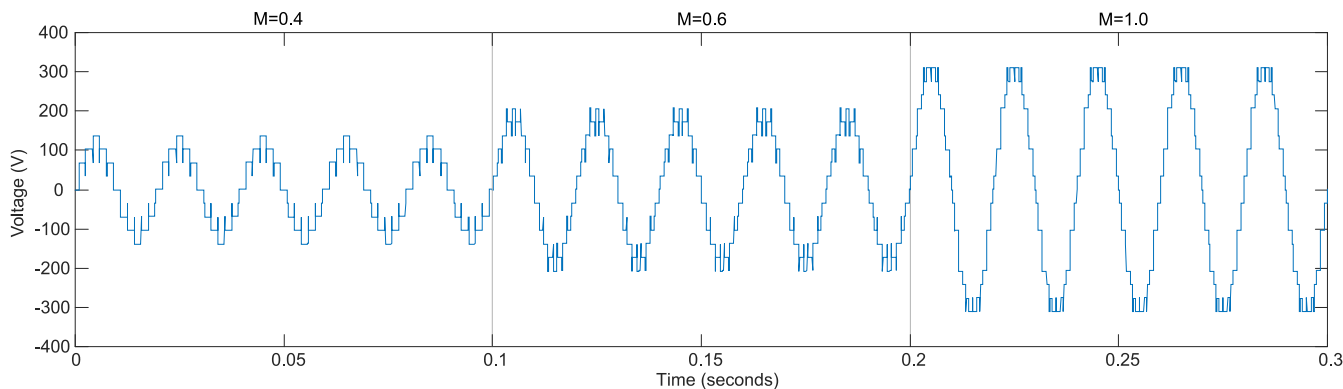


Fig.5. Load voltage waveforms for different modulation indices (7-level): (a) M=0.4, (b) M=0.6, (c) M=1.0

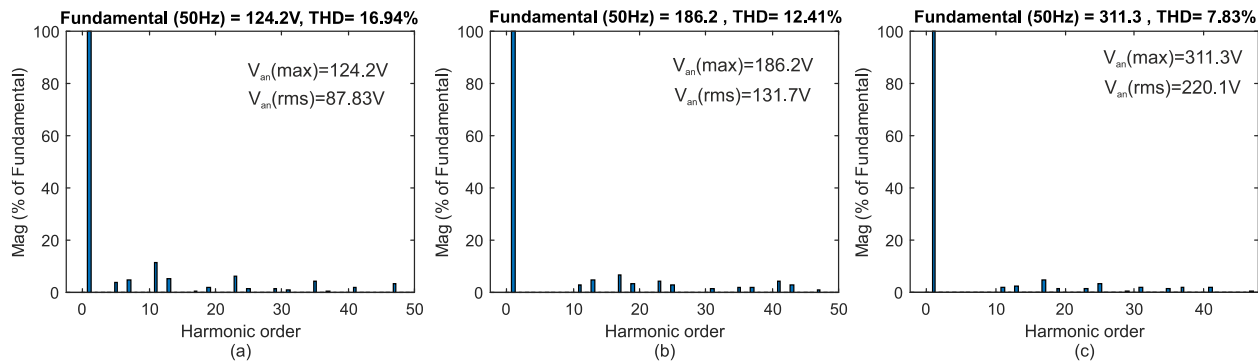


Fig.6. THD values for different modulation indices (7-level) (a) M=0.4, (b) M=0.6, (c) M=1.0

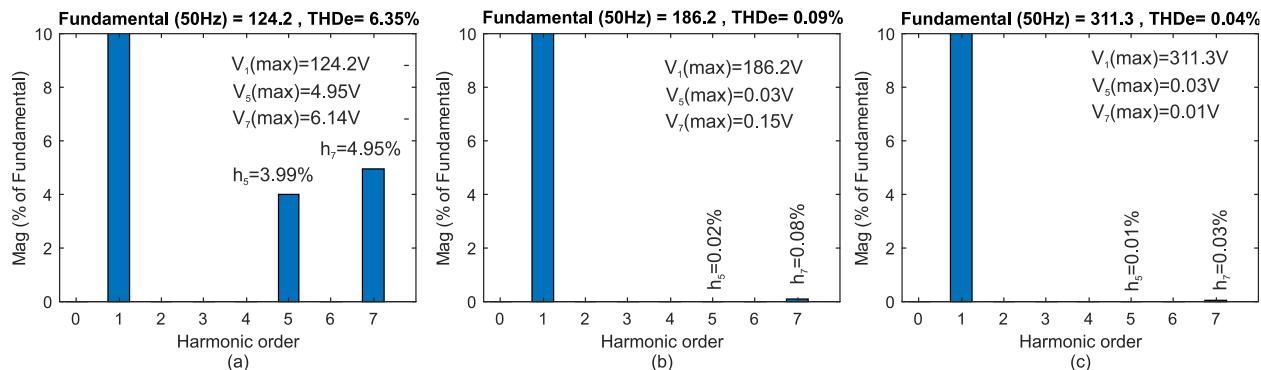


Fig.7. THDe values for different modulation indices (7-level) (a) M=0.4, (b) M=0.6, (c) M=1.0

As shown in Table V, the AVO algorithm has successfully found suitable solutions for the 11-level inverter within the M range of 0.1 to 1.0. It is observed that the desired fundamental voltage is controlled with an error of less than 0.5% compared to the reference value. M range of 0.6 to 1.0 is identified as the effective range for suppressing selected harmonics. Within this range, both the THDe load value and the magnitudes of the 5th, 7th, 11th, and 13th harmonics are below 0.09%.

TABLE IV  
SWITCHING ANGLES CALCULATED WITH AVO (FOR 11 LEVELS)

Modulation Index	Switching Angles (Radians)				
	$M$	$\theta_1$	$\theta_2$	$\theta_3$	$\theta_4$
0.1	1.1672	1.5708	1.5708	1.5708	1.5708
0.2	0.7575	1.5119	1.5708	1.5708	1.5708
0.3	0.7706	1.1356	1.5317	1.5708	1.5708
0.4	0.6859	0.9827	1.3266	1.5704	1.5708
0.5	0.6409	0.8684	1.1454	1.4738	1.5708
0.6	0.6168	0.8196	1.0224	1.2672	1.5331
0.7	0.3426	0.6797	0.9852	1.1093	1.5395
0.8	0.3898	0.6855	0.9197	1.0352	1.2386
0.9	0.1334	0.4805	0.7115	0.9169	1.2745
1.0	0.1332	0.3363	0.5102	0.8244	1.1008

In Fig. 9, the voltage waveforms of the load are shown for modulation indices of 0.4, 0.6, and 1.0. The THD and THDe analyses of the waveforms are presented in Figures 10 and 11, respectively. For a modulation index of 0.4, the THD of the load voltage is calculated to be 12.27%, with a maximum value of  $V_{an(max)} = 124V$  and an RMS value of  $V_{an(rms)} = 87.71V$  (Fig. 10(a)). For a modulation index of 0.6, the THD of the load voltage is calculated to be 6.84%, with a maximum value of  $V_{an(max)} = 185.9V$  and an RMS value of  $V_{an(rms)} = 131.5V$  (Fig. 10(b)). Finally, for a modulation index of 1.0, the THD of the load voltage is determined to be 4.48%, with a maximum value of  $V_{an(max)} = 310.8V$  and an RMS value of  $V_{an(rms)} = 219.8V$  (Fig. 10(c)).

The degree of suppression of the selected harmonics is shown in Fig. 11. For a modulation index of 0.4, the THDe is calculated to be 3.10%, with harmonic magnitudes of  $h_5 = 2.32$  ( $V_5 = 2.88V$ ),  $h_7 = 0.71$  ( $V_7 = 0.89V$ ),  $h_{11} = 1.89$  ( $V_{11} = 2.34V$ ), and  $h_{13} = 1.89$  ( $V_{13} = 2.34V$ ). For a modulation index of 0.6, the THDe is determined to be 0.09%, with harmonic magnitudes of  $h_5 = 0.05$  ( $V_5 = 0.08V$ ),  $h_7 = 0.04$  ( $V_7 = 0.07V$ ),  $h_{11} = 0.02$  ( $V_{11} = 0.04V$ ), and  $h_{13} = 0.02$  ( $V_{13} = 0.03V$ ). Finally, for a modulation index of 1.0, the THDe is found to be 0.04%, with harmonic magnitudes of  $h_5 = 0.02$  ( $V_5 = 0.05V$ ),  $h_7 = 0.01$  ( $V_7 = 0.03V$ ),  $h_{11} = 0.01$  ( $V_{11} = 0.02V$ ), and  $h_{13} = 0.01$  ( $V_{13} = 0.04V$ ).

TABLE V  
SIMULATION RESULTS (FOR 11 LEVELS)

$M$	Vref(peak)	Vref(rms)	V(rms)	Error (%)	THD (%)	THDe (%)	$h_5$	$h_7$	$h_{11}$	$h_{13}$
0.10	31.1	22	21.91	<b>0.41</b>	59.35	55.17	46.09	11.48	0.31	17.01
0.20	62.2	44	43.84	<b>0.36</b>	27.64	18.27	12.34	2.88	12.38	2.19
0.30	93.3	66	65.74	<b>0.39</b>	19.34	8.12	4.32	3.26	0.01	6.06
0.40	124.4	88	87.66	<b>0.39</b>	12.77	3.10	2.32	0.71	1.89	0.39
0.50	155.5	110	109.2	<b>0.73</b>	9.05	<b>0.95</b>	<b>0.47</b>	<b>0.25</b>	<b>0.70</b>	<b>0.34</b>
0.60	186.6	132	131.5	<b>0.38</b>	6.48	<b>0.09</b>	<b>0.05</b>	<b>0.04</b>	<b>0.02</b>	<b>0.02</b>
0.70	217.7	154	153.3	<b>0.45</b>	8.10	<b>0.05</b>	<b>0.02</b>	<b>0.02</b>	<b>0.03</b>	<b>0.01</b>
0.80	248.8	176	175.2	<b>0.45</b>	6.72	<b>0.03</b>	<b>0.01</b>	<b>0.00</b>	<b>0.02</b>	<b>0.01</b>
0.90	279.9	198	197.2	<b>0.40</b>	6.31	<b>0.05</b>	<b>0.02</b>	<b>0.00</b>	<b>0.01</b>	<b>0.02</b>
1.00	311.0	220	219.8	<b>0.09</b>	4.49	<b>0.03</b>	<b>0.02</b>	<b>0.01</b>	<b>0.01</b>	<b>0.01</b>

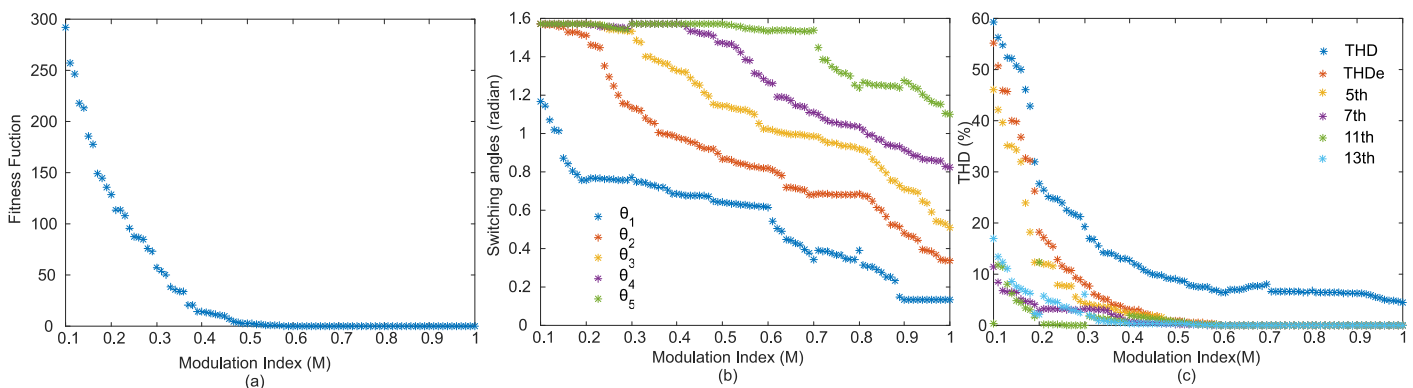


Fig.8. Analysis of the 11-level MLI inverter in terms of (a) fitness function, (b) switching angles, and (c) THD, THDe, 5<sup>th</sup>, 7<sup>th</sup>, 11<sup>th</sup> and 13<sup>th</sup> harmonic characteristics as a function of M.



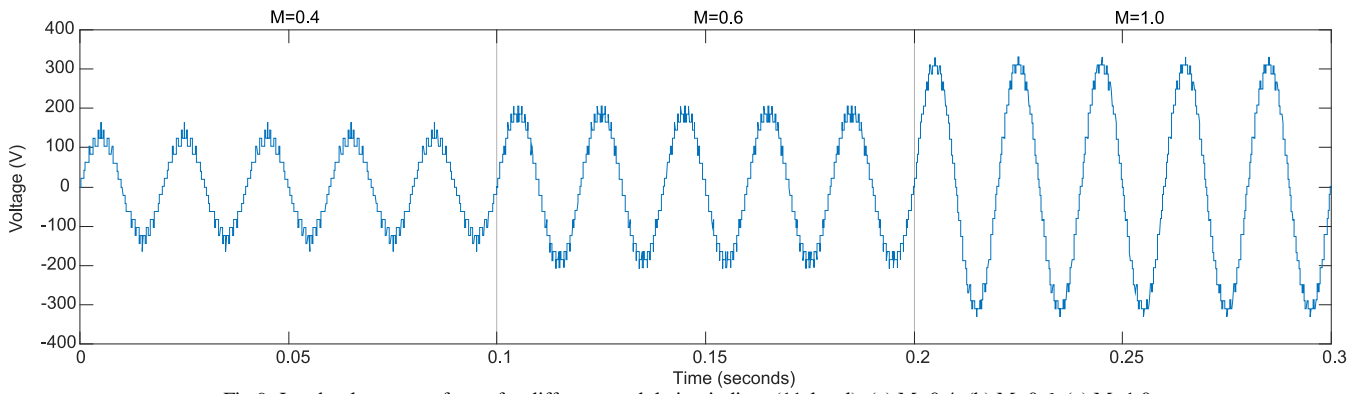


Fig.9. Load voltage waveforms for different modulation indices (11-level) (a) M=0.4, (b) M=0.6, (c) M=1.0

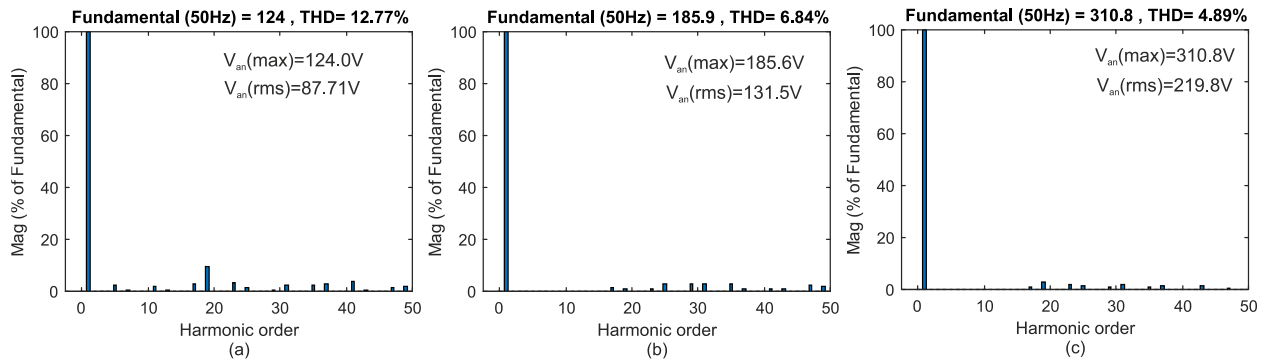


Fig.10. THD values for different modulation indices (11-level) (a) M=0.4, (b) M=0.6, (c) M=1.0

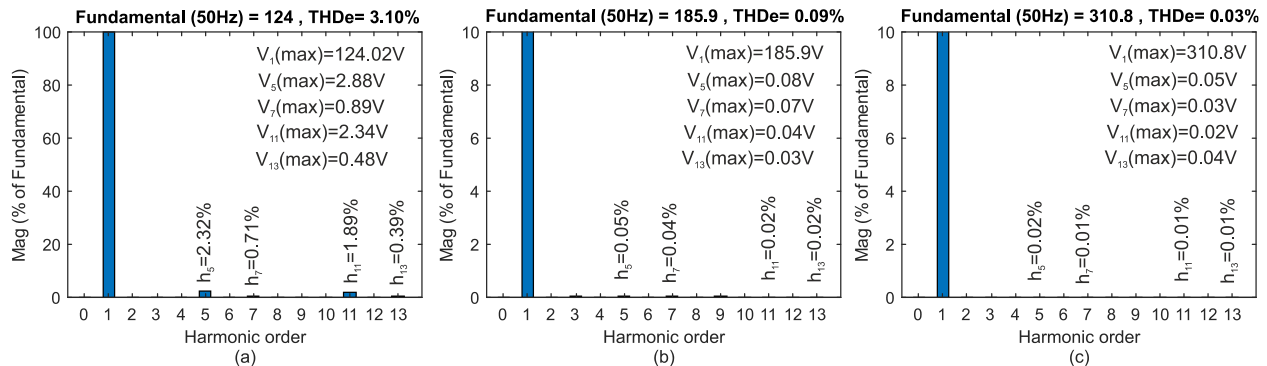


Fig.11. THDe values for different modulation indices (11-level) (a) M=0.4, (b) M=0.6, (c) M=1.0

IV. CONCLUSION

This study utilized the recently developed AVOA (Artificial Virus Optimization Algorithm) to tackle the SHE-PWM (Selective Harmonic Elimination Pulse Width Modulation) issue in multilevel inverters. The effectiveness of AVOA in resolving SHE-PWM was showcased through MATLAB simulations conducted on 7 and 11 level cascade H-bridge inverters. The algorithm has demonstrated its capability to find suitable solutions within the M range of 0.1 to 1.0. The desired reference voltage has been effectively controlled with an error of less than 0.5%. Furthermore, the selected harmonics have been efficiently suppressed within the modulation range of 0.6 to 1.0, as indicated by the THDe values and the magnitudes of

the 5<sup>th</sup>, 7<sup>th</sup>, 11<sup>th</sup>, and 13<sup>th</sup> harmonics, all of which are below 0.09%.

The waveform analysis of the output voltage for different modulation indices (0.4, 0.6, and 1.0) has shown that the THD values decrease as the modulation index increases. The maximum values of the fundamental voltage and its RMS value have also been determined for each modulation index, providing important insights into the performance of the system. Overall, the AVO algorithm has proven its effectiveness in achieving accurate voltage control and harmonic suppression in multilevel inverters, thereby contributing to the improvement of power quality and efficient operation of the system.

## REFERENCES

- [1] A. Raki, Y. Neyshabouri, M. Aslanian, & H. Iman-Eini, "A Fault-Tolerant Strategy for Safe Operation of Cascaded H-Bridge Multilevel Inverter under Faulty Condition". *IEEE Transactions on Power Electronics*.
- [2] B. Karthikeyan, B. Swetha, C. Shahana, B. Vijay, & M. Swathi, "Switched Capacitor Based Multilevel Inverter for/newline PV System". In 2023 9th International Conference on Electrical Energy Systems (ICEES) (pp. 471-475). IEEE, March, 2023.
- [3] N. Padmamalini, P. Deepa, T. Joel, S. Loganayagi, M. F. Banu, & S. Gomathi, "Control of Diode Clamped Multilevel Inverter based STATCOM for Reactive Power Compensation using H-bridge Topology". In 2023 7th International Conference on Intelligent Computing and Control Systems (ICICCS) (pp. 1860-1865). IEEE.
- [4] H. I. Hazim, K. A. Baharin, C. K. Gan, A. H. Sabry, & A. J. Humaidi, Review on optimization techniques of PV/inverter ratio for grid-Tie PV systems. *Applied Sciences*, 13(5), 3155, 2023.
- [5] C. Mayet, D. Labrousse, R. Bkekri, F. Roy, & G. Pongnot, "Energetic Macroscopic Representation and Inversion-Based Control of a Multi-Level Inverter with Integrated Battery for Electric Vehicles". In 2021 IEEE Vehicle Power and Propulsion Conference (VPPC) (pp. 1-6). IEEE, October, 2021.
- [6] C. Pisani, G. Bruno, H. Saad, P. Rault, & B. Clerc, "Functional validation of a real VSC HVDC control system in black start operation", In 2019 AEIT HVDC International Conference (AEIT HVDC) (pp. 1-6). IEEE, May, 2019.
- [7] H. Lin, R. Chen, R. Li, L. Zhu, H. Yan, & Z. Shu, "A flexible and fast space vector pulse width modulation technique for multilevel converters", In 2019 22nd International Conference on Electrical Machines and Systems (ICEMS) (pp. 1-4). IEEE, August, 2019.
- [8] S. A. Azmi, A. A. Shukor, & S. R. A. Rahim, "Performance evaluation of single-phase H-bridge inverter using selective harmonic elimination and sinusoidal PWM techniques", In 2018 IEEE 7th International Conference on Power and Energy (PECon) (pp. 67-72). IEEE, December, 2018.
- [9] S. E. Arslan, F. E. Uzun, K. Gürkan, A. Acar, İ. O. Yildirim, U. Güven, ... & S. Yarman, "Realization of SV-PWM motor control algorithm using ARM Cortex-M4 based microcontroller", In 2016 National Conference on Electrical, Electronics and Biomedical Engineering (ELECO) (pp. 282-285). IEEE, December, 2016.
- [10] E. Bektas, & H. Karaca, "GA based selective harmonic elimination for multilevel inverter with reduced number of switches: an experimental study", *Elektronika ir Elektrotehnika*, 25(3), 10-17, 2019.
- [11] A. Pourdashnia, M. Farhadi-Kangarlu, B. Tousi, & M. Sadoughi, "SHM-PWM technique in a cascaded H-bridge multilevel inverter with adjustable DC-link for wide voltage range applications", *International Journal of Circuit Theory and Applications*, 51(5), 2228-2246, 2023.
- [12] S. Pani, N. Guru, D. Puhana, & A. K. Barisal, "Comparative Performance analysis of Multilevel Inverter through meta heuristics", In 2022 6th International Conference on Intelligent Computing and Control Systems (ICICCS) (pp. 129-135). IEEE, May, 2022.
- [13] B. G. Babu & M. S. Kalavathi, "Hardware Implementation of Multilevel Inverter using NR, GA, Bee Algorithms" In 2021 International Conference on Sustainable Energy and Future Electric Transportation (SEFET) (pp. 1-6). IEEE, January, 2021.
- [14] Y. Bektaş, & H. Karaca, "Red deer algorithm based selective harmonic elimination for renewable energy application with unequal DC sources" *Energy Reports*, 8, 588-596, 2022.
- [15] B. Abdollahzadeh, F. S. Gharehchopogh, & S. Mirjalili, African vultures optimization algorithm: A new nature-inspired metaheuristic algorithm for global optimization problems", *Computers & Industrial Engineering*, 158, 107408, 2021.
- [16] H. Karaca, & E. Bektas, "Selective Harmonic Elimination Technique Based on Genetic Algorithm for Multilevel Inverters", In *Transactions on Engineering Technologies: World Congress on Engineering and Computer Science 2015* (pp. 333-347). Springer Singapore, 2017
- [17] M. YEŞİLBUDAK, "Extraction of photovoltaic cell and photovoltaic module parameters using african vultures optimization algorithm" *Gazi University Journal of Science Part C: Design and Technology*, 9(4), 708-725.
- [18] M. E. Bento, "Design of a Resilient Wide-Area Damping Controller Using African Vultures Optimization Algorithm", In 2021 31st Australasian Universities Power Engineering Conference (AUPEC) (pp. 1-6). IEEE, September, 2021.
- [19] H. E. ALOUACHE, S. SAYAH, & A. HAMOUDA, "Africa vultures optimization algorithm for optimal power flow solution including SVC devices", In 2022 19th International Multi-Conference on Systems, Signals & Devices (SSD). IEEE, May, 2022
- [20] R. Liu, T. Wang, J. Zhou, X. Hao, Y. Xu, & J. Qiu, "Improved African vulture optimization algorithm based on quasi-oppositional differential evolution operator", *IEEE Access*, 10, 95197-95218, 2022.
- [21] D. Gürses, P. Mehta, S. M. Sait, & A. R. Yildiz, African vultures optimization algorithm for optimization of shell and tube heat exchangers", *Materials Testing*, 64(8), 1234-1241, 2022.
- [22] J. Zhang, M. Khayatnezhad, & N. Ghadimi, "Optimal model evaluation of the proton-exchange membrane fuel cells based on deep learning and modified African Vulture Optimization Algorithm". *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, 44(1), 287-305, 2022.
- [23] M. Ahmed, M. Khamies, G. Magdy, & S. Kamel, "Designing Optimal PI λ D μ Controller for LFC of Two-Area Power Systems Using African Vulture's Optimization Algorithm", In 2021 22nd International Middle East Power Systems Conference (MEPCON) (pp. 430-437). IEEE, December, 2021.
- [24] S. Mil'shtein, & D. N. Asthana, "Brief Comparison of Tandem and Cascaded Solar Cells", In 2021 IEEE 48th Photovoltaic Specialists Conference (PVSC) (pp. 2260-2263). IEEE, June, 2021.
- [25] F. S. Gharehchopogh, "An improved Harris Hawks optimization algorithm with multi-strategy for community detection in social network" *Journal of Bionic Engineering*, 20(3), 1175-1197, 2023.
- [26] T. Sonmezocak, O. Akar, & U. K. Terzi, "HIGH PERFORMANCE ADAPTIVE ACTIVE HARMONIC FILTER DESIGN FOR NON-LINEAR LED LOADS" *Light & Engineering*, 30(1), 29-38, 2022.
- [27] A. Demirci, O. Akar, U. K. Terzi, & T. Sönmezocak, "Investigation of International Harmonic Standards in Power Systems", In 4th International Mardin Artuklu Scientific Research Congress (pp. 7-8), 2020, August.

## BIOGRAPHIES



**YASİN BEKAŞ** born in Sivas, Turkey, in 1982. He received a B.S. degree in Electrical Education from Marmara University, Istanbul, Turkey, in 2008, an M.S. degree in Electrical Education from Marmara University, Istanbul, Turkey, in 2011. His MSc thesis

about analysis, development, and computer-based study of Brushless Direct Current (BLDC) motor and its driver circuit training set. He received a B.S. degree in Electrical and Electronics Engineering from Selçuk University, Konya, Turkey, in 2015. He got a PhD degree from Selçuk University, Department of Electrical and Electronics in 2023. He is currently as lecturer at the Electrical and Energy Department of Aksaray University Technical Sciences Vocational High School. His current research interests include power systems, renewable energy, and multi-level inverters. He can be contacted at email: [yasinbektas@aksaray.edu.tr](mailto:yasinbektas@aksaray.edu.tr).

# Designing Effective Models for COVID-19 Diagnosis through Transfer Learning and Interlayer Visualization

Cuneyt Ozdemir

**Abstract**— Creating an optimal model for a specific dataset can be challenging and time-consuming. This article presents an innovative approach to model creation by leveraging transfer learning models and employing the interlayer visualization method. The study aims to overcome the complexities of designing a new model specifically for the COVID-19 dataset. Transfer learning models, which are pre-trained models, offer a practical solution due to their adaptability. However, not all layers in these models may be suitable for a given dataset, emphasizing the importance of identifying and removing unnecessary layers for successful model performance.

Experimental studies were conducted using transfer learning models, including Densenet201 and InceptionV3, on the COVID-19 dataset. The interlayer visualization method was utilized to identify irrelevant layers, resulting in the creation of new models. The evaluation metrics demonstrated that the derived models outperformed the original transfer learning models. The Mixed3 model derived from InceptionV3 achieved a higher accuracy of 98.6%, compared to the original model's accuracy of 98.3%. Similarly, the Pool4 model derived from DenseNet201 achieved an accuracy of 98.8%, surpassing the original model's accuracy of 98.19%. Moreover, the sensitivity for detecting images belonging to the Covid-19 class reached 98.9% and 99.7% for the Mixed3 and Pool4 models, respectively.


The new models obtained through the interlayer visualization method offer several advantages, including lightweight design, faster training processes, and improved performance. This study highlights the effectiveness of leveraging transfer learning models with the interlayer visualization method for creating robust models tailored to specific datasets. The proposed approach serves as a valuable solution to the challenges associated with model creation, particularly in the context of COVID-19 diagnosis using medical imaging data.

**Index Terms**—Covid-19, DenseNet201, InceptionV3, Interlayer visualization, Model pruning, Transfer learning models

## I. INTRODUCTION

TRANSFER LEARNING is a method in machine learning, utilizes pre-trained models to perform new tasks. When designing a new deep neural network architecture, transfer learning models are commonly employed [1]. This approach

CÜNEYT ÖZDEMİR is with Department of Computer Engineering University of Siirt University, Siirt, Turkey, (e-mail: [cozdemir@siirt.edu.tr](mailto:cozdemir@siirt.edu.tr)).

 <https://orcid.org/0000-0002-9252-5888>

Manuscript received Mar 31, 2023; accepted Aug 20, 2023.  
DOI: [10.17694/bajece.1274253](https://doi.org/10.17694/bajece.1274253)

allows the adaptation of a pre-trained model's general features to a new task, proving particularly useful when data scarcity or large datasets hinder training a model from scratch. Transfer learning leverages features learned from one dataset for another, leading to superior results with reduced training data [2].

Transfer learning finds extensive applications in various domains, including agriculture, healthcare, image processing, natural language processing, and speech recognition [3-7]. However, not all transfer learning models are suitable for every dataset, and some models may contain unnecessary layers. Therefore, selecting the most appropriate transfer learning model for a specific dataset and identifying the optimal layers within that model are crucial for achieving successful outcomes.

The size and characteristics of the dataset influence the selection of an appropriate model. Larger datasets benefit from complex and deep models, while smaller datasets are better suited for simpler models. Hence, the choice of transfer learning models should align with the specific characteristics of the dataset.

Despite the advantages of transfer learning, there are certain drawbacks to consider. Firstly, not all layers in a pre-trained model are necessary for a new task, and some layers may introduce redundancy or contribute to overfitting depending on the dataset's characteristics [8]. Therefore, it is essential to remove or modify unnecessary layers to enhance the performance of the pre-trained model. Adapting pre-trained models to match the size and complexity of the dataset is crucial, and having a clear understanding of which layers to modify or remove is essential for optimizing model performance.

In some cases, employing all layers of a pre-trained model can result in longer training times or reduced model efficiency. Thus, pre-trained models should be adaptable to the dataset's size and complexity. Nevertheless, transfer learning models require less time and computational power compared to training a model from scratch for a new task. Moreover, pre-trained models typically capture more general features due to their training on larger datasets. These features can be fine-tuned for the new task, resulting in a more accurate and efficient model.

To address unnecessary or redundant information in transfer learning models, the inter-layer visualization technique, known as model pruning, is employed. Model pruning aims to eliminate unnecessary parameters or layers from a model. It

identifies redundant layers by analyzing the model's layers or filters, eliminating similar or weak activations. Model pruning reduces the model's size, shortens training time, and improves its generalization ability.

Zeiler et al. [9] utilized the AlexNet deep learning model to visualize how each layer learns its features and represents each feature map. This visualization process aimed to comprehend the combination of lower-level features learned by certain layers in higher-level layers, leading to the production of the final output. The study investigated the impact of removing specific layers on the model's performance, revealing a significant drop in performance when the first two layers were removed. This result demonstrated that the initial layers of the model learn lower-level features, while higher-level layers learn higher-level features through their integration.

Urban et al. [10] investigated the effect of removing layers in deep learning models on the model's performance. The study showed that, in certain cases, the number of layers in deep learning models could be reduced, and some models achieved high performance even with a few layers.

Bau et al. [11] introduced an approach to comprehend the features of layers within deep learning models. This approach introduced the concept of "neuron activation" to quantify the relevance of each feature to a specific object or concept.

Li et al. [12] developed a method for visualizing the loss landscape of various deep learning models. This method enables an understanding of the impact of removing or modifying layers on model performance by visually demonstrating the effects of different layers in the model.

This study aims to create new models and enhance their performance by conducting inter-layer visualization on transfer learning models. The inter-layer visualization method can identify layers that carry less information by visualizing the outputs of each layer in the model. When a model is provided with input, inter-layer visualization visualizes the activations generated by each layer of the model. These visualizations assist in determining the importance of each layer and which layers contribute more significantly to the model's performance.

COVID-19 X-ray images were employed to examine the performance of model pruning operations on transfer learning models. COVID-19, a viral disease caused by SARS-CoV-2, was initially identified in Wuhan, China. While the initial cases were reported in Wuhan, the precise origins and transmission of the virus remain subjects of ongoing scientific research and investigation.

COVID-19 primarily affects the respiratory system and belongs to the coronavirus family. It is a novel virus for humans, posing a significant risk, particularly to elderly individuals and those with chronic illnesses. COVID-19 can induce various symptoms, including fever, cough, fatigue, shortness of breath, muscle aches, headaches, and loss of taste or smell, among others. The outbreak of this disease rapidly escalated into a major worldwide health crisis, profoundly impacting healthcare systems and economies of numerous countries. Consequently, COVID-19 continues to be a globally-relevant research topic [13-14].

The health effects of COVID-19 are not limited to the

respiratory system. The virus can cause brain inflammation, psychological disorders, cardiovascular diseases, kidney failure, and damage to other organs. PCR or antigen tests are commonly employed for COVID-19 diagnosis. With the COVID-19 pandemic becoming a global health issue, artificial intelligence techniques, such as deep learning, have started to play a crucial role in addressing this problem [15,16].

Ertugrul et al. [17] proposed a machine learning-based automatic diagnosis system that utilizes X-ray images to detect COVID-19-related diseases. The system employed robust texture features, such as Histogram of Oriented Gradients, Law's Tissue Energy Measure, Gabor Wavelet Transform, Gray Level Co-occurrence Matrix, and Local Binary Pattern, to train a random neural network. This approach facilitated a rapid and robust diagnostic process for COVID-19 by extracting identifying indicators from the two-dimensional space of X-ray images from virus patients.

Kaya et al. [18] presented a novel approach for COVID-19 detection using X-ray images. They introduced the Angle Transform (AT) method, which captures the angle information between pixels in the images. The AT method generated eight different images per dataset image, which were then used to train a hybrid deep learning model combining GoogleNet and LSTM. The proposed approach achieved a high classification accuracy of 98.97% and demonstrated success in COVID-19 detection using chest X-ray images.

This study involves the creation of new models through model pruning operations on transfer learning models using the COVID-19 dataset. Experimental studies have been conducted using these newly obtained models, and the performance of transfer learning models and the new models derived from them through model pruning methods has been observed.

## II. DATASET AND PREPROCESSING

### A. Dataset

A group of researchers from Doha, Qatar and Dhaka University, Bangladesh have shared a dataset that contains images. These images consist of chest X-rays for COVID-19 positive cases, Normal images, and Viral Pneumonia images. Table I provides further details regarding the dataset [19].

TABLE I  
Distribution of images according to classes

Image Class	Number
COVID 19	3616
Normal	10192
Viral Pneumonia	1345

The distribution of visual data among different classes exhibits an imbalance, as illustrated in Table I. This imbalance can lead to incorrect learning by the model and subsequently generate inaccurate results. In an effort to mitigate this issue, only 3500 normal images from the dataset were utilized. The distribution of the dataset is depicted in Figure 1, encompassing a total of 8461 images.



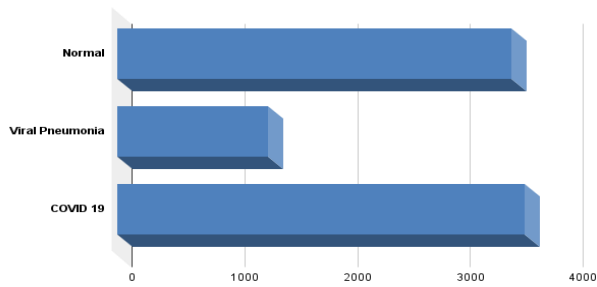


Fig.1. Dataset distribution

Figure 2 displays image examples from three distinct classification clusters within the dataset. These image samples were obtained following the application of data augmentation techniques on the images.

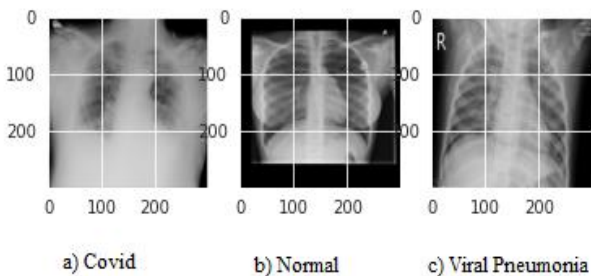


Fig 2. Examples of 3 different image classes

Several studies have been conducted in the literature using this dataset:

Aslan et al. [20] conducted a study where they utilized features extracted from various well-known Convolutional Neural Network (CNN) models, such as AlexNet, Inceptionresnetv2, ResNet18, Inceptionv3, ResNet50, MobileNetv2, Densenet201, and GoogleNet. These extracted features were then used for classification using different machine learning algorithms, including SVM, k-NN, Naive Bayes, and Decision Tree. The DenseNet + SVM method achieved a classification accuracy of 96.29%.

Sohan et al. [21] employed ResNet-18 and VGG16 transfer learning models on the same dataset, achieving an accuracy of 97% in their study.

Sahlol et al. [22] proposed a hybrid classification approach that utilized the Marine Predators algorithm to select a swarm-based feature from the InceptionV3 transfer learning model. In their study, they achieved an accuracy of 98.7% and an F-score of 98.2%.

Abdollahi et al. [23] achieved an accuracy of 97.99% in their study using the VGG16 transfer learning model.

Abdrakhmanov et al. [24] employed few-shot learning techniques to classify images with a small amount of training data and achieved a classification accuracy of 97.7%.

These studies demonstrate the effectiveness of various transfer learning models and approaches in achieving high accuracies on the dataset.

### B. Pre Processing

Before being inputted into the model, the images underwent pre-processing steps. During this stage, the image dimensions were resized to 224x224 pixels. To prevent overfitting and enhance the diversity of the dataset for improved model performance, data augmentation techniques were employed. The data augmentation process included applying a 15% rotation, 10% shifting to the right and left, horizontal flipping, and a 20% zoom in and out of the images.

Following the pre-processing and data augmentation steps, the images were split into three distinct sets: training, validation, and testing. Specifically, 15% of the images were allocated for testing, while the remaining 85% were assigned to the training set. Within the training set, an additional 15% of the images were set aside for validation purposes.

### III. METHODOLOGY

In order to address the challenges and time constraints associated with building a model from scratch for our specific dataset, we turned to the powerful technique of transfer learning. Transfer learning involves utilizing the knowledge and features learned by a pre-trained model on a different task or dataset and applying it to a new task or dataset, aiming to achieve improved performance.

Transfer learning brings several advantages to the table, as it allows us to leverage the wealth of information captured by pre-trained models and adapt them to our specific task. By employing transfer learning, we aimed to overcome the difficulties inherent in starting from scratch and enhance the performance of our model on our dataset.

In our experimental studies, the primary objective was to identify the models that exhibited the best performance on our dataset. To this end, we employed transfer learning models, and the performance metrics of these models when applied to our dataset can be found in Table II. To optimize the training process, we utilized the Adam optimization function with a fixed learning rate of 0.0003. Furthermore, we implemented the ReduceLROnPlateau and EarlyStopping methods, which dynamically reduce the learning rate and automatically halt the training of the model when the accuracy fails to improve over a certain number of epochs.

The utilization of transfer learning, along with the aforementioned optimization techniques, enabled us to effectively leverage pre-existing knowledge and adapt it to our specific task. This approach not only accelerated the training process but also allowed us to achieve superior results on our dataset. The detailed performance metrics of the transfer learning models applied in our experimental studies are presented in Table II, providing insights into the effectiveness of our approach.

TABLE II  
PERFORMANCE OF TRANSFER LEARNING MODELS

	Accuracy	Precision	Recal	F1-Score
Xception	0.9779	0.9784	0.9779	0.9779
<b>InceptionV3</b>	<b>0.9832</b>	0.9832	0.9832	0.9832
ResNet50V2	0.9799	0.9799	0.9799	0.9799
<b>DenseNet201</b>	<b>0.9819</b>	0.9819	0.9819	0.9819

According to Table II, the InceptionV3 and DenseNet201 transfer learning models achieved the best results for the dataset. After identifying these models through initial experimental studies, model pruning was conducted on them. Model pruning involves removing unnecessary layers from transfer learning models that are not essential for the dataset. Various methods can be employed for model pruning. In this study, the inter-layer visualization or data visualization method was utilized. The inter-layer visualization method generates visual representations of the output produced by each layer involved in processing the input image of an image classification model. By analyzing the output of each layer, we gain insights into the features learned by the model at each stage and how these features are propagated to subsequent layers. For

example, it helps us understand that the initial layers generally learn object contours or colors, while deeper layers focus on more specific object features. This information assists in identifying the layers where learning occurs and those where it does not.

In the inter-layer visualization method, the appearance of an image in all layers of the DenseNet201 and InceptionV3 models trained on the dataset was examined. Through this examination, the layers in which learning occurred were determined in these two models, and the layers where learning did not occur were removed from the models. Figure 3 illustrates the image appearance at different layers of the InceptionV3 model for the dataset. As depicted in Figure 3, learning did not occur after the mixed3 layer.

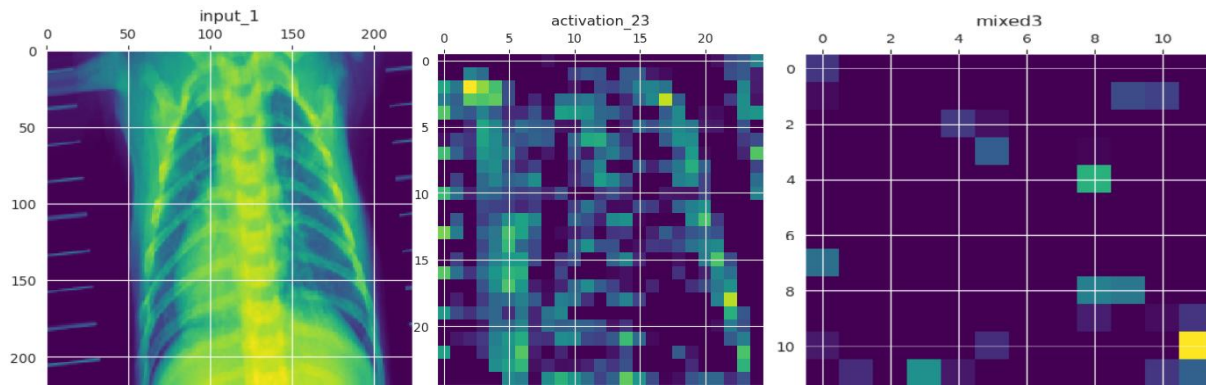


Fig 3. Examples of 3 different image classes

As a result of the process, it was observed that learning occurred up to the mixed3 layer in the InceptionV3 model and up to the pool4 layer in the DenseNet201 model. Based on this observation, new and modified models were created from the transfer learning models.

In the Mixed3 model, only the layers up to the mixed3 layer of the InceptionV3 model were utilized, and the remaining layers were discarded. Similarly, in the Pool4 model, only the layers up to the pool4 layer from the DenseNet201 model were retained, and the rest were removed to obtain a new model.

Figure 4 illustrates the structure of the new Convolutional Neural Network (CNN) model obtained through inter-layer image visualization from the InceptionV3 transfer learning model. The mixed3 model was used for feature extraction in the new model, which were subsequently fed into an artificial neural network. The neural network consists of two layers with 256 and 128 neurons, respectively. The output layer of the artificial neural network consists of three classes, and the Softmax activation function was employed.

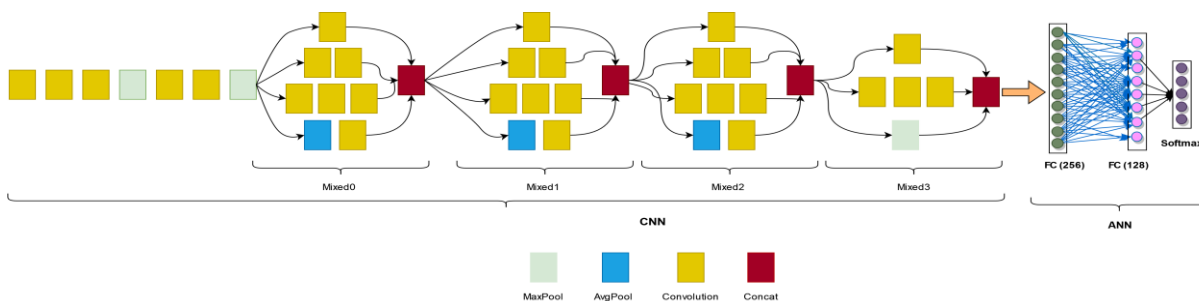


Fig 4. Mixed3 CNN model

After the new models were obtained, experimental studies were performed. The results of the experimental studies with Mixed3 and Pool4 models are shown in Table III. The train and validation, loss and accuracy graphs of the Mixed3 model are shown in Figure 5. The confusion matrix of the Mixed3 model is shown in Figure 6 and the classification report is shown in Table IV.

TABLE III  
ACCURACY RESULTS OF NEW MODELS

	Train Acc	Val Acc	Test Acc
Mixed3	0.998	0.979	0.986
Pool4	0.999	0.988	0.988

TABLE IV  
MIXED3 CLASSIFICATION REPORT

	Precision	Recall	F1-score
Normal	0.98	0.986	0.983
Viral Pneumonia	0.99	0.985	0.988
Covid	0.989	0.986	0.989
accuracy			0.986
macro avg	0.987	0.986	0.986
weighted avg	0.986	0.986	0.986

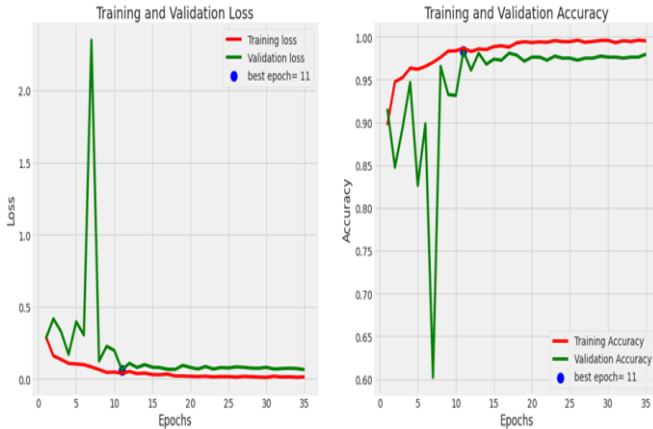


Fig 5. Loss and accuracy graph for Mixed3 model

The evaluation of classification performance was conducted using various metrics, including accuracy, precision, recall, and F1 score for each class. Additionally, the classification report provides average precision, recall, and F1 score, offering an overall assessment of the model's performance across all classes. Table IV demonstrates that the Mixed3 model derived from the InceptionV3 model achieved superior results. While the InceptionV3 model attained an accuracy rate of 98.3% on the same dataset, the Mixed3 model achieved a success rate of 98.6%. Notably, images belonging to the Covid-19 class were detected with a sensitivity of 98.9%. The classification report for the Pool4 model derived from the DenseNet201 model is displayed in Table V. The DenseNet201 model achieved an accuracy rate of 98.19%, while the Pool4 model achieved 98.8% accuracy. Remarkably, images belonging to the Covid-19 class were detected with 99.7% accuracy.

Figure 7 illustrates the train and validation loss and accuracy graph of the Pool4 model, providing insights into the model's performance during the training process. Furthermore, Figure 8 presents the confusion matrix for the Pool4 model, allowing for a visualization of the model's performance in terms of correct and incorrect predictions for each class.

TABLE V  
POOL4 CLASSIFICATION REPORT

	Precision	Recall	F1-score
Normal	0.975	0.997	0.986
Viral Pneumonia	1	0.963	0.981
Covid	0.997	0.989	0.993
accuracy			0.988
macro avg	0.991	0.983	0.987
weighted avg	0.988	0.988	0.988

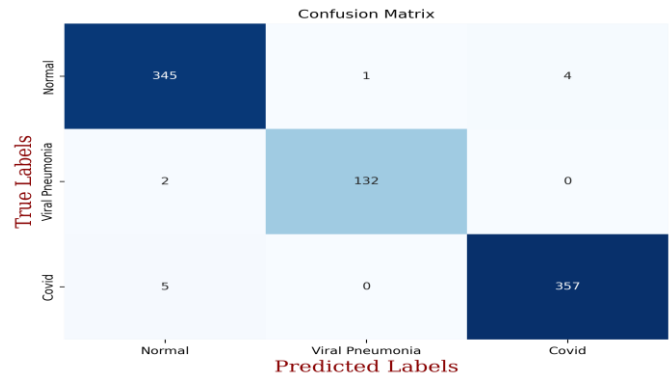


Fig 6. Confusion matrix of the Mixed3 model

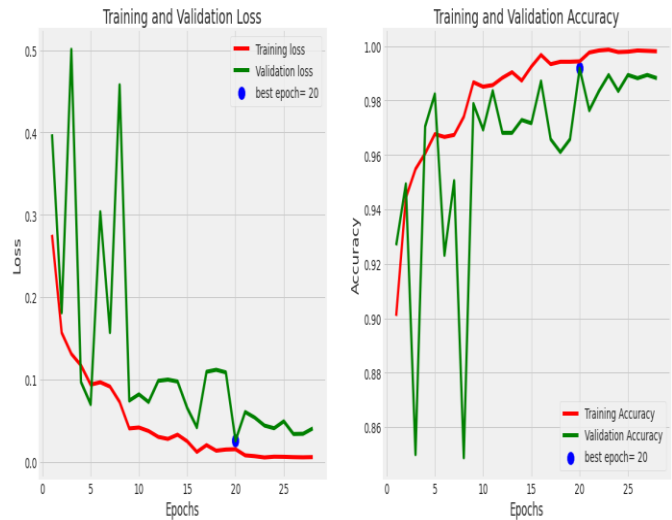


Fig 7. Loss and accuracy graph for Pool4 model

The experimental studies demonstrated that the new models derived from transfer learning models yielded more successful results. The InceptionV3 model achieved an accuracy of 98.32%, while the Mixed3 model achieved an even higher accuracy of 98.6%. Similarly, the DenseNet201 model achieved an accuracy of 98.19%, while its derived Pool4 model achieved a higher success rate of 98.8%. These results indicate that lightweight and fast models obtained through the interlayer image visualization method can yield more successful outcomes.

IV. CONCLUSIONS

This study proposes an approach that utilizes the interlayer visualization method from model pruning techniques to derive new models from transfer learning models, instead of designing a new model specifically for the COVID-19 dataset. The aim is to address the challenges associated with creating a new model, which is often time-consuming and laborious. By leveraging transfer learning models, the focus is on adapting them to the dataset and creating a new model. It is crucial to identify the layers in transfer learning models that may not extract relevant features from the dataset, as this significantly impacts the model's success and complexity.

To evaluate the performance, transfer learning models were employed on the COVID-19 dataset, rather than building a

model from scratch. In these experimental studies, the Densenet201 and InceptionV3 transfer learning models achieved the highest scores. The interlayer visualization method was utilized to identify unnecessary layers in the transfer learning models. The experimental studies conducted with the derived new models demonstrated more successful results.

The new models created using the interlayer visualization method are lightweight, fast, and have fewer parameters. They facilitate faster training processes and yield improved performance in terms of model accuracy. The InceptionV3 model achieved an accuracy rate of 98.3%, while the Mixed3 model derived from it achieved a higher accuracy of 98.6%. Moreover, images belonging to the Covid-19 class were detected with a sensitivity of 98.9%. Similarly, the DenseNet201 model achieved an accuracy rate of 98.19%, and the Pool4 model derived using the interlayer visualization method achieved an even higher accuracy of 98.8%. Images belonging to the Covid-19 class were detected with a sensitivity of 99.7%.

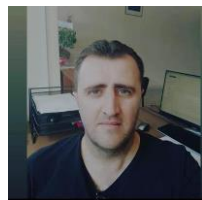
This study demonstrates that the interlayer visualization method can effectively design the most suitable CNN model for a given dataset by creating new models from transfer learning models. Leveraging the interlayer visualization method provides an effective approach to overcome the challenges associated with creating a new model.

#### REFERENCES

- [1] H. E. Kim, A. Cosa-Linan, N. Santhanam, M. Jannesari, M. E. Maros, and T. Ganslandt, 'Transfer learning for medical image classification: a literature review', *BMC Med. Imaging*, vol. 22, no. 1, p. 69, Apr. 2022.
- [2] S. Atasever, N. Azginoglu, D. S. Terzi, and R. Terzi, 'A comprehensive survey of deep learning research on medical image analysis with focus on transfer learning', *Clin. Imaging*, vol. 94, pp. 18–41, Feb. 2023.
- [3] H. S. Nogay, T.C. Akinci, & M. Yilmaz, "Comparative experimental investigation and application of five classic pre-trained deep convolutional neural networks via transfer learning for diagnosis of breast cancer". *Advances in Science and Technology*, 15(3), 2021.
- [4] Y. Dogan and H. Y. Keles, 'Iterative facial image inpainting based on an encoder-generator architecture', *Neural Comput. Appl.*, vol. 34, no. 12, pp. 10001–10021, Jun. 2022.
- [5] M. Ataş, C. Özdemir, İ. Ataş, B. Ak, and E. Özeröglü, 'Biometric identification using panoramic dental radiographic images with few-shot learning', *TURK. J. OF ELECTR. ENG. COMPUT. SCI.*, vol. 30, no. 3, pp. 1115–1126, Jan. 2022.
- [6] Y. Dogan and H. Yalim Keles, 'Stability and diversity in generative adversarial networks', in 2019 27th Signal Processing and Communications Applications Conference (SIU), Sivas, Turkey, 2019.
- [7] C. Ozdemir, M. A. Gedik, and Y. Kaya, 'Age estimation from left-hand radiographs with deep learning methods', *Trait. Du Signal*, vol. 38, no. 6, pp. 1565–1574, Dec. 2021.
- [8] M. Iman, K. Rasheed, and H. R. Arabnia, 'A review of Deep Transfer Learning and recent advancements', *arXiv [cs.LG]*, 18-Jan-2022.
- [9] M. D. Zeiler and R. Fergus, 'Visualizing and understanding convolutional networks', in *Computer Vision – ECCV 2014*, Cham: Springer International Publishing, 2014, pp. 818–833.
- [10] G. Urban et al., 'Do deep convolutional nets really need to be deep and convolutional?', *arXiv [stat.ML]*, 17-Mar-2016.
- [11] D. Bau, B. Zhou, A. Khosla, A. Oliva, and A. Torralba, 'Network dissection: Quantifying interpretability of deep visual representations', in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, 2017.
- [12] H. Li, Z. Xu, G. Taylor, C. Studer, and T. Goldstein, 'Visualizing the loss landscape of neural nets', *arXiv [cs.LG]*, 28-Dec-2017.
- [13] H. Harapan et al., 'Coronavirus disease 2019 (COVID-19): A literature review', *J. Infect. Public Health*, vol. 13, no. 5, pp. 667–673, May 2020.

- [14] J. Elliott, M. Whitaker, B. Bodinier, O. Eales, S. Riley, H. Ward, ... & P. Elliott, Predictive symptoms for COVID-19 in the community: REACT-1 study of over 1 million people. *PLoS medicine*, 18(9), e1003777., 2021.
- [15] U. Jain, Effect of COVID-19 on the Organs. *Cureus*, 12(8), 2020
- [16] D. L. Weiner, V. Balasubramaniam, S. I. Shah, J. R. Javier, and Pediatric Policy Council, 'COVID-19 impact on research, lessons learned from COVID-19 research, implications for pediatric research', *Pediatr. Res.*, vol. 88, no. 2, pp. 148–150, Aug. 2020.
- [17] Ertuğrul, Ö. F., Emrullah, A. C. A. R., Öztekin, A., & Aldemir, E. (2021). Detection of Covid-19 from X-ray images via ensemble of features extraction methods employing randomized neural networks. *European Journal of Technique (EJT)*, 11(2), 248-254.
- [18] Kaya, Y., Yiner, Z., Kaya, M., & Kuncan, F. (2022). A new approach to COVID-19 detection from X-ray images using angle transformation with GoogleNet and LSTM. *Measurement Science and Technology*, 33(12), 124011
- [19] Kaggle. COVID-19 Radiography Database. <https://www.kaggle.com/datasets/tawsiurrahman/covid19-radiography-database>. Accessed 18 January 2023
- [20] M. F. Aslan, K. Sabanci, A. Durdu, and M. F. Unlarsen, 'COVID-19 diagnosis using state-of-the-art CNN architecture features and Bayesian Optimization', *Comput. Biol. Med.*, vol. 142, no. 105244, p. 105244, Mar. 2022.
- [21] M. F. Sohan, A. Basalamah, and M. Solaiman, 'COVID-19 detection using machine learning: a large scale assessment of x-ray and CT image datasets', *J. Electron. Imaging*, vol. 31, no. 04, Mar. 2022.
- [22] A. T. Sahlol, D. Yousri, A. A. Ewees, M. A. A. Al-Qaness, R. Damasevicius, and M. A. Elaziz, 'COVID-19 image classification using deep features and fractional-order marine predators algorithm', *Sci. Rep.*, vol. 10, no. 1, p. 15364, Sep. 2020.
- [23] J. Abdollahi and L. Mahmoudi, 'An artificial intelligence system for detecting the types of the epidemic from X-rays: Artificial intelligence system for detecting the types of the epidemic from X-rays', in 2022 27th International Computer Conference, Computer Society of Iran (CSICC), Tehran, Iran, Islamic Republic of, 2022.
- [24] R. Abdrakhmanov, M. Altynbekov, A. Abu, A. Shomanov, D. Viderman, and M.-H. Lee, 'Few-shot learning approach for COVID-19 detection from X-ray images', in 2021 16th International Conference on Electronics Computer and Computation (ICECCO), Kaskelen, Kazakhstan, 2021.

#### BIOGRAPHIES



**CÜNEYT ÖZDEMİR** He graduated from the Computer Engineering Department of Firat University in 2004. He pursued his M.S degree in the Computer Engineering Department of Firat University in 2013. In 2022, he received Ph.D. degrees in the Department of Electrical and Electronics Engineering at Siirt University. Between 2004 and 2009, he gained professional experience in the private sector. In 2010, he assumed the position of a lecturer at Siirt University Vocational School. Since 2022, he has been serving as an Assistant Professor in the Department of Computer Engineering at Siirt University. His research interests encompass the fields of machine learning, deep learning, computer vision, image processing and signal processing



# Intelligent Video Surveillance System Using Faster Regional Convolutional Neural Networks

Olayemi Mikail Olaniyi, Shefiu Olusegun Ganiyu, Efenedo Gabriel Ilori, Sunday J Akam

**Abstract**— Insecurity remains a major challenge in our society. Government, private organizations, and individuals strive to ensure their possessions are kept safe from intruders. Automated surveillance system plays a key role to ensure that the environment is safe with little human intervention. Therefore, object detection, classification, and tracking are vital in building a robust and remote intelligent video surveillance system to aid security in physical environments. Previous studies used enhanced background subtraction techniques for object detection which recorded notable achievements but performance issues in distinguishing humans, pets and vehicles. For insecurity to be solved more intelligently, deep neural network techniques are employed. In this paper, an intelligent video surveillance system that detects only human intrusion and sends an SMS notification to the user with the registered mobile number was developed. The results of the system performance evaluation recorded an accuracy of 96%, a precision of 94%, and a recall of 98%. The experimental results showed that the intelligent system was suitable for detecting human intrusion, thereby contributing to the safety of physical environments.

**Index Terms**— CNN, Deep learning, Detection, Surveillance Video.

## I. INTRODUCTION

THE NEED for day-to-day security of immediate environment cannot be over-emphasized. For a long time now, video surveillance has been an important aspect of security system that plays a vital role in ensuring that lives and properties are kept safe. The early implementation of this system relied on humans for its operations [1]. Thus, surveillance systems have increased in number as the demand for this system increases over the years. With the advent of development in video surveillance systems, governments, individuals, and various organizations across society use the systems to keep track of various activities for the sole aim of security and safety [2]. In today's smart cities, video

**O. M. OLANIYI** is with Department of Computer Science, National Open University of Nigeria, Abuja, Nigeria, (e-mail: [omoolaniyi@noun.edu.ng](mailto:omoolaniyi@noun.edu.ng)).

 <https://orcid.org/0000-0002-2294-5545>

**S.O. GANIYU**, is with Computer Science Department, Kampala International University, Uganda, (e-mail: [ganiyu.shefiu@kiu.ac.ug](mailto:ganiyu.shefiu@kiu.ac.ug)).

 <https://orcid.org/0000-0003-4182-3890>

**G. I. EFENEDO**, is with Department of Electrical and Electronics Engineering, Delta State University, Abraka, Nigeria, (e-mail: [giefenedo@delsu.edu.ng](mailto:giefenedo@delsu.edu.ng)).

 <https://orcid.org/0000-0002-2831-068X>

**S. J. AKAM**, is with Department of Computer Engineering, Federal University of Technology, Minna, Nigeria, (e-mail: [mailto:sundayjames115@gmail.com](mailto:mailto:sundayjames115@gmail.com))

 <https://orcid.org/0000-0002-2294-5545>

Manuscript received Dec 22, 2022; accepted Sep 15, 2023.

DOI: [10.17694/bajece.1223050](https://doi.org/10.17694/bajece.1223050)

surveillance is used for inventory control in retail outlets, home monitoring (neighbourhood watch), security on corporate and educational campuses.

Video surveillance systems by human operators to detect an intrusion is an inefficient or even impractical solution because human resources are expensive and have limited capabilities [3]. Intelligent video surveillance systems are integrated systems that include electronic (sensing devices), pattern recognition, and computer vision, networking, artificial intelligence, and communication [2]. Therefore, the goal of an intelligent video surveillance system is to automatically monitor people, property, and the environment without the need for human intervention. As a result, this monitoring task entails automatically detecting and classifying objects (either humans or household pets), as well as performing additional analysis and taking actions. Especially, image processing and artificial intelligence (deep learning) techniques are important in the development of intelligent video systems [4].

With advancements in deep learning, particularly convolutional neural network (CNN), in computer vision applications, the accuracy of object detection and classification has improved dramatically for intelligent video surveillance [5]. Neural network algorithms offer state-of-the-art performance in classification and object detection widely used in intelligent video surveillance for intrusion detection in restricted environments. The specific contribution of this paper is to evolve an algorithm that supports the design and development of an intelligent surveillance system, which is based on Faster R-CNN for human and non-human detections for accurate intrusion detection while reducing false (increase precision) alarm rate as anticipated in [3].

## I. RELATED WORKS

Several researchers have worked on intelligent video surveillance and object detection, classification, and tracking. An object detector using multi-detector regional convolutional neural networks (RCNN) was developed in [6]. Object detection and tracking play a very important role in surveillance for traffic control, object counting, and physical aspect of security. The system adopted a multi-detector model based on faster RCNN, a combination of CNN and Amulet is use to extract the raw feature from the image, region proposal network (RPN) is used to predict the expected region of interest (RoI). Thus, multiple detections are used to detect the image. Similarly, An intelligent smartphone-enabled surveillance system was designed and implemented in [7]. The system uses a passive infrared (PIR) sensor and a microcontroller (MCU) attached to a smartphone through the MCU for motion detection. When a motion is detected, video is captured and the

footage is sent to the user via short message service (SMS). The surveillance record is stored in a cloud and the link to the record is also sent to the user via email. The developed system ensures efficient use of memory by storing the record in a cloud. It is cost-effective and also offers efficient energy use as the camera is only activated when motion is detected by the PIR sensor. However, the developed system cannot efficiently differentiate radiation changes between humans, household pets, or other animals.

Also, an intelligent surveillance system for a low-cost CNN design was developed in [8]. The developed system makes use of a hardware accelerator known as Neural Compute Stick (NCS) with ROCK64 for high-speed calculation of images. A lightweight MobileNet network is used to extract the features and classify the image. Authors in [8] used the NCS to load a single shot multibox detector (SSD) network for human detection. Also, the Darknet architecture of You Only Look Once (YOLO) is used for extraction and classification of images and combine with SSD to create a bounding box for the region of interest of the detected images. A simple mail transfer protocol was used to send email to deliver the detected object. Furthermore, an enhanced background subtraction algorithm for a smart surveillance system using adaptive gaussian mixture technique was developed in [9]. The smart system can efficiently detect motion and detect an object by means of background subtraction with illumination change. However, the developed system cannot differentiate between humans and home pets. In addition authors in [10] developed a real-time Action Detection in Video Surveillance using Sub-Action Descriptor with Multi-CNN. The system presented a novel real-world surveillance video dataset and a new approach to real-time action detection in video surveillance system. The joint space of the sub-action descriptor was not considered. Also, more powerful temporal feature methods, such as skin-color MHI or optical flow, and other deep architectures of CNNs are not considered.

Similarly, [11] developed an activity recognition using temporal optical flow convolutional features and multi-player LSTM. The activity recognition framework for industrial systems proposed with a trained map CNN model help to select only the salient region that are activated for persons in the video frame which reduce verbosity and ambiguity of information in video frame. Surveillance video analysis for store-base using deep learning techniques proposed was by [12]. A skeleton recognition algorithm is adopted in place of object detection algorithm to conquer occlusion problem for gathering sufficient customer information and realizing crowd counting and density map drawing. For human tracking and counting, multiple human tracking algorithm and human re-identification (ReID) technology are adopted.

Also, [4] developed a people tracking system by Detection Using CNN features. They represented each person with 4096 Faster-RCNN feature vectors, and the Euclidean distance method was used to calculate the distance between two feature vectors of each input pair. A pair is considered the same person if their Euclidean distance is a minimum. This is due to the assumption that convolutional features of similar objects generated by Faster-RCNN should be quite similar compared to features of dissimilar objects. Furthermore, [13] suggested a

General Purpose Intelligent Surveillance System for Mobile Devices using deep learning. The developed system module was divided into two: a detection and a classification module. The detection module combined background subtraction techniques, optical flow and recursively estimated density. The classification module is based on a CNN used to classify objects into one of the seven predefined categories using a pre-trained CNN.

In addition, [14] designed and developed an Edge Intelligence-Assisted Smoke Detection in Foggy Surveillance Environments. The system was developed using the architecture of CNN for detecting smoke in video streams. Pre-trained MobileNet model was trained on ImageNet dataset which focus on trying to achieve accuracy and eliminating rate of false alarm in Foggy Surveillance Environments. Also, [15] implemented an Intelligent Surveillance System Using Background Subtraction Technique for unattended or abandoned object detection. In the developed system, threshold is applied to separate red, green and blue then the use of blob-based algorithm for detecting the change in video scene, the technique detects, analyze and track object motion.

Similarly, [16] presented the use of Adaboost and CNN in crowded surveillance environment for people counting based on head detection. In this system three modules were used to achieve people counting. The module includes: Two off-line training and one online detection stage. The first off-line training, Adaboost algorithm is adopted to learn a fast-cascaded head detector with Histogram of Oriented Gradients (HOG) feature. In addition, [17] developed a Smart Surveillance as an Edge Network Service: from Harr-Cascade, Support vector Machine (SVM) to a Lightweight CNN. The system used histogram Oriented Gradient (HOG) and SVM algorithm for fast and accurate human detection. The system also used Harr cascade, Harr-like feature made up of three shapes: two rectangular features, three rectangular features and four rectangular features alongside Lightweight CNN as the classifier trained with keras dataset. Authors in [18] implemented a video structured description technology-based intelligence analysis of surveillance videos for public security applications. A pre-trained CNN architecture was adapted for tracking and re-identification of people and analysed the architecture with CUHK03 dataset. The researchers provided both manually cropped images and automatically detected bounding boxes with DPM detector, which respectively contains 13,164 images of 1360 pedestrians captured by six surveillance cameras.

Also, an efficient CNN based summarization of surveillance videos for resource-constrained devices was presented in [19]. The study investigated deep features for shot segmentation and intelligently divided the video stream into meaningful shots. The deep features were extracted from two consecutive frames to determine whether the underlying frames belong to the same or different shot. The Features were extracted from the fully connected layer (FC7) of CNN model which is trained using MobileNet architecture (version 2) on ImageNet dataset.

In addition, a Kernel ELM and CNN based Facial Age Estimation was developed in [20]. A two-level system for apparent age estimation from facial images. Then first classify

samples into overlapping age groups. Within each group, the apparent age is estimated with local repressors, whose outputs are then fused for the final estimate. We use a deformable parts model-based face detector, and features from a pre-trained deep convolutional network. Kernel extreme learning machines are used for classification.

Also, a Surveillance System Using CNN for Face Recognition with Object, Human and Face Detection was developed in [21]. The region of object they considered in an entire image was picked by object detection and discriminate whether the area is human or Otherwise and the movement of the detected object was analyzed. Similarly, [22] implemented a vegetable Category Recognition System Using Deep Neural Network. They implemented a Caffe framework based on CNN for the system classification and used Deep Neural Network (DNN) for the vegetable category recognition.

Also, [23] presented an adaptive Feature Learning CNN for Behavior Recognition in Crowd Scene. A 3D scale convolutional neural network (3DSCNN) was implemented on crowd video scene, the 3D-CNN was used in a large-scale supervised crowd dataset which optimized convolutional architectures settings. The outcomes from 3DS-CNN captured information related to objects, scenes, and actions in a video, making them useful for different applications that do not fine tune the architectural setup. In addition, [24] designed HOG-CNN Based Real Time Face Recognition to recognize faces. Histogram of Oriented Gradient (HOG) was used as the feature extractor, also for detecting all the faces in the image and the CNN was used as the training algorithm for classifying the images.

Furthermore, an Engineering Vehicles Detection Based on Modified Faster R-CNN for Power Grid Surveillance was developed in [25]. CNN methods were divided into two categories, one is the two-stage methods based on region proposal and the other is the one-stage methods based on regression. The feature extraction part of these methods was implemented by the CNN. Some methods based on region proposal such as R-CNN, Fast R-CNN and SPPnet, which adopted selective search algorithm to generate candidate boxes were utilised. Also, YOLO was used as the topmost feature map to predict confidences and bounding boxes for all categories over a fixed grid. SSD detects multiple categories by a single evaluation of the input image.

Machine Learning for Gender Detection using CNN on Raspberry Pi Platform was presented by [26]. The implementation of the system is based on the architecture of CNN and the solution permits users to extract some relevant information from the visual data containing image labelling, face and landmarks detection, optical character recognition (OCR). REST API was used to interact with Google's cloud vision platform. The real-time implementation of the hardware as well as software solution were implemented and executed on a Raspberry Pi 3 model B+ board with Pi Camera module. This paper tries to tackle such limitations presented by [27-32], [2], [10] by making use of raspberry pi and faster object detection and classification technique to improve video surveillance system

### III PROPOSED DESIGN AND METHODOLOGY

This section describes in detail, the methodological steps employed in the design and implementation of the proposed intelligent surveillance system. Hence, the proposed system utilized flow design tool, flowchart, block diagram and circuit diagram for the implementation to describe the technical phase of the methodology. The system uses a trained faster regional convolutional neural network classifier for its object detection. It also makes use of Twilio API for SMS notification. With the above techniques the system was able to solve the problem of detecting home pet and humans as intruder. All these techniques were used to improve the efficiency, performance and intelligence of the proposed surveillance system.

#### A. Methodology

The implementation of the proposed system for intelligent video surveillance is guided by the block diagram in Figure 1. Foremost, Faster R-CNN architecture is considered because it is fast, accurate and suitable for detecting and classifying human [31] objects and home pets [4]. The Faster R-CNN architecture as shown in Figure 9 is divided into two modules: The Region Proposal Network (RPN) and a Fast R-CNN Detector. The RPN and the Fast R-CNN detector share the same convolutional layers. Faster R-CNN, by consequence, could be considered as a single and a unified network for object detection. To generate high quality object proposal, a highly descriptive feature extractor in the convolutional layers can be used. The Fast R-CNN detector uses many regions of interest (ROIs) as input. Then, the ROI pooling layer extracts a feature vector for each ROI. This feature vector will constitute the input for a classifier formed by a series of fully connected (FC) layers.

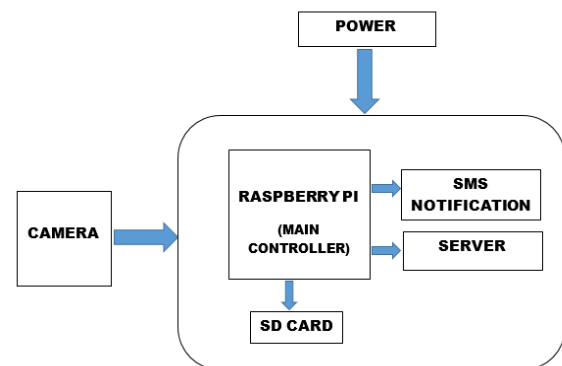


Fig. 1. Proposed system block diagram

The embedded system of the proposed system, include Raspberry Pi 3B 8MP camera module, Raspberry Pi 3B as the main controller for all the object detection and programming for the whole system, SMS notification, buzzer (Alarm notification) and power supply.

The system comprises of five basic components

- Raspberry Pi 3
- Raspberry Pi 3 8MP Camera module
- SMS notification
- Power supply
- Raspberry microSD card.

The Raspberry Pi 3B is a basic module for processing images/videos, executing object detection on acquired video frames to detect objects. The board has ARM cortex A53 clocked at 1.2GHz, 4000MHz Video Core IV multimedia GPU,

1Gb memory, power supply, HDMI, USB ports and other features.

The camera module takes in video stream then the Raspberry Pi 3B module implement the object detection on the captured frames. Figure 2 shows the flow diagram of the proposed system;

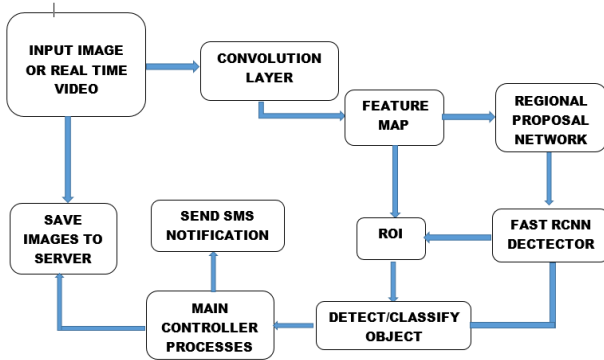


Fig. 2. Flow diagram of the proposed system

The electronic components to be used are; Raspberry pi 3B, PiCamera module SMS notification and power supply. The camera and the power supply all the required inputs to the Raspberry Pi 3B, while the SMS notification act as the output for the system. Once frames/video are acquired from the camera and fed into the Raspberry Pi 3B controller, the image is being processed using the faster regional convolutional neural network as stored within the programmed Raspberry pi 3B.

**B. Mathematical Modelling of Faster Regional Convolutional Neural Network for Intelligent Video Surveillance**

Regional convolutional neural network remains the first notable CNN-based object detection techniques which was found to outperformed traditional object detection algorithm with 30% improvement over the years. It relies on two separate CNN to perform its object detection. CNN computation is based on the basic computation performed by an artificial neuron which is the sum of products between weights and values of a layer. Furthermore, the computation operates in two modes - single input image and multiple input images.

**1. Single input image**

If  $(l_{x,y})$  denote image or pooled feature values, depending on the layer, then the convolution value at any point  $(x, y)$  for a single network in the input is given by:

$$w * l_{x,y} = \sum_i \sum_k w_{ik} l_{x-i,y-k} \tag{1}$$

Where  $w$  is the weight,  $l$  is the feature value,  $i$  and  $k$  are the dimensions of the kernel. For a  $3 \times 3$  weight ( $w$ ) equation (1) becomes,

$$w * l_{x,y} = \sum_i \sum_k w_{ik} l_{x-i,y-k} = w_{1,1}l_{x-1,y-1} + w_{1,2}l_{x-1,y-2} + \dots + w_{3,3}l_{x-3,y-3} \tag{2}$$

Labelling the subscript on  $w$  and  $l$ ,

$$w * l_{x,y} = w_1l_1 + w_2l_2 + \dots + w_9l_9 = \sum_{i=1}^9 w_i l_i \tag{3}$$

Since equation (2) and (3) are identical, if bias is added to the equation and the result is equated to  $T$ ,

$$T = \sum_{j=1}^9 w_j l_j + b = w * l_{x,y} + b \tag{4}$$

**2. Multiple input images**

The equation forward pass-through CNN,

$$T_{xy}(\vartheta) = \sum_i \sum_k w_{i,k}(\vartheta) l_{x-i,y-k}(\vartheta - 1) + b(\vartheta) = w(\vartheta) * l_{x,y}(\vartheta - 1) + b(\vartheta) \tag{5}$$

And

$$l_{xy}(\vartheta) = h(l_{xy}(\vartheta)) \tag{6}$$

Where  $\vartheta = 1, 2, \dots, Lc$ , and  $Lc$  is the number of convolutional layers, and denotes the values of pooled features in convolutional layer  $\vartheta$ , where  $h$  is the activation function.

When  $\vartheta=1$ ,  $l_{xy}(0)$  = (values of pixels in the input image(s))

When  $\vartheta=Lc$ , then:

$l_{xy}(Lc)$  = (values of pooled features in the last layer of the CNN)

The Equations of Backpropagation Used to Train CNNs:

$$\delta_{x,y}(\vartheta) = h'(T_{x,y}(\vartheta)) \left[ \delta_{x,y}(\vartheta + 1) * \sqrt{180(w(\vartheta + 1))} \right] \tag{7}$$

$$\vartheta = Lc - 1, Lc - 2, \dots, 1 \tag{8}$$

Finally, the update parameter updates the weights and bias for each feature map using,

$$w_{i,k}(\vartheta) = w_{i,k}(\vartheta) - l\delta_{i,k}(\vartheta) * \sqrt{180(l(\vartheta - 1))} \tag{9}$$

And

$$b(\vartheta) = b(\vartheta) - l \sum_x \sum_y \partial_{x,y}(\vartheta); (\vartheta = 1, 2, \dots, Lc) \tag{10}$$

**3. Faster Regional Convolutional Neural Network**

Faster R-CNN belongs to the family of Regional convolutional neural network and it is becoming a replacement for fast R-CNN. The evolution within the different versions of R-CNN was usually in terms of computational efficiency (integrating the different training stages), reduction in test time, and improvement in performance (mAP). These networks usually consist of

- 1) A region proposal algorithm to generate “bounding boxes” or locations of possible objects in the image;
- 2) A feature generation stage to obtain features of these objects, usually using a CNN;
- 3) A classification layer to predict which class this object belongs to; and
- 4) A regression layer to make the coordinates of the object bounding box more precise.

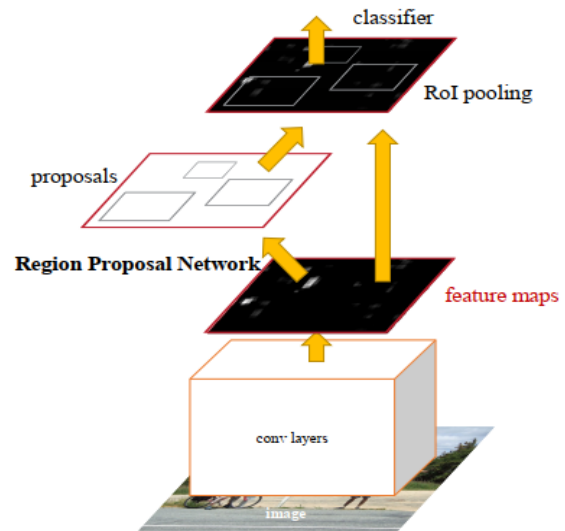


Fig. 3. Architecture of Faster R-CNN [38]



The Faster R-CNN architecture is illustrated in Figure 3. It has commonly used two-step approach to object detection. The first stage generates a set of class agnostic region of interest (RoI) or regional proposal network, where each RoI is defined by a bounding box location and an abjectness score. The second step then classifies each RoI and predicts a refined limit box position. From a general point of view, Faster R-CNN is composed of two subnetworks:

- Regional proposal Network (RPN)
- Region of interest box network (RoI)

The algorithm for simulating a multi-human object detection is presented in Figure 4.

**ALGORITHM 1**

1. Take input image from camera.
2. Pass the image to the ConvNe layer and returns feature maps for the image.
3. Apply Region Proposal Network on the feature maps and gets object proposals.
4. Apply RoI pooling layer to bring down all the proposals to the same size.
5. Pass these proposals to a fully connected lay.
6. Classify and predict the bounding boxes for the image.
7. Print the predicted object.

Fig. 4. Simulation on multi human object

4. Regional Proposal Network

The region proposal network (RPN) starts with the input image being fed into the backbone CNN. The input image is first resized such that its shortest side is 600px with the longer side not exceeding 1000px. It takes an image (of any size) as input and outputs a set of rectangular object proposals, each with an objective score. To generate region proposals, we slide a small network over the convolutional feature map output by the last shared convolutional layer. This small network takes as input an  $n \times n$  spatial window of the input convolutional feature map as shown in Figure 5.

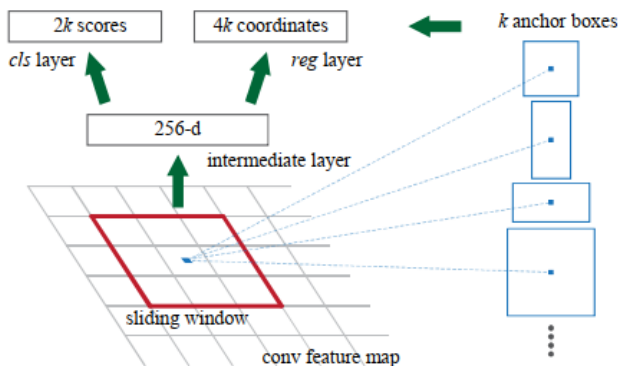


Fig. 5. Regional proposal network [39]

**ALGORITHM 2**

1. Initialise the input camera
  - a. Get image/video frames
  - b. Send the frame to raspberry pi
2. Raspberry pi loads the model to begin object detection for intelligent surveillance
3. Pass the object through faster RCNN
4. Is the object detected is a human or home pet
  - If home pet,
    - a. No intruder found go to
  - If human intruder detected,
    - b. Intruder detected
5. Send SMS using Twilio API
6. Save the footage to SD card
7. Send the footage to server

Fig. 6. Algorithm for the proposed video surveillance system

Each sliding window is mapped to a lower-dimensional feature (256-d for ZF and 512-d for VGG, with ReLU following). This feature is fed into two sibling fully connected layers—a box-regression layer (reg) and a box-classification layer (cls). We use  $n = 3$  in this research, noting that the effective receptive field on the input image is large (171 and 228 pixels for ZF and VGG, respectively). The algorithm that explains the flow diagram of the overall system is shown in Figure 6.

5. System Implementation

The system simulation was executed on a personal computer having TensorFlow installed in a virtual environment of an Anaconda programming environment. All the codings were done using Python. The system webcam was used to capture images which were later transformed into frames for processing. Also, the same computer was used to process the frames, results of the implemented Faster R-CNN and the results from the integration of the software and hardware part of the developed system. Also, the performance and accuracy of the system using precision and recall for faster R-CNN in intelligent surveillance were conducted as part system evaluation. Thus, the overall system algorithm was implemented using python programming language, TensorFlow and OpenCV APIs to detect different object of humans, cat and dog (i.e., dog and cat as home pet). The training of the model was done using TensorFlow API and faster R-CNN algorithm.

A working prototype of the developed system is shown in Figure 7. Also, a sample screenshot of the system’s user interface that showed an alert message when intruder was detected is presented in Figure 8.



Fig. 7. Prototype of intelligent surveillance system using Faster RCNN

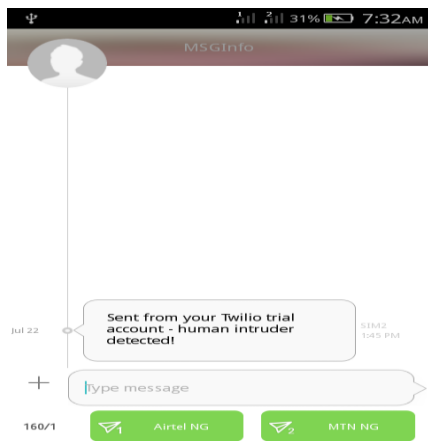


Fig. 8. User interface showing intrusion message



Fig. 9. Simulation on multi-human objects

IV. RESULTS AND DISCUSSION

The trained model was tested with images that produced outputs that contained the classes ID number, the detection confidence and the anchor boxes. This is used to identify and classify the object and determine the location of the object using  $y_{max}$ ,  $x_{max}$  and  $y_{min}$ ,  $x_{min}$  coordinates. Using the input frames, the

model locates and recognizes the category in which the object belongs, i.e., human, cat or dog. The tasks of locating and categorizing objects is achieved with principles of detection, localization and classification. The model was also tested on multi-human objects as presented in Figure 9, as well as cat and dog as home pet as shown in Figure 10.

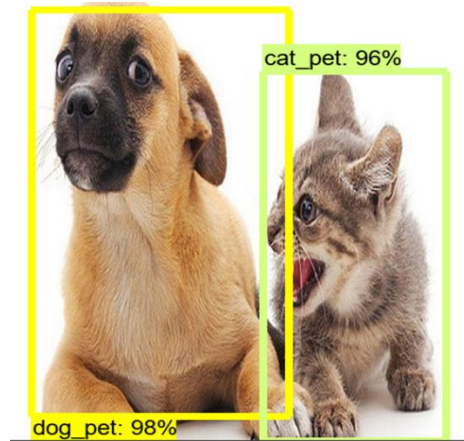


Fig. 10. Simulation on home pet

The result of the performance evaluation conducted on the implemented system is shown in Table 1. The system achieved highest precision ratio and recall ratio of 94% and 98% respectively. These results demonstrated the appropriateness of the image characteristics such as quality, type and size for study. Furthermore, the results revealed the suitability of the algorithms and APIs utilized for the system implementation.

TABLE I  
DETECTION RESULT FOR INTELLIGENT VIDEO SURVEILLANCE SYSTEM

S/N	Precision (%)	Recall (%)
1	90	98
2	88	96
3	91	97
4	92	94
5	89	95
6	93	97
7	94	96
8	90	98
9	94	97
10	89	91

V. CONCLUSION

Security remains a major concern to everyone. Everybody wants to be protected from being attacked, and the means to prevent this has been a challenge over the years. A lot of solution has already been put in place to tackle insecurity. Thus, this study provided an additional approach to already existing motion and object detection techniques. The was implemented with Faster-RCNN and it intelligently detected people, dogs, and cats in a bid to raise intrusion alerts via SMS when human intrusion is detected within an environment. Faster R-CNN was adopted as an object detection model because it utilizes a regional proposal network (RPN) for faster and more accurate detection. Also, the intelligent surveillance system is portable and requires minimum expertise to operate. Overall, it provides another opportunity to enhance physical security at home and place of work against human intruders with a good performance.

## REFERENCES

- [1] Y. Kurylyak, "A Real-Time Motion Detection for Video Surveillance System," IEEE International Workshop on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications, Rende, Italy, 2009, pp. 386-389, doi: 10.1109/IDAACS.2009.5342954.
- [2] S. W. Ibrahim, "A comprehensive review on intelligent surveillance systems", CST, vol. 1, no. 1, pp 7-14, May 2016
- [3] O.M. Olaniyi, S. Ganiyu and S. J. Akam. Intelligent Video Surveillance Systems: A Survey. Balkan Journal of Electrical and Computer Engineering (BAJECE).1(1).pp 57-53
- [4] A. A. Shafie, F. Hafizhelmi, and K. Zaman, "Smart Video Surveillance System for Vehicle Detection and Traffic Flow Control". Journal of Engineering Science & Technology (JESTEC). Vol.13 no 7. 2195-2210
- [5] B. Benjdira, T. Khurshed, A. Koubaa, A. Ammar, and K. Ouni, "Car Detection using Unmanned Aerial Vehicles: Comparison between Faster R-CNN and YOLOv3," 2019 1st Int. Conf. Unmanned Veh. Syst., pp. 1–6, 2019
- [6] W. Tan, "Object Detection with Multi-RCNN Detectors," pp. 193–197.
- [7] A. H. Sanoob, J. Roselin, and P. Latha, "Smartphone Enabled Intelligent Surveillance System," no. c, pp. 1–7, 2015, doi: 10.1109/JSEN.2015.2501407.
- [8] L. W. Yang and C. Y. Su, "Low-cost CNN Design for Intelligent Surveillance System," 2018 Int. Conf. Syst. Sci. Eng., pp. 1–4, doi: 10.1109/ICSE.2018.8520133.
- [9] . M. Olaniyi, J. A. Bala, S. O. Ganiyu, and P. E. Wisdom, "A Systematic Review of Background Subtraction Algorithms for Smart Surveillance System," vol. 8, no. 1, pp. 35–54, 2020
- [10] C. Jin, S. Li, and H. Kim, "Real-Time Action Detection in Video Surveillance using Sub-Action Descriptor with Multi-CNN," pp. 1–29.
- [11] A. Ullah, K. Muhammad, J. Del Ser, S. W. Baik, and V. Albuquerque, "Activity Recognition using Temporal Optical Flow Convolutional Features and Multi-Layer LSTM," IEEE Trans. Ind. Electron., vol. PP, no. c, p. 1, 2018, doi: 10.1109/TIE.2018.2881943.
- [12] H. Kaya, H. Dibeklio, and A. A. Salah, "Kernel ELM and CNN based Facial Age Estimation," pp. 80–86.
- [13] A. Antoniou, "A General Purpose Intelligent Surveillance System For Mobile Devices using Deep Learning," pp. 2879–2886, 2016
- [14] . Muhammad, S. Khan, S. Member, and V. Palade, "Edge Intelligence-Assisted Smoke Detection in," IEEE Trans. Ind. Informatics, vol. PP, no. c, p. 1, 2019, doi: 10.1109/TII.2019.2915592
- [15] . Hargude and M. T. It, "i-surveillance: Intelligent Surveillance System Using Background Subtraction Technique," vol. 1.
- [16] C. Gao, P. Li, Y. Zhang, J. Liu, and L. Wang, "Author ' s Accepted Manuscript People counting based on head detection combining environment Reference: To appear in: Neurocomputing," Neurocomputing, 2016, doi: 10.1016/j.neucom.2016.01.097.
- [17] . Y. Nikouei, Y. Chen, S. Song, R. Xu, B. Y. Choi, and T. Faughnan, "Smart surveillance as an edge network service: From harr-cascade, SVM to a Lightweight CNN," Proc. - 4th IEEE Int. Conf. Collab. Internet Comput. CIC 2018, pp. 256–265, 2018, doi: 10.1109/CIC.2018.00042.
- [18] Z. Xu, C. Hu, and L. Mei, "Video structured description technology based intelligence analysis of surveillance videos for public security applications," 2015, doi: 10.1007/s11042-015-3112-5.
- [19] T. Hussain, K. Muhammad, A. Ullah, Z. Cao, S. W. Baik, and V. H. C. De Albuquerque, "Cloud-assisted multiview video summarization using CNN and bidirectional LSTM," IEEE Trans. Ind. Informatics, vol. 16, no. 1, pp. 77–86, 2020, doi: 10.1109/TII.2019.2929228.
- [20] H. Kaya, H. Dibeklio, and A. A. Salah, "Kernel ELM and CNN based Facial Age Estimation," pp. 80–86.
- [21] Y. Byeon and S. Pan, "A Surveillance System Using CNN for Face Recognition with Object, Human and Face Detection," pp. 975–984, doi: 10.1007/978-981-10-0557-2.
- [22] Nogay, H.S. T.C. Akinci, and M. Yilmaz. "Detection of invisible cracks in ceramic materials using by pre-trained deep convolutional neural network." Neural Computing and Applications 34.2 (2022): 1423-1432.
- [23] A. N. Shuaibu, A. S. Malik, and I. Faye, "Adaptive Feature Learning CNN for Behavior Recognition in Crowd Scene," pp. 357–361, 2017
- [24] H. Ahamed, I. Alam, and M. Islam, "HOG-CNNBasedRealTimeFaceRecognition," 2018 Int. Conf. Adv. Electr. Electron. Eng., pp. 1–4, 2018.
- [25] X. Xiang, N. Lv, X. Guo, S. Wang, and A. El Saddik, "Engineering vehicles detection based on modified faster R-CNN for power grid surveillance," Sensors (Switzerland), vol. 18, no. 7, 2018, doi: 10.3390/s18072258.
- [26] M. H. Gauswami, "Implementation of Machine Learning for Gender Detection using CNN on Raspberry Pi Platform," 2018 2nd Int. Conf. Inven. Syst. Control, no. Icisc, pp. 608–613, 2018.
- [27] D. Chahyati, M. I. Fanany, and A. M. Arymurthy, "Tracking People by Detection Using CNN Features," Procedia Comput. Sci., vol. 124, pp. 167–172, 2018, doi: 10.1016/j.procs.2017.12.143.
- [28] D. Chahyati, M. I. Fanany, and A. M. Arymurthy, "Tracking People by Detection Using CNN Features," Procedia Comput. Sci., vol. 124, pp. 167–172, 2018, doi: 10.1016/j.procs.2017.12.143.
- [29] H. C. Shin and J. Y. Lee, "Pedestrian Video Data Abstraction and Classification for Surveillance System," 9th Int. Conf. Commun. Technol. Converg. ICT Converg. Powered by Smart Intell. ICTC 2018, pp. 1476–1478, 2018, doi: 10.1109/ICTC.2018.8539426.
- [30] A. Ullah, K. Muhammad, J. Del Ser, S. W. Baik, and V. Albuquerque, "Activity Recognition using Temporal Optical Flow Convolutional Features and Multi-Layer LSTM," IEEE Trans. Ind. Electron., vol. PP, no. c, p. 1, 2018, doi: 10.1109/TIE.2018.2881943.
- [31] L. Du, R. Zhang, and X. Wang, "Overview of two-stage detection algorithms," 2020, doi:10.1088/1742-6596/1544/1/012033
- [32] R. Arti, "Animal Detection Using Deep Learning Algorithm," vol. 7, no. 1, pp. 434–439, 2020
- [33] O., Türk, A. Çalışkan, Acar, E. and B. Ergen. "Palmprint recognition system based on deep region of interest features with the aid of hybrid approach. SIVIP " 17, 3837–3845. 2023.

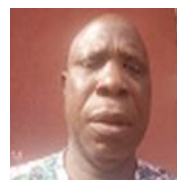
## BIOGRAPHIES



**Olayemi Mikail Olaniyi** is a Professor in the Department of Computer Science at National Open University of Nigeria, Abuja, Nigeria. He obtained his B.Tech. and M.Sc. in Computer Engineering and Electronic and Computer Engineering respectively. He had his Ph.D. in Computer Science and Engineering (Computer Security) from Ladoké Akintola University of Technology, Ogbomosho, Oyo State, Nigeria. He has published in reputable journals and learned conferences. His areas of research include Computer Security, Intelligent/Embedded Systems design and Applied Medical Informatics.



**Shefiu Olusegun Ganiyu** is a Senior Lecturer in the Department of Computer Science, Kampala International University, Uganda. He holds a Ph.D. in Computer Science, which focused on risk-aware access control for pervasive environments. Similarly, he obtained Bachelor degree in Mathematics/Computer Science and Master degree in Information Science from Federal University Minna and the University of Ibadan respectively. Prior to joining the academic environment, he acquired valuable work experience as a programmer/information system developer. His research interests include information security risk management, dynamic access control, user behaviour analytics, security of pervasive computing including Bring Your Own Device (BYOD) strategy, and physical security. Also, he has participated in several projects involving information systems security and development.



**Efenedo Gabriel Ilori** is a Senior Lecturer in the Department of Electrical and Electronics Engineering, Delta State University, Abraka, Oleh Campus, Delta State, Nigeria. He obtained his BEng

degree in Electrical and Electronics Engineering as well MEng and PhD in Electronics and Telecommunication Engineering from the University of Benin, Benin-City, Edo State, Nigeria. He has publications in reputable journals and conferences. His area of research includes: Intelligent, cellular communication and information system.



**Sunday J Akam** obtained his Bachelor's Degree in Computer Engineering from the Federal University of Technology, Minna, Niger State, Nigeria. He is a promising computer vision and embedded systems developer. He has keen passion for AI, machine vision and developing advanced vision models.



# Automated Detection of Collagen Bundles in Second Harmonic Generation Microscopy Images

Cihan Bilge Kayasandik

**Abstract**— Collagen is one of the most abundant proteins in the body. It is essential for the structure, functionality, and strength of the connective tissue such as skin, bone, tendon, and cornea. It is known that a change in the arrangement or morphology of these fibrillar structures relates to multiple dysfunctions including corneal diseases and various cancer types. Due to their critical roles in wide-range abnormalities, there is an increasing interest in the pattern analysis of collagen arrangements. In recent years, Second Harmonic Generation (SHG) microscopy is proven to be an efficient imaging modality for visualizing unstained collagen fibrils. There are plenty of studies in the literature on the analysis of collagen distribution in SHG images. However, the majority of these methods are limited to detecting simple, statistical and non-local properties such as pixel intensity and orientation variance. There is a need for a method to detect the local structural properties of collagen bundles. This paper is to introduce an automated method to detect collagen bundles in 3-dimensional SHG microscopy images. The origin of the proposed method is based on multiscale directional representation systems. The proposed method detects the collagen bundles by measuring the dominant orientation of local regions and an orientation-based connected component analysis. Through more local analysis and the detection of collagen bundles separately, the proposed method would lead to the extraction of more detailed structural information on collagen bundle distribution.

**Index Terms**—collagen detection, cornea analysis, image analysis, machine learning, SHG.

## I. INTRODUCTION

**C**ORNEA is the transplant front region of the outer casing of the eye [1]. Although one of the main functions as protecting the interior content, the cornea has multiple roles in maintaining the human vision system properly. The highly complex and organized structure of the cornea is critically important for satisfying transparency and refraction [2]. Cornea is dominantly formed by structures called collagen. Collagen gives the cornea the ability to be strength required to fulfill its role of producing a tough container for the inner contents of the eye with precise curvature and a high level of transparency to visible wavelengths. The collagen fibers are arranged in parallel bundles called fibrils. These fibrils are packed in layers or lamellae. The stroma of the human eye contains 200–250 distinct lamellae, each layer arranged at right angles relative to fibers in adjacent lamella [3]. It is known that a change in the arrangement or morphological properties of these fibrils relates to dysfunction in the vision system [4]. Since collagen is one of the main structural proteins found in connective

● **Cihan Bilge Kayasandik** is with the Department of Computer Engineering Istanbul Medipol University, Istanbul, Turkey, e-mail: cbkayasandik@medipol.edu.tr

tissues, besides the vision system, collagen structures are known to be related to various abnormalities such as breast and ovarian cancer [5], [6]. Due to this critical relationship between collagen structures and wide-range abnormalities, there is an increasing interest in the pattern analysis of images of collagen arrangements [4], [5], [6].

Multiphoton microscopy is recently recognized as a powerful imaging technique to visualize unstained samples. With no need for staining and advantages for diagnostic procedures, Multiphoton microscopy and its variants have been used increasingly in biomedical imaging [7]. Multiphoton microscopy includes Second Harmonic Generation (SHG) microscopy which is based on an absorption-free process [8]. In recent years, SHG microscopy is proven to be an efficient imaging modality for visualizing unstained collagen fibrils. The Type I collagen which is found on tendons, skin, and cornea is visualized bright in SHG images without any staining [9]. Besides, it produces a better visualization of collagen structures than fluorescent imaging of stained samples [10], [11]. As a result of these advantages, SHG microscopy has become the most preferred imaging modality for the analysis of collagen structures.

There are plenty of studies in the literature on the analysis of collagen distribution in SHG images. When we focus on the computational method papers, majority of these methods work in 2-dimensional space and their abilities are limited to detect basic properties such as pixel value or non-local statistical properties. Some of these studies focus on global or windowed texture analysis like directional variance [12], [13], [14], Fourier Transform and grey level co-occurrence matrix (GLCM) [15]. However, these methods are unable to detect the significant morphological features like fiber size or number [16]. Ogura et al. [17] used Fourier transform to quantify 2-dimensional SHG corneal collagen images. Fourier transform is widely used in texture analysis and gives successful results [18]. However, due to the nature of this method, local information about the morphology of collagen fibers cannot be detected. Therefore, such a technique will not be sufficient to characterize the detailed structure of collagen fibrils. In another study, Hu et al. [19] used the co-occurrence matrix method to distinguish different collagen tissues from each other. The co-occurrence matrix is a statistical method to measure the texture complexity of the given image [20]. This method, which does not give detailed information about individual collagen fibers and only looks at the general image texture, will not be sufficient to measure the features of collagen bundles individually. Besides these non-local analysis studies, there is a limited number of attempts conducted to detect collagen

fibers and extract more local properties. Yong Park et al. [21] used an automatic thresholding method to detect individual collagen fibers. Although this method aims to collect local information despite the previous ones, the thresholding method fails in many object segmentation tasks and does not give an efficient result in complex and low-contrasted images such as an SHG image of collagen. After separating the fibers from the background with thresholding, the authors applied texture analysis to several predetermined patches. However, the failure of the segmentation obtained by thresholding, as well as the analysis of small patches instead of the whole image, makes this method unreliable. Besides, this analysis was applied in 2-dimensional images instead of 3-dimensional images, so, there is a loss of information.

Liu et al. [13] aimed to detect the 3-dimensional structure of collagen bundles. However, since 3-dimensional analysis is computationally expensive, they applied the method referred to as 2.5 convolutions in the literature to reduce the computational load. This method aims to calculate the 2-dimensional projections of 3-dimensional images on three perpendicular planes, and then work with these three 2-dimensional images. Although this technique is suitable for plain images (for example, a single neuron cell image, a single isotropic object image, etc.), it can cause significant information to be lost in the collagen image by detecting separate bundles as a whole. This causes miscalculation of quantification, and creates a possible bias between images and misleads the data analysis.

In a recent study [22] SHG quantitative properties are used to distinguish morphological changes in collagen distribution in different basal cell carcinoma sub-types. For the Quantitative analysis the study focuses on the texture and the directional distribution of collagen bundles. For that purpose, beside the non-local frequency-based analysis, CurveAlign and CT-FIRE methods were used [23], [24]. Curve Align aims to quantify all fiber angles within a specified region [25]. On the other hand, CT-Fire performs a local directional analysis. CT-Fire aims to determine vessel-like structures in the input. As a result, CT-Fire is theoretically applicable for the current collagen bundle detection purpose, and this method forms a comparable alternative for the presented method. However CT-Fire is very sensitive to the contrast in the images. As a

result of that, objects can be detected in several disconnected components and due to that it may result with high number of false positive detection.

Some other methods from the literature are out of focus of this paper since the tasks and/or data acquisition methods are not comparable [26], [27], [28], [29], [30], [31], [32], [33], [34], [35], [36].

In brief, despite the significance of automated quantification of collagen bundles, current methods in the literature are insufficient to detect local properties of 3-dimensional of these structures. This paper aims to introduce a directional analysis-based quantification method to detect collagen bundles in the 3-dimensional SHG microscopy images. Through that, the 3-dimensional structure of collagen bundles will be preserved, and local geometrical properties of bundles will be possible to be collected.

## II. METHOD

All numerical analysis was performed in MATLAB 2023a. The proposed method is specified for the analysis of 3-dimensional SHG images of collagen distribution. All presented numerical results are for healthy data. However, the proposed method can be applied for diseased samples as well through optimizing the parameters accordingly.

The analysis steps and associated sub-steps can be seen in Fig. 1. The proposed method can be grouped into two main steps:

- A. Preprocessing.
- B. Collagen bundle detection.

### A. Preprocessing step

Preprocessing is designed to take full advantage of the capabilities of instrumentation by reducing sources of image degradation such as blurring and noise. However, SHG images of collagen bundles require more than that. The collagen bundles in SHG images seem like a cluster of noise, and even by eye, it is very difficult to determine the exact boundaries of bundle regions. Hence, before the application of

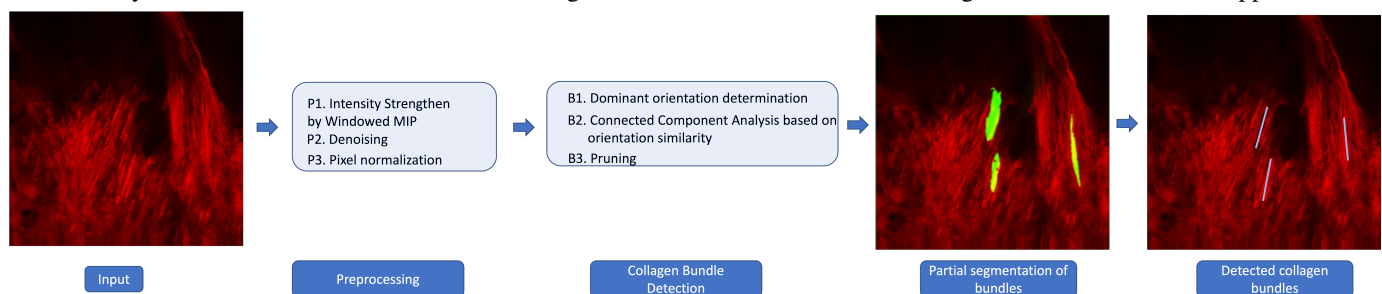


Fig. 1. The framework of the proposed method: The 3-dimensional input is firstly pre-processed by Windowed Maximum intensity projection (MIP), denoising and pixel intensity normalization. Then the pre-processed image is proceed to next step for collagen bundle detection. The output of this step is segmented regions which correspond to a part of collagen bundle. These segmentation are not guaranteed to be an accurate segmentation of bundles. However, they can be safely used for detection purpose. Hence, for next step, to form a better visualization of detected bundles, new output image is generated where detected bundles are represented with rectangular markers.

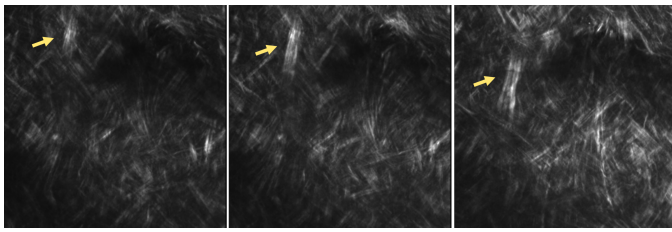


Fig. 2. The partition of a 3-dimensional bundle structure in the subsequent slices. As it is seen the bundle (marked with a yellow arrow) is oriented in the X-Z direction, as a result, it is not possible to detect the bundle in any 2-dimensional slice. However, by maximum filtering, it is possible to increase the visibility of bundles in individual slices (see Fig.3).

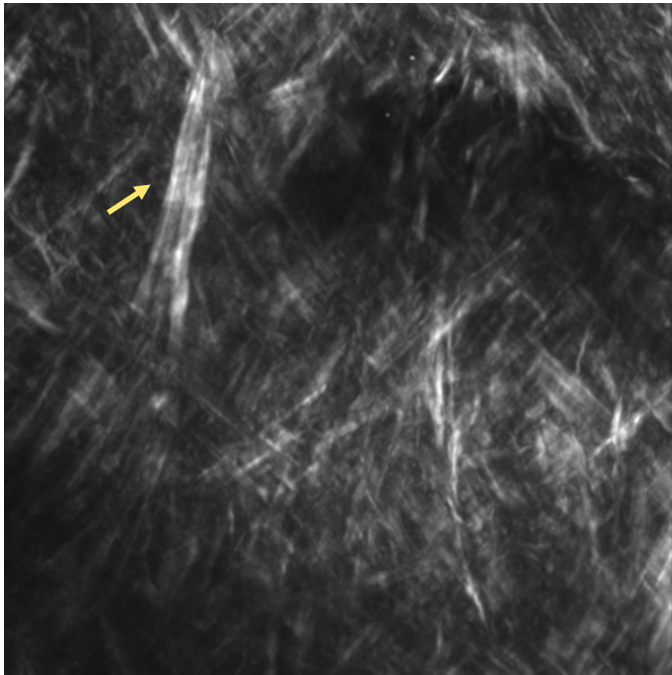


Fig. 3. After pre-processing of the slices in Fig. 2. As a result of maximum filtering on a certain window size, it is possible to visualize the 3-dimensional structure of collagen bundles more clearly in 2-dimensional slices. The bundle is marked with the yellow arrow as in Fig. 2.

noise reduction, an additional step to strengthen the pixel intensities on bundle regions is necessary. The 3-dimensional structure of collagen bundles helps to distinguish bundle regions from artifacts. Through that lead, as the first step of preprocessing, a maximum filter is applied in the X-Z-direction of input volume with a scale depending on the X-Z-direction resolution of the input. This process can be considered as a windowed maximum projection of the input volume. Through this strategy, the pixel brightness in bundle regions increases, and as a result, noise and blurring artifacts get fader. Besides, images are compressed in the X-Z-direction which benefits reducing the cost of the remaining computational analysis (see Fig. 2&3).

Through increased contrast of bundle regions, artifacts can be distinguished from collagen bundles. Thus, pre-processing continues with a denoising step. One of the most common denoising routines is Gaussian smoothing [37], which is based on convolving the input image with a Gaussian function. While

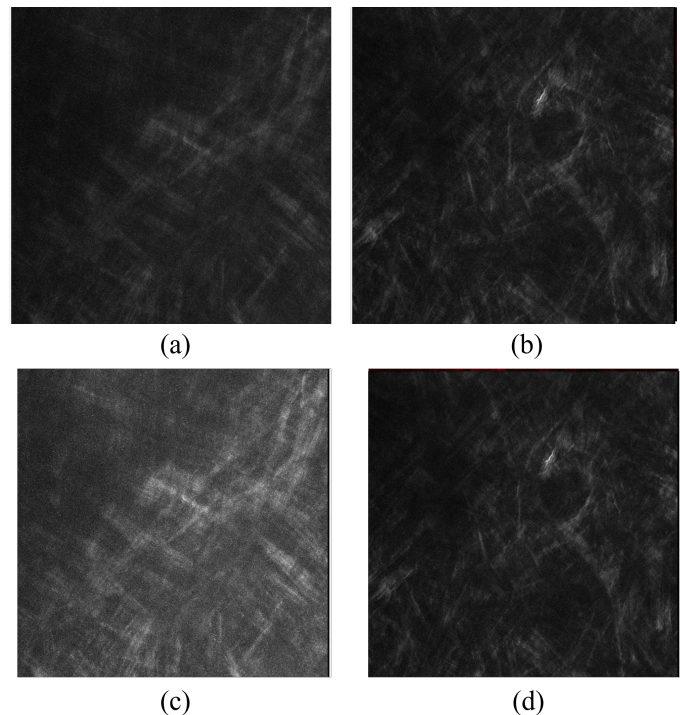


Fig. 4. Pixel normalization for each slice individually affects the visual quality significantly. The top images (a-b) are from the deeper, and the bottom images (c-d) are from surface regions. The first column (a-c) shows the volume-based normalization results, while the second column (b-d) shows slice-based normalization results. As seen slice-based normalization is significantly improving the contrast for deeper regions, while it has minimal effect on surface slices.

Gaussian smoothing is computationally highly efficient, it also has disadvantages such as edge surpassing due to the nature of the Gaussian function. However, since the SHG images of collagen bundles have weak edge information, and edges are not critical for the proposed method, that disadvantage is not a significant issue for the current task. Besides, due to the large image size, the low computational expense is a critical advantage. Hence 3-dimensional small-scale Gaussian filtering is used to smooth and denoise the images. As the last step of the pre-processing, the pixel intensity normalization is applied on each slice of the input volume, individually. Intensity normalization is significantly important when a threshold-based analysis or data clustering is included. In our analysis, besides its necessity for further steps, normalization is used to increase the contrast and have a more obvious view of bundles. The whole volume normalization is not efficient in SHG images since the pixel intensity constantly decreases while going deeper slices (while Z values are increasing). As a result, in the deeper regions, the detection of bundles is difficult even by eye (see Fig. 10). In order to handle that drawback and extract all possible information from each depth, the introduced method applies pixel-normalization for each 2-dimensional slice separately.

### B. Collagen bundle detection

The proposed method is designed to detect and quantify the bundles only in a certain shape and visual condition.



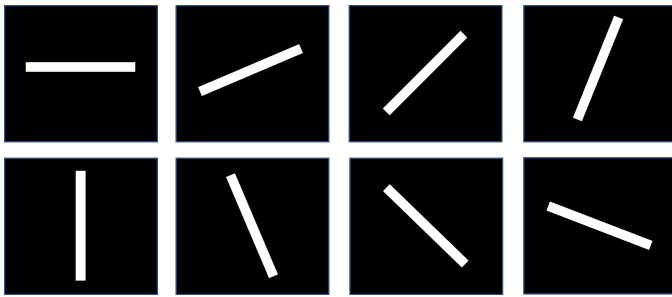


Fig. 5. The oriented rectangular filters evenly distributed into the X-Y plane.

In the SHG images, the collagen bundles are supposed to be in an anisotropic shape due to their structural nature. Hence, the proposed method ignores the isotropic high pixel intensity regions by accepting them as an artifact. Hence, the pixel intensity is not enough to detect bundles, but the intensity properties must be supported by geometric features. Since the images are noisy, and not all bundle regions are distinguishable, individual detection of each bundle is a challenging task. The proposed method is based on measuring the similarity of pixels based on their spatial and direction features. For that purpose, orientable unit filters are used to extract those geometrics features. Orientable unit filters are rectangular in shape and rotated in multiple orientations (see Fig. 5); thus, they can determine the dominant orientation of a close neighborhood of a point. Collagen bundle detection will be done in three steps:

- B1. Determination of dominant orientation for pixels in bundle region
- B2. Orientation similarity based Connected Component Analysis
- B3. Shape-constrained Pruning

Now, these three main steps will be explained in detail.

**B.1. Dominant orientation determination of pixels in the bundle region:** For that purpose, the input image is firstly convolved with the 3-dimensional rectangular prism directional filters. These filters are rotated in X-Y and X-Z plane with user determined angles (see Fig. 5 for rectangular filters rotated in X-Y plane.). The filter has size  $a \times b \times c$ , where  $a$  is the length,  $b$  is the height and  $c$  is the width. The filter size must be optimized according to the input image. After convolving the input image with filters in different orientations, the method determines the dominant orientation of the local neighborhood of each pixel by determining the direction of the maximum filtering response. As seen in Fig. 6, points have the maximum filtering response while the filter direction is close to the dominant orientation of the close neighborhood of the point. As a result, the orientation of a region can be determined by finding the general maxima of the filtering response. This idea is successfully used before for neurite orientation detection [38]. However, due to the high level of noise and complexity of the collagen images, further steps will be needed to use a similar idea for the current task. There are two major problems 1) intersecting bundle regions, and 2) large bundle regions which would give maximum filtering response at consecutive

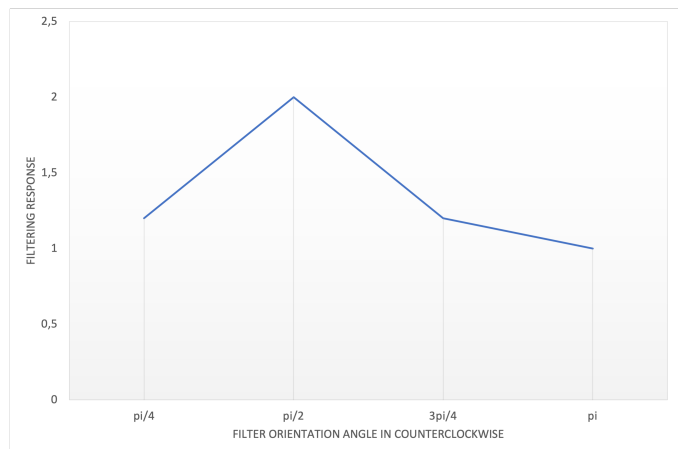
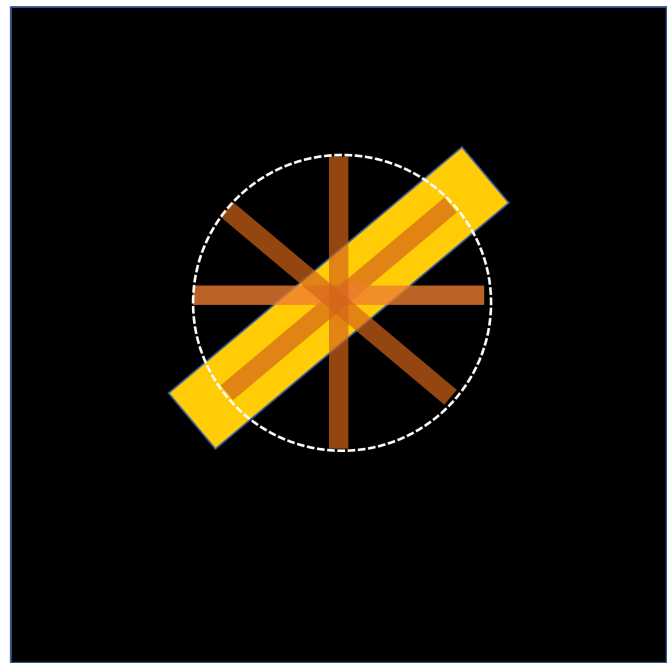


Fig. 6. The bundle orientation detection method on a synthetic image. (Top) A synthetic image where the collagen bundle is shown as the yellow anisotropic region on black background. Orange rectangular regions are the unit filters in four different orientations. (Bottom) The filtering response for each of the four filtering by (orange) oriented filters. The maximum filtering response is measured at the angle which is closest to the bundle direction ( $\pi/4$ ). That angle gives the dominant orientation of the close neighborhood (circle with white dashed boundary) of the pixel in the center.

directions.

In the case of intersecting bundles, which is a common situation for SHG images of collagen structures, the pixels in the overlapping region have multiple dominant orientations (see Fig. 7). Subsequently, determining the general maxima of the filtering response would be misleading for such cases. By considering these problems, the proposed method is designed to detect at most up to two intersecting bundles. For that purpose, instead of detecting one direction as a dominant orientation, the method collects two directions with the largest filtering responses. If the angle between two dominant directions is larger than a certain threshold value, then that pixel is



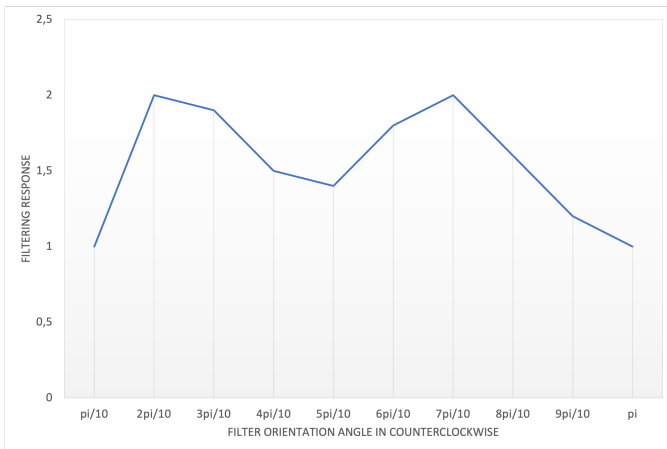
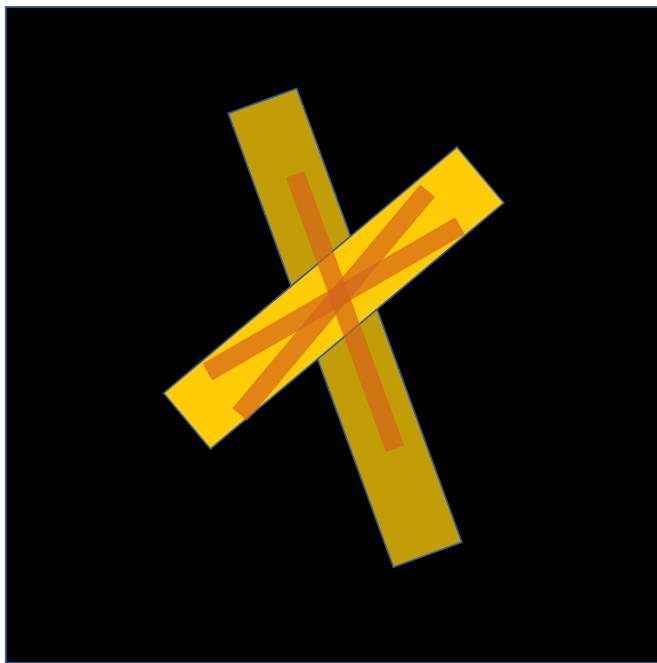


Fig. 7. The bundle orientation detection method on a synthetic image in the case of intersecting bundles or large bundle regions. (Top) A synthetic image where the collagen bundles are shown as the yellow anisotropic region on black background. Orange rectangular regions are the unit filters in different orientations. (Bottom) The filtering response of the centroid of the intersection of the bundles for ten oriented filters. As it is seen, the intersection point has two dominant orientations through two bundles at the angles  $2\pi/10$  and  $7\pi/10$ . Besides the filtering response at angle  $3\pi/10$  is close to the maximum although it is misleading due to the wide bundle region.

recorded with two dominant orientations. Otherwise, only one direction with the largest response is kept as the orientation of the neighborhood of the pixel. By examining the images in the dataset, more crowded intersection scenarios are neglected. The aim of using a lower threshold for the angle distance is to distinguish intersecting bundles from the single extra wide bundles. The extra wide bundles might have maximum filtering response in two consecutive angles even though they have a single dominant orientation (see Fig. 7). To avoid such problems, the algorithm assigns multiple dominant orientations to a pixel only if these orientations are distant from each other.

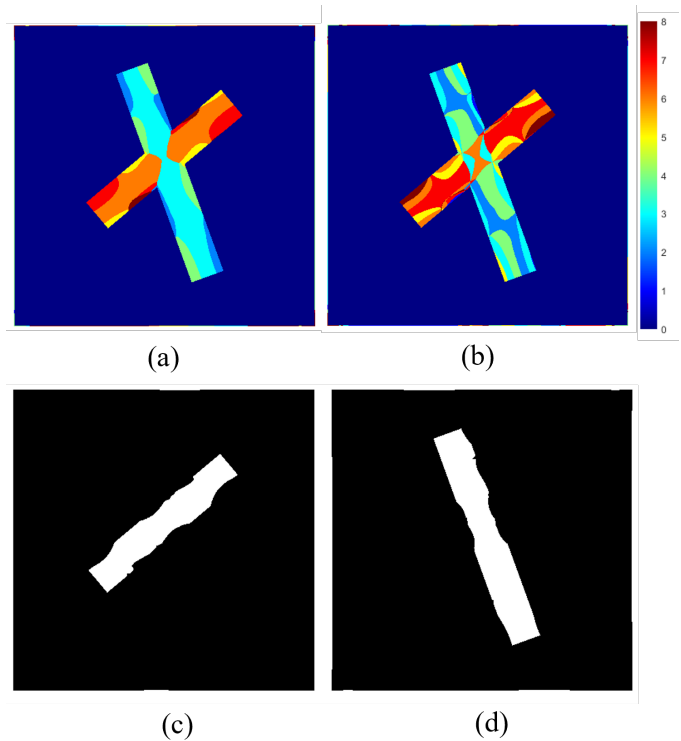


Fig. 8. Detection and segmentation of collagen bundles on a 2-dimensional synthetic image. Heat maps are showing the directions which yield maximum (a) and second maximum (b) filtering responses. These values show the dominant orientation of points in their local neighborhood. Due to the intersection of two bundles, there is a problem with assigning dominant orientations in their intersecting region. However, by considering these two filtering responses together and getting the connected components merged, both bundles are segmented with minimum error on boundaries (c-d).

**B.2. Connected Component Analysis based on dominant orientation similarity:** After the dominant orientation of each pixel is determined, pixels with the same dominant orientation are connected as being part of a collagen bundle. Through this process, the points with multiple dominant orientations are preserved in their nature and all possible connection paths are examined. As a result of this process, even the intersecting bundles are segmented separately by the proposed algorithm (Fig. 8).

**B.3. Pruning:** Since the nature of the collagen bundles is anisotropic, the post-processing step starts with determining the anisotropy level of segmented bundle regions through the Directional Ratio (DR). DR is a multiscale geometric descriptor to measure the isotropy level of the close neighborhood of a pixel. DR is calculated as the proportion of minimum to maximum filtering responses of a pixel over filters in multiple orientations [39]. DR approaches 1 while regions get more isotropic, and it approaches 0 while regions get more anisotropic [40], [41], [42]. It is used in previous studies to distinguish blob-like objects from vessel-like objects [40], [38], [39], [41]. For pruning, the algorithm calculates the DR of each connected component as the average DR of each pixel contained in the inner part of the region. For any region, even if the region is perfectly isotropic, the DR values of pixels will approach 0 while getting closer to the boundary [42]. Hence,

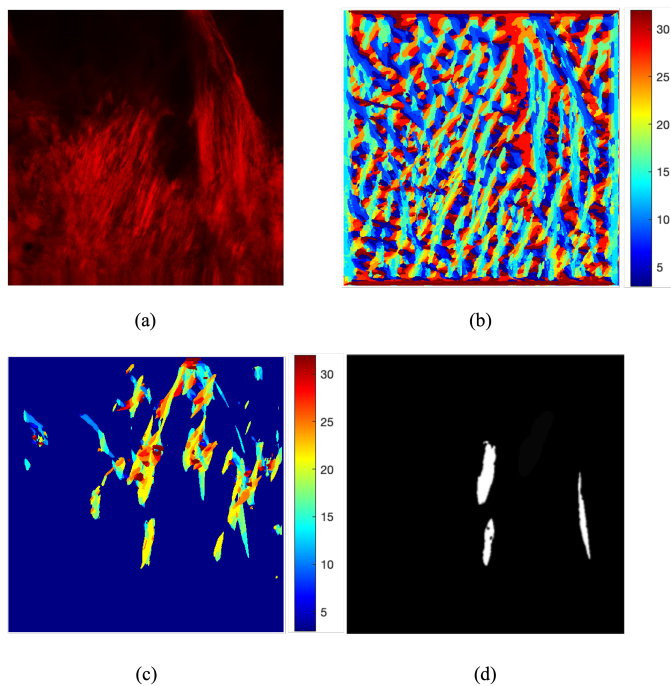


Fig. 9. Sample outputs of the intermediate steps in collagen bundle detection: a) Preprocessed image, b) Output of step B1. Dominant orientation determination of pixels in the bundle region. Since the orientations are enumerated as 1-32, this image has pixel intensities between 1-32. c) Output of step B2. Connected Component Analysis based on dominant orientation similarity. d) output of step B3. Pruning: Components in c) are eliminated if detections are isotropic and smaller than a certain volume threshold. Please note that these outputs are actually 3-dimensional images, however for visualization 2dimensional projections are given.

pixels close to the boundary are misleading to calculate the isotropy level of a region. Therefore, these misleading pixels are eliminated through removing the close-boundary parts of regions by morphological erosion. The remaining pixels' DR values are averaged to assign a DR value for each connected component. This DR value shows the anisotropy level of the whole region. To eliminate the isotropic regions a lower DR value threshold is applied. If the DR value is larger than the threshold that means the component is more isotropic than a usual collagen bundle. Hence, it will be accepted as an artifact and eliminated. As a result, the output image contains the segmented anisotropic collagen bundles. The computational cost of this part is neglectable since the filtering responses were already calculated in part B.1. for Dominant orientation determination of pixels.

### III. RESULTS

The performance of the proposed method is measured on a publicly available dataset of SHG volume collected from three healthy rats [43]. The dataset contains around 4000 scans, but some of these images do not show the orientated nature of collagen bundles. Such images are eliminated and as a result, four volumes are selected to measure the performance of the method. A few slices of the samples which are ignored for the analysis can be seen in Fig. 10. As it is seen in the figure, these slices do not have oriented anisotropic patterns. Hence, such images are out of the scope of the presented method.

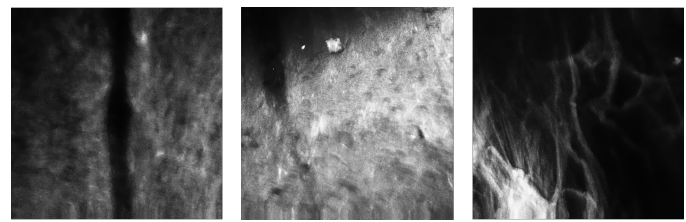


Fig. 10. Example images from the dataset that cannot be processed by the proposed method. These images do not have orientation information, bundles are not detectable by the eye and there are artifacts. Such images are ignored for sample selection for analysis.

The ground truth of the dataset was for semantic segmentation of similarly oriented regions. The pixels in the ground-truth images are in three categories as similar orientation, dissimilar orientation, and not interest. There was no label for direction or bundle detection. The presented method generates more local information as the pattern and orientation detection. Hence, the given ground truth was not useful for measuring the performance of the method. For that purpose, the detection and orientation labels are determined by the experts at Istanbul Medipol University, and the performance of the method is measured according to that.

The method has several parameters to be tuned for the given input. For the presented results, those parameters are optimized through the whole dataset, and these values are set as Default in the method. Filter size is set as 40x4x4 (with respect to x,y,z coordinates) pixels to catch the large anisotropic bundles efficiently. The analysis is performed in 8 directions in the X-Y-plane and 4 directions in the X-Z-plane, a total of 32 orientations. Since the images are so crowded and the same collagen bundle is visible in multiple slices, ground truth was collected for random 80 slices from samples by considering their 3-dimensional environment. The connected components whose volume is smaller than 3000 pixels are accepted as artifacts and eliminated. For each of the remaining connected components, the DR value is calculated, as explained in the Method section. The components with DR larger than 0.9 are eliminated since they are more isotropic than an accepted bundle. Then, the centroid of each connected component is calculated and detection results are demonstrated with a line that is oriented with the same angle of the bundle and centered at the centroid of the segmented region. As a result, large and vessel-like detections are collected as collagen bundles. The results of a few slices can be seen in Fig. 11, 14 and 15.

To measure the performance of the method Recall, Precision, and F1 score are used. Recall shows the performance of avoiding False negative detections, Precision shows the performance of avoiding False positive detections; while the F1 score shows the overall performance considering False Positives and Negatives. All these measurements range between 0 and 1, where 1 is the perfect score for each [44].

The method correctly detected 42 collagen bundles, out of 51 (Table I). The Recall score of the method is measured as 0.82. The Precision rate is measured low due to high False Positive detections. However, about half of the False detections were due to multiple detections of wide bundles. If such

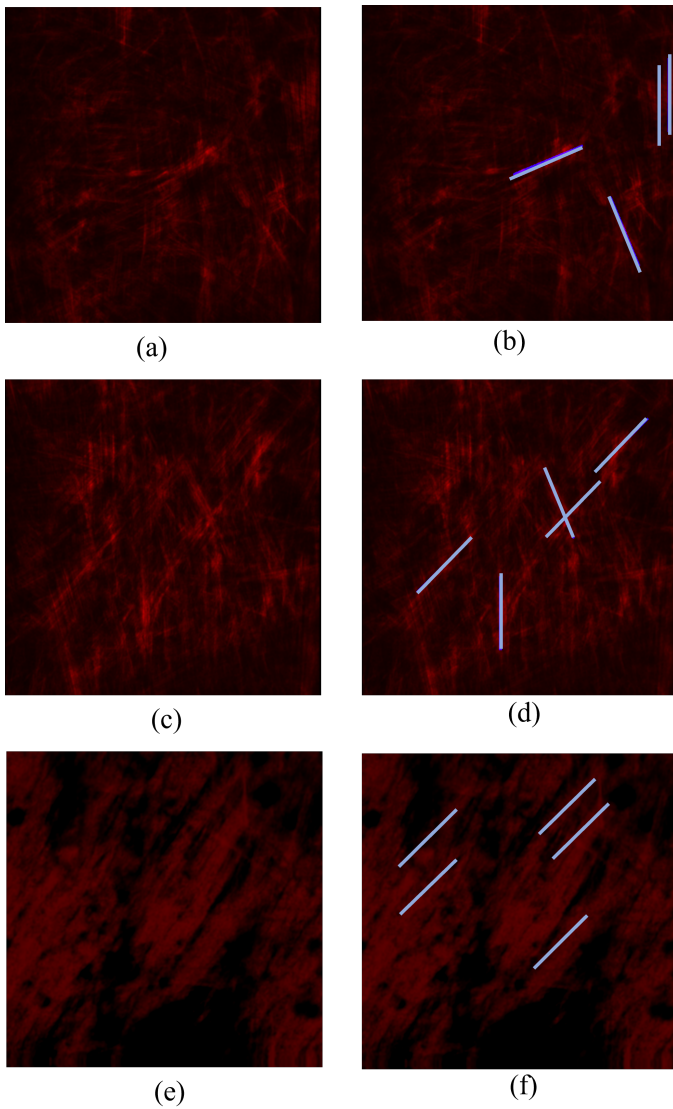


Fig. 11. Results of proposed method for collagen bundle detection. The raw image slices are shown in the right column (a, c, e), and associated detection results are given on the left (b, d, f) as an overlay on the original slices. Detected bundles are visualized with blue lines. The orientation of these lines indicates the dominant orientation of the detected collagen region.

	TP	FP	FN
multiple detections are included	42	19	9
multiple detections are excluded	42	11	9

TABLE I

DETECTION RESULTS OF THE METHOD WITH AND WITHOUT THE ELIMINATION OF FALSE POSITIVES DUE TO MULTIPLE DETECTIONS OF WIDE BUNDLES. TP STANDS FOR TRUE POSITIVE, FP STANDS FOR FALSE POSITIVE, AND FN STANDS FOR FALSE NEGATIVE DETECTIONS.

false detections were eliminated the Precision rate increased to 0.79 from 0.69 (Table II). Since orientation is the core of the proposed detection method, the accuracy of orientation detection for detected bundles was 100%.

#### IV. DISCUSSION

The SHG images of collagen bundles are highly complex and low-contrasted. Since the edge information is almost

	P	R	F1
multiple detections are included	0.69	0.82	0.75
multiple detections are excluded	0.79	0.82	0.81

TABLE II

PERFORMANCE RESULTS OF THE METHOD WITH AND WITHOUT ELIMINATION OF FALSE POSITIVES DUE TO MULTIPLE DETECTIONS OF WIDE BUNDLES. P STANDS FOR PRECISION, R STANDS FOR RECALL, AND F1 STANDS FOR F1 SCORE.

missing, collagen bundles seem like regions with uncertain boundaries. As a result, most detection methods fail for such cases. However, the presented method gives promising results through its advantage of using a connected component analysis through orientation similarity. This idea skips the edge or boundary information and focuses on local texture, which results in successful detection performance for the current task. As it is seen in Table II, the method has a high Recall score. False Positive detections are high despite low False Negative detections, which results in a lower Precision score (Table I & II). It is observed that some of the False Positives are due to multiple detections of one wide collagen bundle region. For example, in Fig. 11 (b), the rightmost two vertical detections are indicating the same region. Although this is not a significant error of the method, it could cause a problem if the detected bundle number is in focus since it will be overcalculated. In order to handle that an additional post-processing step can be included. When these multiple detections of wide bundles are ignored, the False Positive count reduces by almost 50% (Table I) and the F1 score increases from 0.75 to 0.81 (Table II).

On the other hand, False Positive detections are higher for the images where bundles are crossing each other wildly (Fig. 14). In such images, patterns are formed by short anisotropic regions, and large-scale filters cause the mixing of the information of multiple bundles. As a result, high False Positive detections occur. In order to handle that, a small-scale filter or multi-scale analysis could be used. By using multi-scale filters, short patterns can be detected as well as longer patterns.

The collagen bundles close to each other may have different orientations. As a result, in the SHG images, multiple collagen bundles crossing each other can be seen. Since this is a common issue, the proposed method is specifically designed to handle such cases. For that purpose, the method always examines if the points have two dominant orientations, as explained in detail in Section II. However, this may be misleading in some cases. For example, when the contrast is critically low, the method can be faulted to see noisy regions as if they are intersection regions of multiple bundles. This causes False Positive detections (Fig. 14 (d)). In order to handle that, a lower threshold for pixel intensity can be used. If the determined bundle region has an average pixel intensity lower than the threshold, it can be considered a noisy region instead of a bundle.

Another important aspect of the method is that not all bundle regions are targeted by the presented method. The bundles with certain shape and intensity properties can be detected. However, the method gives a consistent highly local texture



measurement. As a result, the number of detections as well as the orientation of detections can be significant indicators of multiple disorders.

The proposed method has multiple steps and for each step there are parameters to be adjusted for the input. Throughout the analysis it is observed the method's sensitivity to the selection of these parameters are different (see Table III). In the pre-processing step the method is sensitive to one parameter which is "Maximum filtering projection (MIP) window scale". That parameter must be adjusted according to the X-Z direction resolution of the input image. If the resolution is high default value can be increased to optimize the computational cost. But, the detection result is not highly sensitive to that parameter. Similarly, method is sensitive or highly sensitive to the several threshold values. When the input image has wildly crossing and thin bundle regions, "Threshold to determine intersecting bundles" in step B1 can be reduced. Otherwise multiple distinct bundles in close orientation can be detected as a whole (This would be a risk as well if the number of orientations is increased. So this threshold parameter is closely related with the user-determined number of orientation.). "The threshold to eliminate small regions" in step B2 can be decreased if the input image is highly noisy. "Threshold to eliminate isotropic regions" must be adjusted by observing the results. The complete list of these parameters and their default values are given in Table III.

During the analysis, it is observed that the method is highly sensitive to the number of orientations of the analysis and the scale of the filters. This is due to high noise and uncertain boundaries of collagen bundles. It is observed that when the number of orientations are increased the method tends to detect the same bundle multiple times with slightly different orientations. As a result of that, False Positives increase and the Precision rate decreases; or bundles are detected with incorrect orientation (see Fig. 12 (a)). On the other hand, when the number of orientations is increased a highly sensitive analysis results with one bundle region detected in several disconnected and differently oriented partitions. These partitions with different orientations can not be attached to each other by the connected component analysis (step B2). As a result of that, some large bundle regions can not be detected. As it is seen in (Fig. 12 (c)) two bundle regions are skipped by the method when the number of X-Y plane orientation is set to 20. For the current dataset, 8-10 orientations in the X-Y plane and 4 orientation on the X-Z plane was optimal.

Similarly, it was observed that method is sensitive to the selection of filter scale. When the filter scale is small the case of multiple detection of the same bundle is more probable. As a result of that, the number of False Positives increase and the Precision rate decreases. However, when the bundles are crossing with each other widely smaller-scale filters would work better. In order to measure the proposed method's sensitivity to the filter scale selection, 40 slices from two volumes are selected to reduce the computational burden. These samples were chosen carefully to cover all different situations to be fair on the sensitivity analysis. These two volumes were analysed with varying filter scales and the performance metrics are compared with each other. As it is seen in Fig. 13, for any

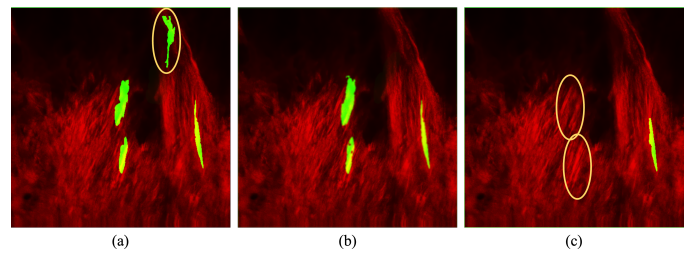


Fig. 12. Detected collagen bundles outputs when the proposed method is performed with different number of orientations. A) X-Y plane orientation is 4, B) X-Y plane orientation is 8 and C) X-Y plane orientation is 20. For all cases X-Z plane orientation number was fixed to 4. Since the resolution of volumes in X-Z plane is smaller than X-Y, the parameter sensitivity analysis on X-Z plane is ignored. Errors are shown by yellow ellipses.

Method Step	Parameter	Sensitivity of method to the selection	Default
Preprocessing	MIP window size	Sensitive	3 pixels in XZ direction
	Denoising filter scale	Low sensitivity	1
Collagen Bundle Detection B1	Threshold to determine intersecting bundles	Highly sensitive	$\pi/8$
	Filter scale	Highly sensitive	40 x 4 x 4
	Orientation number	Sensitive	8 x 4
Connected Component Analysis B2	Threshold to eliminate small regions	Sensitive	5000 pixels
Collagen Bundle Detection B3	Threshold to eliminate isotropic regions	Sensitive	0.93
	Erosion filter scale	Low sensitivity	4 pixels

TABLE III  
LIST OF PARAMETERS TO BE ADJUSTED FOR INPUT

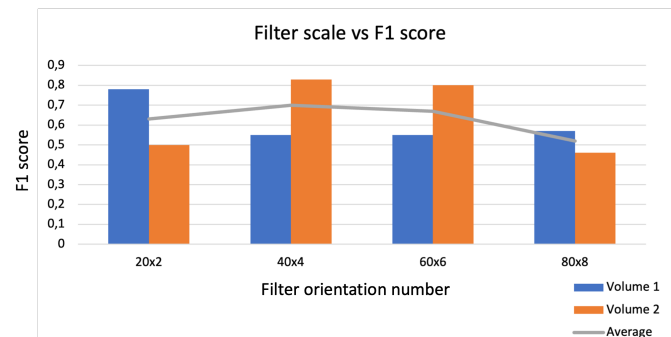


Fig. 13. The change in F1 score with the analysis with different filter scales. X-Z plane analysis is ignored since the resolution in X-Z direction is small. Given filter scales are for the length and width respectively. For each case height was set as same with width. As it is seen the average F1 score is maximized when 40x4 scale filters are used. However, for volume 1 the performance is maximized with the smallest scale 20x2.

filter scale the variation of F1 scores between the analysis of different volumes were critically high. Besides, the results show that the accuracy of the method is highly sensitive to the selection of filter scale. As it is seen in Fig. 13, in Volume 1 the maximum performance is obtained with smallest scale filter, while that scale yielded the worst performance result in the Volume 2. As a result, in order to show the overall performance of the method with optimal effort, parameters are numerically optimized and set to a one value for the whole dataset. Hence, the reported numerical results can be improved by optimizing parameters individually for each input volume. For the further studies, an automated parameter optimization step could be included.

The dataset originally had ground-truth labels for semantic



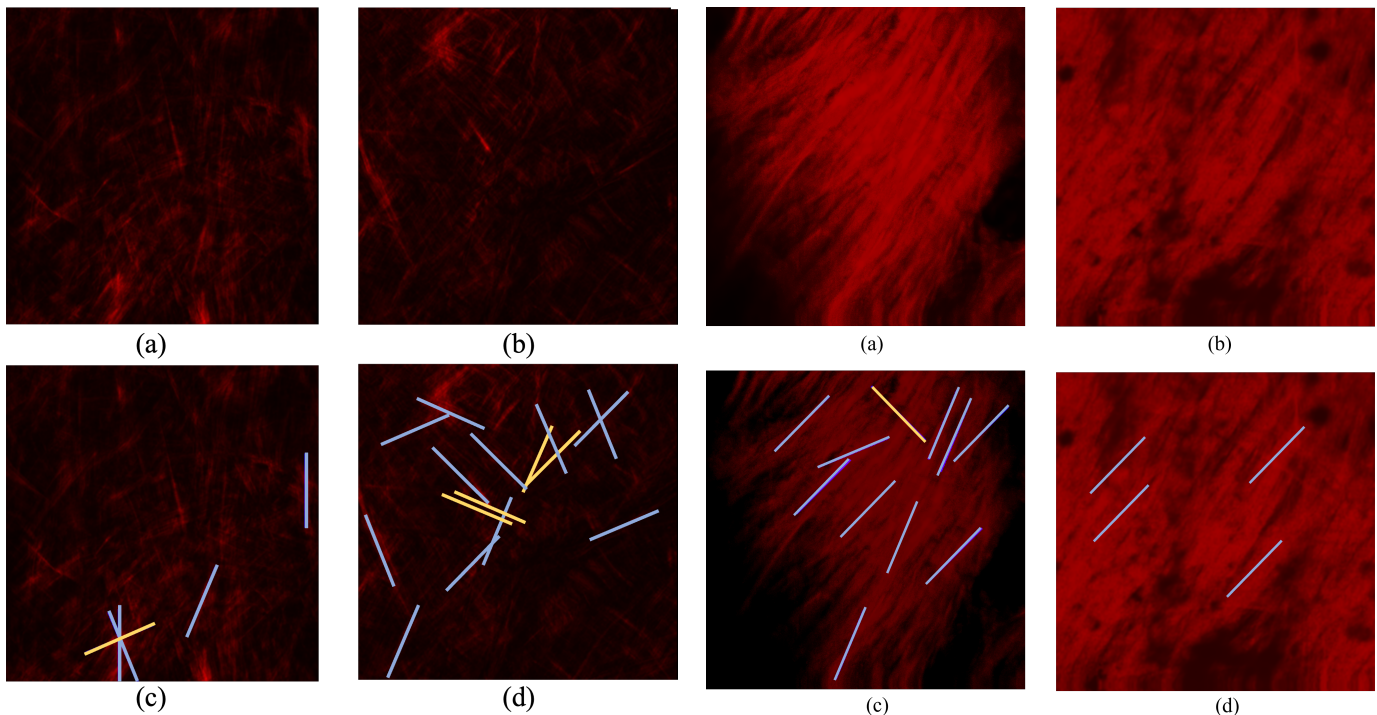


Fig. 14. Some examples for False Positive detection: Raw images are given in (a), (b), and the corresponding detection are given (c), (d). True Positives are represented with blue lines while False Positives are represented with yellow.

segmentation. Since the proposed method has no application for pixel segmentation, the given labels could not be used. However, they can still be used to show the consistency of the presented method. As it is seen in Fig. 15 (e-f), the correctly detected bundles in the green region, which indicates the ground-truth labels for similarly oriented regions, are in highly close orientations.

In order to compare the performance of the proposed method with the state of the art methods, CT-FIRE is selected as a reference. CT-FIRE is the best candidate since it has the similar objectives with the presented method. Besides CT-FIRE has applications in the literature of collagen bundle analysis [22]. Since the CT-FIRE is analyzing volumes slice by slice the 3-dimensional analysis was computationally highly expensive. As a result, instead of analyzing the whole dataset, the performance comparison is done only on the half of the dataset. For that purpose, 40 slices from each volume is collected instead of 80. By observing the results, it can be said that comparison is fair due to the consistent behavior of CT-FIRE based on errors and successes. To have a fair comparison, parameters of CT-FIRE are optimized numerically for the all test samples. When we compared the performance of CT-FIRE with the proposed method, it is observed that CT-FIRE tends to have high False Positive detections (see Table IV). One major reason is that CT-FIRE fails to detect regions close to the image boundaries as a bundle due to high contrast change (Fig. 16 (a), (c)). As a result, for all test images, CT-FIRE yield a very large False Positive detection value when it is compared with the proposed method. Besides, CT-FIRE detected some bundle regions with wrong angle.

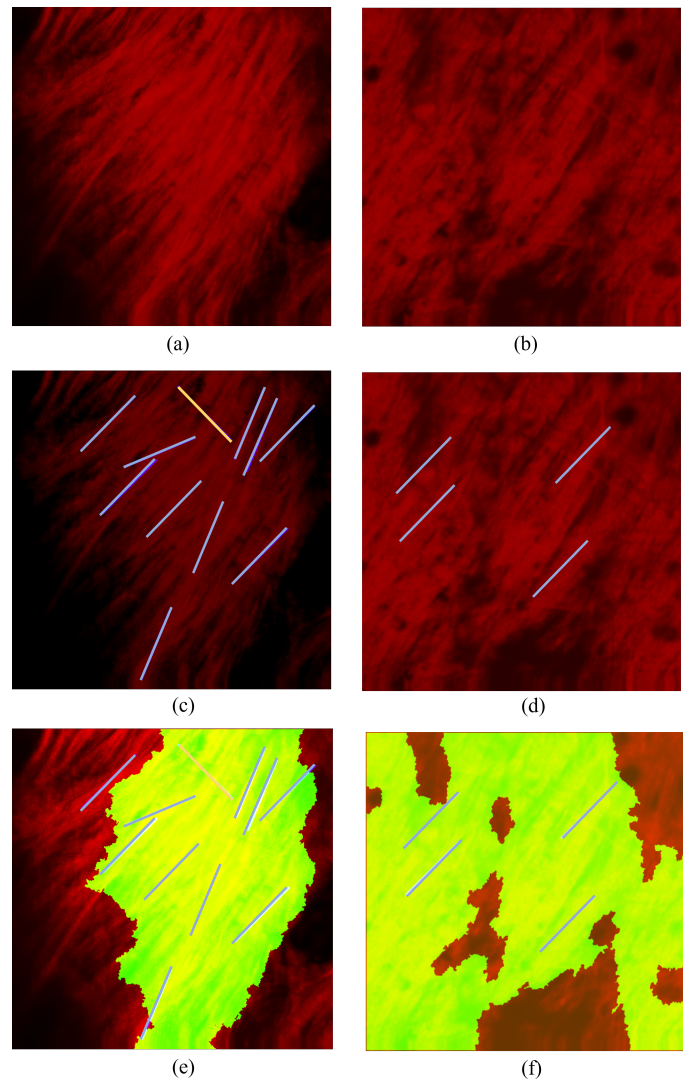


Fig. 15. Comparison of detection results with the ground-truth pixel-wise orientation similarity labels. a-b) original raw images, c-d) overlay of the presented method's detections in blue with the raw images. There is only 1 False Positive detection, which is represented with the orange line. e-f) overlay of ground-truth labels for similarly oriented regions in the green channel, detected bundles in blue, and bundle regions in red.

These detections were counted as False Positive again, since the bundle orientation is a significant morphological feature. As it is seen in the Table IV the proposed method has F1 score 94 % while CT-FIRE reaches 44 %. Although CT-FIRE is successful to detect vessel-like structures, this method is so sensitive to contrast change. As a result, in the noisy images such as SHG images of collagen bundles, it could not reach the performance of proposed method.

The performance analysis of the method is done on samples from healthy rats. This could be a limitation since the method has not been applied to any sample with an abnormality. The structural properties of fibrillar collagen bundles tend to vary due to countless diseases and abnormalities. One of these properties is known as the curvature of the bundle both on small and large scales [45], [46]. Although the presented method is designed only to detect linear bundles, it is observed

Filter scale	TP	FN	FP	P	R	F1
CT-FIRE	11	11	17	0.39	0.5	0.44
Presented Method	21	2	1	0.96	0.91	0.94

TABLE IV

COMPARISON OF PERFORMANCE METRICS RESULT FOR CT-FIRE AND PRESENTED METHOD.

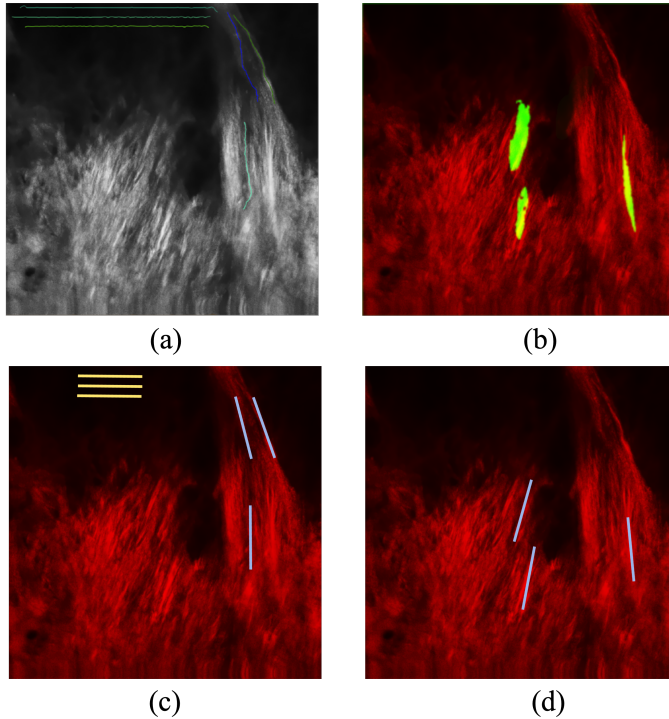


Fig. 16. Comparison of outputs of CT-FIRE and presented algorithm: a-b outputs of CT-FIRE and presented method respectively; c-d a better visualization of detections results with representative rectangular markers. Correct detections are represented with blue and wrong detections represented with orange rectangular marker.

that the method can also detect curvatures on a large scale (Fig. 17). This leads to the idea that by adjusting the filter scale the presented method would be able to determine curvy collagen bundles which are seen as a sign of multiple abnormalities [45], [46]. Hence, even if the method is only validated with healthy collagen bundles, it still has the ability to detect abnormalities in morphology by tuning the parameters. As a result, the presented method can be used for classification purposes to distinguish healthy samples from diseased samples.

## V. CONCLUSION

Due to the advantages of SHG and the critical importance of collagen bundles, several studies in the literature focus on SHG images of collagen bundles. However, most of these studies are designed to detect non-local statistical properties of collagen distribution. This led to the loss of local morphological information collagen bundles which are known to be highly significant indicators of multiple abnormalities. The introduced method uses the orientation information to detect individual collagen bundles in 3-dimensional SHG images even if they intersect each other. Hence, the presented method gives deeper

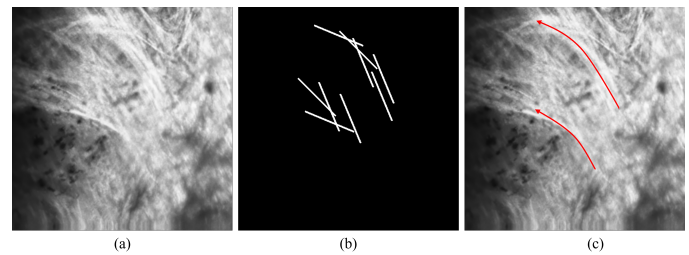


Fig. 17. Curvature detection: (a) raw image, (b) corresponding detected bundles given in white lines, (c) curvy regions path is shown with red arrow. As it is seen through multiple detections of bundles the large curvatures can be determined by the method. This led to the idea that by using small-scale filters it would be possible to detect curvy collagen bundles' structures.

and more detailed information about the distribution and structure of collagen bundles in SHG images. That would lead to more accurate data analysis methods for collagen analysis, which finally lead to better diagnosis methods for collagen-related abnormalities such as cornea diseases and some cancer types.

## VI. ACKNOWLEDGMENT

The author thanks Dr. Cafer Tanrıverdi, Mehmet Şerif Aydın, and Olgu Enis Tok for their expertise and valuable assistance for validating the proposed method.

## REFERENCES

- [1] K. M. Meek and C. Knupp, "Corneal structure and transparency," *Progress in retinal and eye research*, vol. 49, pp. 1–16, 2015.
- [2] C. Raoux, M. Schmeltz, M. Bied, M. Alnawaiseh, U. Hansen, G. Latour, and M.-C. Schanne-Klein, "Quantitative structural imaging of keratoconic corneas using polarization-resolved shg microscopy," *Biomedical optics express*, vol. 12, no. 7, pp. 4163–4178, 2021.
- [3] M. S. Sridhar, "Anatomy of cornea and ocular surface," *Indian journal of ophthalmology*, vol. 66, no. 2, p. 190, 2018.
- [4] H.-Y. Zhou, Y. Cao, J. Wu, and W.-S. Zhang, "Role of corneal collagen fibrils in corneal disorders and related pathological conditions," *International journal of ophthalmology*, vol. 10, no. 5, p. 803, 2017.
- [5] G. A. Di Lullo, S. M. Sweeney, J. Korkko, L. Ala-Kokko, and J. D. San Antonio, "Mapping the ligand-binding sites and disease-associated mutations on the most abundant protein in the human, type i collagen," *Journal of Biological Chemistry*, vol. 277, no. 6, pp. 4223–4231, 2002.
- [6] S. Xu, H. Xu, W. Wang, S. Li, H. Li, T. Li, W. Zhang, X. Yu, and L. Liu, "The role of collagen in cancer: from bench to bedside," *Journal of translational medicine*, vol. 17, pp. 1–22, 2019.
- [7] R. M. Martínez-Ojeda, M. D. Pérez-Cárceles, L. C. Ardelean, S. G. Stanciu, and J. M. Bueno, "Multiphoton microscopy of oral tissues," *Frontiers in Physics*, vol. 8, p. 128, 2020.
- [8] E. A. Gibson, O. Masihzadeh, T. C. Lei, D. A. Ammar, and M. Y. Kahook, "Multiphoton microscopy for ophthalmic imaging," *Journal of ophthalmology*, vol. 2011, 2011.
- [9] R. LaComb, O. Nadiarykh, and P. J. Campagnola, "Quantitative second harmonic generation imaging of the diseased state osteogenesis imperfecta: experiment and simulation," *Biophysical journal*, vol. 94, no. 11, pp. 4504–4514, 2008.
- [10] S. V. Plotnikov, A. C. Millard, P. J. Campagnola, and W. A. Mohler, "Characterization of the myosin-based source for second-harmonic generation from muscle sarcomeres," *Biophysical journal*, vol. 90, no. 2, pp. 693–703, 2006.
- [11] S.-W. Chu, S.-Y. Chen, G.-W. Chern, T.-H. Tsai, Y.-C. Chen, B.-L. Lin, and C.-K. Sun, "Studies of  $\chi(2)/\chi(3)$  tensors in submicron-scaled bio-tissues by polarization harmonics optical microscopy," *Biophysical journal*, vol. 86, no. 6, pp. 3914–3922, 2004.
- [12] Z. Liu, K. P. Quinn, L. Speroni, L. Arendt, C. Kuperwasser, C. Sonnenschein, A. M. Soto, and I. Georgakoudi, "Rapid three-dimensional quantification of voxel-wise collagen fiber orientation," *Biomedical optics express*, vol. 6, no. 7, pp. 2294–2310, 2015.
- [13] Z. Liu, D. Pouli, D. Sood, A. Sundarakrishnan, C. K. H. Mingalone, L. M. Arendt, C. Alonzo, K. P. Quinn, C. Kuperwasser, L. Zeng *et al.*, "Automated quantification of three-dimensional organization of fiber-like structures in biological tissues," *Biomaterials*, vol. 116, pp. 34–47, 2017.
- [14] E. C. Rentchler, K. L. Gant, R. Drapkin, M. Patankar, and P. J. Campagnola, "Imaging collagen alterations in stics and high grade ovarian cancers in the fallopian tubes by second harmonic generation microscopy," *Cancers*, vol. 11, no. 11, p. 1805, 2019.
- [15] J. M. Watson, P. F. Rice, S. L. Marion, M. A. Brewer, J. R. Davis, J. J. Rodriguez, U. Utzinger, P. B. Hoyer, and J. K. Barton, "Analysis of second-harmonic-generation microscopy in a mouse model of ovarian carcinoma," *Journal of Biomedical Optics*, vol. 17, no. 7, pp. 076002–076002, 2012.
- [16] D. S. James and P. J. Campagnola, "Recent advancements in optical harmonic generation microscopy: Applications and perspectives," *BME Frontiers*, vol. 2021, 2021.
- [17] Y. Ogura, Y. Tanaka, E. Hase, T. Yamashita, and T. Yasui, "Texture analysis of second-harmonic-generation images for quantitative analysis of reticular dermal collagen fibre in vivo in human facial cheek skin," *Experimental Dermatology*, vol. 28, no. 8, pp. 899–905, 2019.
- [18] T. Hsu, A. Calway, and R. Wilson, "Texture analysis using the multiresolution fourier transform," *Bristol, UK, Tech. Rep.*, 1993.
- [19] W. Hu, H. Li, C. Wang, S. Gou, and L. Fu, "Characterization of collagen fibers by means of texture analysis of second harmonic generation images using orientation-dependent gray level co-occurrence matrix method," *Journal of biomedical optics*, vol. 17, no. 2, pp. 026007–026007, 2012.
- [20] C. C. Gotlieb and H. E. Kreytzig, "Texture descriptors based on co-occurrence matrices," *Computer vision, graphics, and image processing*, vol. 51, no. 1, pp. 70–86, 1990.
- [21] C. Y. Park, J. K. Lee, and R. S. Chuck, "Second harmonic generation imaging analysis of collagen arrangement in human cornea," *Investigative ophthalmology & visual science*, vol. 56, no. 9, pp. 5622–5629, 2015.
- [22] M. Sendín-Martín, J. Posner, U. Harris, M. Moronta, J. Conejo-Mir Sánchez, S. Mukherjee, M. Rajadhyaksha, K. Kose, and M. Jain, "Quantitative collagen analysis using second harmonic generation images for the detection of basal cell carcinoma with ex vivo multiphoton microscopy," *Experimental Dermatology*, vol. 32, no. 4, pp. 392–402, 2023.
- [23] Y. Liu, A. Keikhosravi, G. S. Mehta, C. R. Drifka, and K. W. Eliceiri, "Methods for quantifying fibrillar collagen alignment," *Fibrosis: methods and protocols*, pp. 429–451, 2017.
- [24] J. S. Bredfeldt, Y. Liu, C. A. Pehlke, M. W. Conklin, J. M. Szulcowski, D. R. Inman, P. J. Keely, R. D. Nowak, T. R. Mackie, and K. W. Eliceiri, "Computational segmentation of collagen fibers from second-harmonic generation images of breast cancer," *Journal of biomedical optics*, vol. 19, no. 1, pp. 016007–016007, 2014.
- [25] Y. Liu, A. Keikhosravi, C. A. Pehlke, J. S. Bredfeldt, M. Dutson, H. Liu, G. S. Mehta, R. Claus, A. J. Patel, M. W. Conklin *et al.*, "Fibrillar collagen quantification with curvelet transform based computational methods," *Frontiers in bioengineering and biotechnology*, vol. 8, p. 198, 2020.
- [26] J. Liu, M.-y. Xu, J. Wu, H. Zhang, L. Yang, D.-x. Lun, Y.-c. Hu, and B. Liu, "Picosirius-polarization method for collagen fiber detection in tendons: A mini-review," *Orthopaedic Surgery*, vol. 13, no. 3, pp. 701–707, 2021.
- [27] Y. Zhang, Y. Chen, B. Zhao, J. Gao, L. Xia, F. Xing, Y. Kong, Y. Li, and G. Zhang, "Detection of type i and iii collagen in porcine acellular matrix using hplc-ms," *Regenerative biomaterials*, vol. 7, no. 6, pp. 577–582, 2020.
- [28] X. Wang, B. Le, N. Zhang, K. H. Bak, Y. Zhang, and Y. Fu, "Off-flavour compounds in collagen peptides from fish: Formation, detection and removal," *International Journal of Food Science & Technology*, vol. 58, no. 3, pp. 1543–1563, 2023.
- [29] J. Zhang, Y. Ning, H. Zhu, N. J. Rotile, H. Wei, H. Diyabalanage, E. C. Hansen, I. Y. Zhou, S. C. Barrett, M. Sojoodi *et al.*, "Fast detection of liver fibrosis with collagen-binding single-nanometer iron oxide nanoparticles via t 1-weighted mri," *Proceedings of the National Academy of Sciences*, vol. 120, no. 18, p. e2220036120, 2023.
- [30] M. Salarian, H. Yang, R. C. Turaga, S. Tan, J. Qiao, S. Xue, Z. Gui, G. Peng, H. Han, P. Mittal *et al.*, "Precision detection of liver metastasis by collagen-targeted protein mri contrast agent," *Biomaterials*, vol. 224, p. 119478, 2019.
- [31] O. Y. Ibhagui, D. Li, H. Han, G. Peng, M. L. Meister, Z. Gui, J. Qiao, M. Salarian, B. Dong, Y. Yuan *et al.*, "Early detection and staging of lung fibrosis enabled by collagen-targeted mri protein contrast agent," *Chemical & Biomedical Imaging*, 2023.
- [32] C. Roa, V. N. Du Le, M. Mahendroo, I. Saytashev, and J. C. Ramella-Roman, "Auto-detection of cervical collagen and elastin in mueller matrix polarimetry microscopic images using k-nn and semantic segmentation classification," *Biomedical Optics Express*, vol. 12, no. 4, pp. 2236–2249, 2021.
- [33] M. Zaffar and A. Pradhan, "Assessment of anisotropy of collagen structures through spatial frequencies of mueller matrix images for cervical pre-cancer detection," *Applied Optics*, vol. 59, no. 4, pp. 1237–1248, 2020.
- [34] J. Li, Y. Chen, W. Zhi, and Q. Cheng, "Photoacoustics spectral analysis for in vivo detection of collagen contents in cancers," in *2022 IEEE International Ultrasonics Symposium (IUS)*. IEEE, 2022, pp. 1–4.
- [35] J.-E. Cota, A. Spadigam, and A. Dhupar, "Detection of type vii collagen in odontogenic keratocyst: An immunohistochemical study," *Journal of clinical and experimental dentistry*, vol. 11, no. 4, p. e310, 2019.
- [36] J. T. Stefano, L. V. Guedes, A. A. A. de Souza, D. S. Vanni, V. A. F. Alves, F. J. Carrilho, A. Largura, M. Arrese, and C. P. Oliveira, "Usefulness of collagen type iv in the detection of significant liver fibrosis in nonalcoholic fatty liver disease," *Annals of hepatology*, vol. 20, p. 100253, 2021.
- [37] R. C. Gonzalez, *Digital image processing*. Pearson education india, 2009.
- [38] C. Kayasandik, P. Negi, F. Laezza, M. Papadakis, and D. Labate, "Automated sorting of neuronal trees in fluorescent images of neuronal networks using neurotree-tracer," *Scientific reports*, vol. 8, no. 1, p. 6450, 2018.
- [39] D. Labate, F. Laezza, P. Negi, B. Ozcan, and M. Papadakis, "Efficient processing of fluorescence images using directional multiscale representations," *Mathematical modelling of natural phenomena*, vol. 9, no. 5, pp. 177–193, 2014.
- [40] C. B. Kayasandik and D. Labate, "Improved detection of soma location and morphology in fluorescence microscopy images of neurons," *Journal of neuroscience methods*, vol. 274, pp. 61–70, 2016.

- [41] B. Ozcan, P. Negi, F. Laezza, M. Papadakis, and D. Labate, "Automated detection of soma location and morphology in neuronal network cultures," *PLoS one*, vol. 10, no. 4, p. e0121886, 2015.
- [42] C. Kayasandik, K. Guo, and D. Labate, "Directional multiscale representations and applications in digital neuron reconstruction," *Journal of computational and applied mathematics*, vol. 349, pp. 482–493, 2019.
- [43] L. Schmarje, C. Zelenka, U. Geisen, C.-C. Glüer, and R. Koch, "2d and 3d segmentation of uncertain local collagen fiber orientations in shg microscopy," in *Pattern Recognition: 41st DAGM German Conference, DAGM GCPR 2019, Dortmund, Germany, September 10–13, 2019, Proceedings 41*. Springer, 2019, pp. 374–386.
- [44] D. M. Powers, "Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation," *arXiv preprint arXiv:2010.16061*, 2020.
- [45] J. N. Ouellette, C. R. Drifka, K. B. Pointer, Y. Liu, T. J. Lieberthal, W. J. Kao, J. S. Kuo, A. G. Loeffler, and K. W. Eliceiri, "Navigating the collagen jungle: the biomedical potential of fiber organization in cancer," *Bioengineering*, vol. 8, no. 2, p. 17, 2021.
- [46] E. Brown, T. McKee, E. DiTomaso, A. Pluen, B. Seed, Y. Boucher, and R. K. Jain, "Dynamic imaging of collagen and its modulation in tumors in vivo using second-harmonic generation," *Nature medicine*, vol. 9, no. 6, pp. 796–800, 2003.



**Cihan Bilge Kayasandik** is an assistant professor in the Department of Computer Engineering at Istanbul Medipol University, where she has been a faculty member since 2019. She graduated with B.Sc. degree from the Department of Mathematics at Bilkent University in 2012. Then she completed her Ph.D. degree at the Department of Mathematics at University of Houston in 2017. She worked as a researcher in the same university for one more year. Her current research areas include medical image and data analysis.



# Geo-Location Spoofing on E-Scooters; Threat Analysis and Prevention Framework

Ahmet Saim Yilmaz, Haydar Cukurtepe and Emin Kugu

**Abstract**— Geo-location services are widely used by a large number of applications and devices. E-scooters are one of the recent prevalent devices that use these services. Geo-location services are highly vulnerable to spoofing attacks due to their open communication. Spoofing detection and prevention techniques have been researched for a long time. Most of the research has focused on smart devices such as smartphones, unmanned aerial vehicles, or other vehicles. Due to intrinsic application limitations, these threats need to be addressed on a per-environment basis. Geo-location spoofing also poses significant threats to e-scooters in terms of user security, costs, and reliability. In this study, we analyze location spoofing threats in an e-scooter-sharing environment and propose a spoofing prevention framework. The proposed framework monitors the e-scooter's location and tracks significant changes, and its service-based structure provides differentiated levels of awareness and capability.


**Index Terms**—E-Scooter, Geo-location Spoofing, GPS, Location Security, Wi-Fi Positioning

## I. INTRODUCTION


GEO-LOCATION SERVICES are widely used by a large number of applications (i.e. navigation, delivery systems, tracking systems, etc.) and devices such as smartphones, smart home devices, and unmanned aerial vehicles (UAV). Recently, electric scooters (e-scooters) have joined the family of geo-location-enabled devices, where e-scooters are made available for short-term rentals.

Geo-location services employ different methods to calculate relative positions, namely Global Navigation Satellite Systems (GNSS), Wi-Fi, and network-based geo-location services.


**AHMET SAIM YILMAZ** is with the Department of Computer Engineering University of Ted University, Ankara, Turkey, (e-mail: asaaim.yilmaz@tedu.edu.tr).

 <https://orcid.org/0000-0002-4279-684X>

**HAYDAR ÇUKURTEPE** is with the Department of Computer and Information Science of Valparaiso University, Valparaiso, USA, (e-mail: haydar.cukurtepe@valpo.edu)

 <https://orcid.org/0000-0002-4670-4877>

**EMIN KUĞU** is with the Department of Software Engineering University of Ted University, Ankara, Turkey, (e-mail: emin.kugu@tedu.edu.tr).

 <https://orcid.org/0000-0001-7829-8087>

Manuscript received Jan 9, 2023; accepted Sep 2, 2023.  
DOI: [10.17694/bajece.1231384](https://doi.org/10.17694/bajece.1231384)

GNSS is a term that refers to the International Multi-Constellation Satellite System, and one broad implementation of GNSS is Global Positioning (GPS) (these terms are used interchangeably). Receivers (i.e. e-scooter) continuously listen to the signals from multiple GPS satellites and calculate their position. When GPS signals are weakened or are even blocked by solid structures, a Wi-Fi positioning system (WPS) steps in to find the signals of nearby Wi-Fi hotspots and wireless access points, thus its own location. Devices also exchange Received Signal Strength (RSS) information to calculate how far away the Wi-Fi access point is from the device. RSS is also attenuated while passing through solid objects, such as walls. Geo-location services use trilateration to compute positions. The other geo-location services (s.a. Bluetooth low energy, network-based, RFID based geo-location) services are dependent on the provider and additional hardware. Thus, these types of services are not considered in this study. In a 3D space, the intersection of 3 spheres gives the point of interest. An e-scooter continuously listens to the GPS signals and measures distance to several Wi-Fi access points, then combines this information with a propagation model to determine its position.

A spoofing attack is a malicious activity in which a program is successfully identified as someone else by forging the actual data [1], or creating fake ones. Location spoofing is a serious threat to geo-location enabled devices. An adversary can make the victim appear in a place irrelevant to the original location. The primary reason for being a serious threat is the ease of method. This threat does not require physical interference or jailbreak-root. More advanced devices (e.g. Smartphones) use various technologies together for security, with predefined priority order [2], [3] or encrypted communication [4]. Relatively easier spoofing techniques pose significant risks on e-scooters. Neither these devices are equipped with encrypted communication capability nor they employ advanced techniques to prevent geo-location spoofing attacks.

In this paper, we analyze location spoofing threats in e-scooter sharing systems and propose a spoofing prevention framework. It takes predefined prevention steps by monitoring the network properties of devices and checking if any significant changes have been made. We analyze different spoofing threats and their applicability in the e-scooter sharing environment. Although geo-location spoofing attacks on smart devices (e.g. UAVs, smartphones) have been studied in detail, this distinctive e-scooter sharing system has not been studied in detail yet. In [5], location spoofing threats against e-scooters are mentioned as a subsection, but no detailed threat analysis or a detection framework is presented. In [6], authors

analyze the threats of location spoofing in e-scooter sharing environments by examining various scenarios and the principles of GPS and Wi-Fi location systems. Authors point out possible negative user experiences, compromised security, maintenance delays, and reputational damages to companies but no spoofing prevention mechanism is proposed.

Spoofing detection and prevention have been studied by many researchers, and various solutions have been proposed to disentangle the legitimate signals from spoofed ones. Jovanovic [7] detects spoofing by observing changes in the clock offset of the GNSS receiver. The authors in [8] use the data from the GNSS technologies to cross-check the data from the base station of the neighboring cell of the mobile cellular network. In [9], an algorithm is proposed for detecting Wi-Fi signal spoofing attacks using RSS. E-scooter systems are different from the other geo-location enabled devices in that they are always on-site and are not equipped with advanced computing power. Thus, e-scooter systems require their own solutions. While previous research has examined spoofing detection and prevention techniques for the devices such as smartphones, unmanned aerial vehicles or other vehicles, this study addresses the unique challenges of the e-scooter sharing environment. This focused analysis study on e-scooter sharing systems and the proposed prevention framework makes this study different from the ones in the literature.

In the rest of this study, we first explain the working principles of location services: GPS and WPS. Section 4 explains how geo-location spoofing attacks are carried out. In Section 5, location spoofing threats against e-scooter systems are presented and analyzed. Section 6 evaluates the applicability of these threats. Section 7 presents the proposed spoofing prevention framework, and Section 8 concludes the study.

## II. GLOBAL POSITIONING SYSTEM

Global Positioning System (GPS) is a service given by dedicated satellites, which provides location, velocity, and time synchronization. Devices receive GPS signals from satellites to calculate their location, speed, and elevation. The technique used by GPS is called trilateration.

The GPS has 3 parts: satellites, ground stations, and receivers. Ground stations make sure the satellite is in the correct position. Receivers (e.g. e-scooters) constantly listen to the signals from satellites and calculate the position.

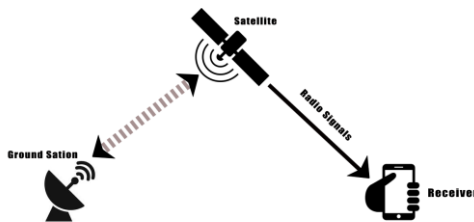


Fig. 1 Basic Concept Representation of GPS.

In Fig. 1, ground stations serve as command and control services for satellites to make sure that they are in the correct position and the receiver collects the signals coming from satellites. GPS satellites broadcast radio signals to enable GPS receivers to calculate position according to the generated signals. GPS receivers do not work well indoors or in crowded environments such as city centers, due to weakened or blocked signals. This situation makes receivers extremely vulnerable to spoofing attacks.

There are different types of GPS services for different purposes. Publicly available GPS Service provides 4.9 meters of accuracy with low cost but it does not provide high accuracy or encrypted signals. Military GPS services and Galileo provide encrypted satellite signals, but e-scooters are not equipped to receive these signals.

### A. How does it work

GPS satellites use the L-band to transmit their signal, and the time information it uses is coordinated with the atomic clock located inside the satellite. GPS' signal structure is publicly accessible, making attacks easier to execute. The signal generated by a GPS satellite travels by line of sight. They are able to travel through water and clouds, but not solid objects such as walls, tunnels, or mountains. A GPS signal consists 3 different types of information;

- Pseudorandom code; this code holds information about the GPS itself, it tells from which satellite the signal is generated. It is the Identification Code of the GPS.
- Ephemeris data; Holds information about position and speed of the GPS. Also holds information about the satellite's health situation.
- Almanac data; Gives orbital information about GPS and tells where GPS would be located at any time of the day.

At the receiver, the user receives the GPS signal using an antenna with an RF front-end GPS receiver. After the signal is received, it goes through a number of steps such as filtering, amplification, and digitizing. Then, the received signal is converted into a baseband signal. This trimmed GPS signal is then processed to make a position computation. Its pseudorandom code is used to get information for all satellites in view, and the position is calculated based on Ephemeris and Almanac data. To calculate its position, the receiver requires 3 basic pieces of information: The location of the satellite, receiving time of each signal, transmission time of each satellite signal, this information is included in the Ephemeris and Almanac data of the GPS signal. The receiver requires 4 satellite signals to ensure accuracy.

Time of arrival (TOA) is the absolute time that a signal is received by the receiver. Position information is calculated by using differences in TOAs which is called Time Difference of Arrival (TDOA). The mathematical form is:

$$\hat{d}_{u_i} = c(\tau_{t,u_i} - \tau_{r,u_i} + \Delta n) = \sqrt{(x_u - x)^2 + (y_u - y)^2 + (z_u - z)^2} \quad (1)$$

- Where, n is the satellite id (e.g. if there are 4 satellites, it goes as 1,2,3,4)
- Variable c is the speed of light which is 3x10<sup>8</sup> m/s
- $t_{t,n}$  is the time GPS satellites transmit their signal. e.g. the signal time transmitted from GPS satellite number 1 is  $t_{t,1}$
- $t_{r,1}$  is the time the GPS receiver segment receives the signal coming from a specific satellite
- $\Delta n$  is receiver clock error
- $x, y, z$  are the respective coordinates of the satellite.

Since the equation has 3 unknown variables and an additional time variable, the receiver requires at least 4 GPS signals to solve the equation.

### III. WI-FI POSITIONING SYSTEM

Wi-Fi is a set of network protocols (IEEE 802.11 family), it allows digital devices to exchange data using radio waves. There are estimated to be around 14 billion devices having Wi-Fi capabilities [10].

The Wi-Fi positioning system receives location data from nearby Wi-Fi hotspots and wireless access points to discover the location of a device. It is especially useful in calculating locations when a device is located indoors, or at locations where the satellite signals are weak. To calculate the position, the device needs to listen to signals coming from Wi-Fi access points and analyze them to decide of the position.

The exchanged data between devices includes Received Signal Strength (RSS). It provides information about how far the Wi-Fi access point is relative to the device. While strong RSS shows the Wi-Fi access point is close to the device, weak RSS shows that device is further away. RSS weakens while passing through solid objects such as walls and windows. So, having weak RSS sometimes may not mean that the Wi-Fi access point is further away.

#### A. How it works

The strength of RSS can be modeled as shown in Equation 2;

$$P(d) = P_0 - 10n \log_{10} \left( \frac{d}{d_0} \right) \quad (2)$$

In this equation,  $P(d)$  is the received power in dB, where, d is the distance and  $P_0$  is the received power in dB at a short reference distance  $d_0$ . Fingerprinting is an indoor positioning technology to determine a user's position. Fingerprinting approach uses RSS to make a position computation.

Position computation in Wi-Fi positioning systems is based on measuring distance to several Wi-Fi access points then combining this information with a propagation model to determine the position of the device. Trilateration is being

used to make a computation. The disadvantage of this technique is that it provides an accuracy around 10 meters.

The trilateration algorithm corresponds to sphere centers. In a 3D world, the intersection of 3 spheres gives our point of interest. In order to calculate it we make use of the following equation:

$$r^2 = (x - x_a)^2 + (y - y_a)^2 + (z - z_a)^2 \quad (3)$$

Let one sphere as an origin point which is A1 in this example and simplify the equations. After radiuses of the spheres become  $r_1$ ,  $r_2$  and  $r_3$  the simplified equations become:

$$r_1^2 = x^2 + y^2 + z^2 \quad (4)$$

$$r_2^2 = (x - x_2)^2 + y^2 + z^2 \quad (5)$$

$$r_3^2 = (x - x_3)^2 + (y - y_3)^2 + z^2 \quad (6)$$

In short, trilateration is used to obtain a 2D point and make a position computation with a low accuracy.

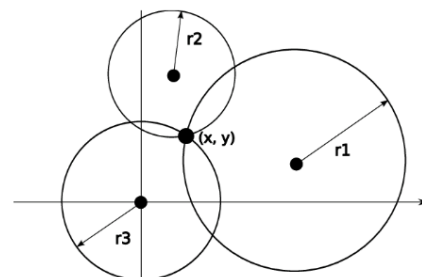


Fig. 2 Trilateration.

As shown in Fig. 2, the circle with radius r3 taken as origin point and trilateration is used accordingly. Intersection point of 3 circles represents the location.

### IV. GEO-LOCATION SPOOFING

#### A. GPS Signal Spoofing

In GPS spoofing, the malicious user uses a radio transmitter to send counterfeit GPS signals to a receiver antenna to interfere with legitimate GPS satellite signals. Most of the location systems are designed to use the strongest GPS signals, and the strong fake signals are able to override the weaker legitimate signals. In [11], authors launched GPS spoofing attacks on various devices and successfully falsified the location of the device. In [12], the authors successfully execute a spoofing attack to the navigation system using a satellite simulator and manage to navigate a victim to locations 1 kilometer away from the original destination. Aside from the navigation system, authors have also successfully spoofed devices (i.e. iPhone 6, Samsung S7) using Hack-RF based spoofers.

The GPS spoofing techniques can be classified into two:

1) *Spoofing with GPS Signal Simulator*: In this type of attack, Software-Defined Radio (SDR) is needed to generate baseband signals and then transmit them to the receiver using an antenna. Time synchronization is not necessary with the real GPS signals. The important thing is that the generated GPS Signal strength must be stronger than authentic signal. The attack model using GPS Signal Simulators is as follows: First step is to choose an SDR platform to generate baseband signals. There are couple of options (i.e. BladerRF, RTL-SDR, LimeSDR, SDRplay, USRP, YARD stick etc.). HackRF is an open source SDR platform operating on frequency bands between 10MHz and 6GHz. After selecting the proper SDR, signals are generated, then it comes to a point where the generated signals are transmitted using an antenna.

2) *Spoofing with Receiver-Based Spoofers*: Receiver Based Spoofers synchronizes the signals with target GPS signals. It extracts time, velocity and position information in order to have a 3D vector of the transmit antenna. Unlike GPS Signal Simulators, the generated signal is authentic and synchronized to the real GPS signals.

When the adversary has complete knowledge of the location of the GPS receiver, the adversary can potentially place multiple RF based antennas to generate GPS signals. Let's call authentic satellite  $S_j$  as the satellite that an adversary wants to impersonate and use  $S_j^a$  to represent the satellite that is emulated. Let  $S^a$  be the entire set of emulated satellites making each fake GPS signal coming from satellite  $S_j^a$ . Now the adversary needs to assign values to these variables in order to execute a GPS spoofing attack.

The next step is to summarize the adversary's variables for each emulated satellite  $S_m^a$  in order to generate fake GPS signals to falsify the GPS receiver's position. Let  $\hat{l}_m^a$  be the falsified location and  $\hat{t}_m^a$  is falsified time. First the adversary needs to announce its orbital information and claim to be at  $\hat{l}_m^a$  at  $\hat{t}_m^a$  second. Let  $l_i^a$  be the true location and  $t_{i,m}^a$  be the true time on the instant of reception, the adversary sets the true location for the antenna as  $l_m^a$  and true time as  $t_m^a$  to represent when the signal is sent.  $t_m^a$  is associated with  $\hat{t}_m^a$  by clock delay making the equation  $\hat{t}_m^a = t_m^a + \delta_m^a$ . Finally, local timestamp becomes  $\hat{t}_{i,m}^a = t_{i,m}^a + \delta_{i,m}^a$ . The claimed time of arrival equation of the adversary can be derived as  $c(\hat{t}_{i,m}^a - t_m^a) = |l_i^a - l_m^a|$ .

This represents the adversary's intended time that is generated from Receiver-Based Spoofer.  $l_i^a$  is calculated location of the receiver when the message is received from  $S_m^a$ . The same equation for the authenticated GPS signal can be shown as  $c(\hat{t}_{i,m}^a - \hat{t}_m^a) = |\hat{l}_i^a - \hat{l}_m^a|$ .

The left-hand side of this equation represents the range between satellite and GPS receiver which is called Pseudorange.

The adversary connects with its fake satellite data with the intended locations and time offsets for which a receiver would be spoofed to solve for the variable. All variables can be related by taking the difference of 2 equations.

$$|l_i^r - l_m^a| + c(\hat{\delta}_{i,m}^r - \delta_m^a) = |\hat{l}_i^r - \hat{l}_m^a| \quad (7)$$

The adversary's goal is to solve for the variable sets that satisfy the following equation. Once all variables are solved, the adversary can successfully falsify the GPS receiver's location in a synchronized way using receiver based spoofer.

### B. Wi-Fi Signal Spoofing

Fabricating APs from another location, the WPS on a smartphone can be fooled into thinking that the device is currently in that specified location. As discussed in the earlier sections, WPS infers the location based on Basic Service Set Identifiers (BSSIDs) or MAC addresses of nearby APs. It is possible to generate BSSIDs and falsify the Wi-Fi positioning system. Those generated APs are called fake Wi-Fi access points which are set up by configuring a wireless card to act as an access point [13].

## V. PROBLEM STATEMENT AND ANALYSIS

E-scooter systems let users unlock and book through mobile applications. Locations of the nearby e-scooters are shown in the app to assist users find available e-scooters. Companies running these systems (i.e Lime, Voi scooters, Link and Marti), plan their maintenance and logistic activities according to the location of the devices. Safety regulations are applied for user safety, and precautions are applied against theft or sabotage risks. For instance, an e-scooter company may restrict the scooter's speed on specific roads or prohibit riding a scooter on sidewalks to ensure pedestrian safety. In this paper, we based our analysis on common terms of use of major companies to meet on a common ground. Geofencing technology is being widely used by e-scooter sharing companies.

Geo-fencing is a location-based service for setting up a virtual boundary where the user is allowed to ride. When it is crossed, actions are taken by the companies such as warning the user or stopping the e-scooter for pedestrian safety. Adversaries may attack the e-scooter environment to harm user experience and e-scooter companies.

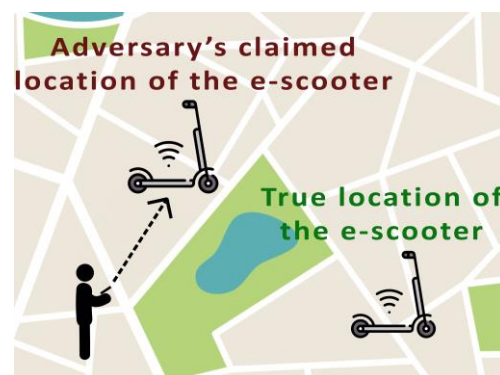


Fig. 3 Exploiting the Pre-Book Feature & Misleading the User.

### A. Threats

1) *Exploiting the pre-book feature & misleading the user*: E-scooter companies allow users to pre-book e-scooters and



unlock them, preventing others from booking the same e-scooter. Each minute the e-scooter is pre-booked, the user is charged respectively. An adversary may spoof the location of the e-scooter and mislead the user. Both time and money would be wasted until the user figures out the e-scooter is not in the supposed location.

As shown in Fig. 3, the adversary exploits the pre-book feature using location spoofing procedures in section 4 to fool the user and rent an e-scooter irrelevant to position showing in the app.

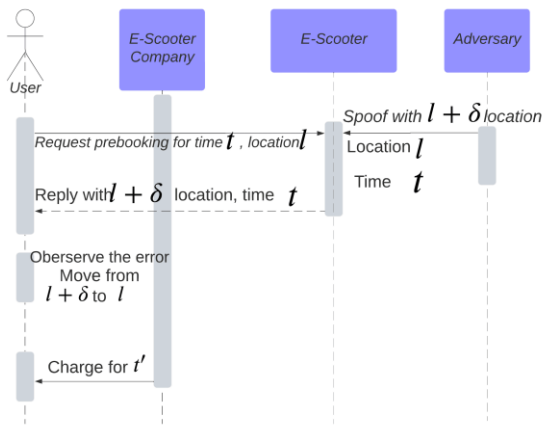


Fig. 4 Exploiting the Pre-Book Feature-Sequence.

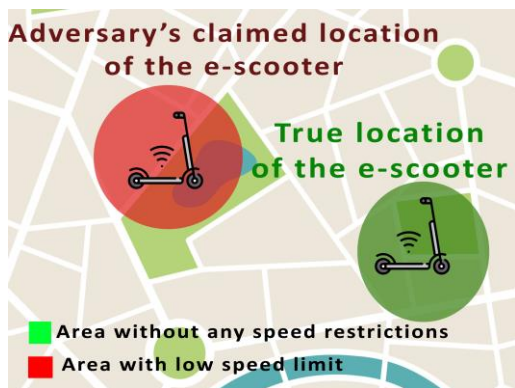


Fig. 5 Prevent User from Ending the Ride.

In Fig. 5, the adversary uses spoofing location methods to falsify e-scooter’s location and show up in an area where the user is not allowed to ride.

In Fig. 4, let  $l$  be the true location of the e-scooter and  $t$  be the real time. Adversary uses location spoofing procedures in section 4 to falsify the location of the e-scooter and show up in another location which is real location  $l$  plus the adversary’s spoofed location  $S$ . This would mislead the user to go the wrong direction. The time  $t'$  and money would be wasted until the user figures out the e-scooter is not located in the supposed location which shows up in the app.

2) *Prevent User from Ending the Ride:* There is a virtual boundary where the user is able to ride the e-scooter. When the user wants to end the ride out of those allowed areas, a

message is prompted by the app telling “return the area to end the ride”. Adversaries may spoof the location of the e-scooter, making it show up in the unallowed area and prevent the user from ending the ride.

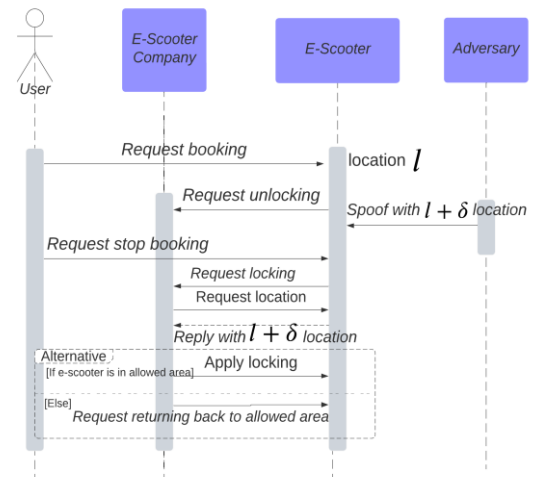


Fig. 6 Prevent User from Ending the Ride-Sequence.

In Fig. 6, user wants to end the ride by locking e-scooter at location  $l$ . Adversary uses location spoofing to make the e-scooter show at  $l + S$  which is not an allowed zone for users to end the ride. The user won’t be able to end the ride until the e-scooter is at a position with allowed boundaries.

3) *Limiting the Scooter’s Top Speed on the Ride:* There are specific places where e-scooter companies limit the speed of e-scooter to protect pedestrians and users. Adversary may falsify the location so that the device is supposedly in the areas with speed limits constantly limiting the speed of the e-scooter vice versa.



Fig. 7 Limiting the Scooter’s Top Speed on the Ride.

In Fig. 7, the adversary uses spoofing location methods to falsify e-scooter’s location and show up in an area with low speed rules making it slow down.

In Fig. 8, user rides e-scooter at location  $l$ . Adversary uses location spoofing to make the e-scooter show at  $l + S$  which is

a low speed zone. By this way the adversary makes the e-scooter slow down and harm user experience.

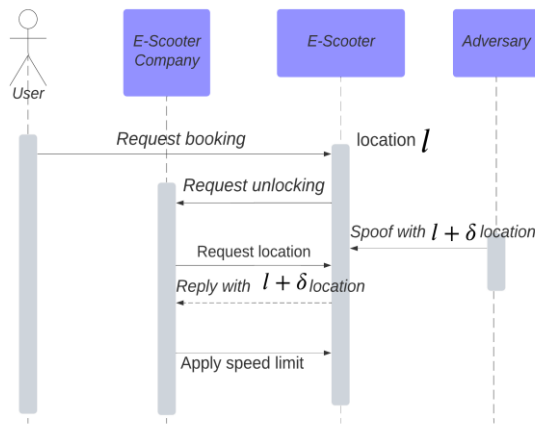


Fig. 8 Limiting the Scooter’s Top Speed on the Ride-Sequence.

4) *Preventing companies from making e-scooters maintenance*: E-scooter companies need to constantly charge the scooters and make maintenance. A typical e-scooter battery lasts around hours of driving. This requires companies charging them on a daily basis. Adversaries could potentially use both GPS and WPS spoofing depending on the situation of the e-scooters and prevent companies seeing the e-scooters at a falsified position.

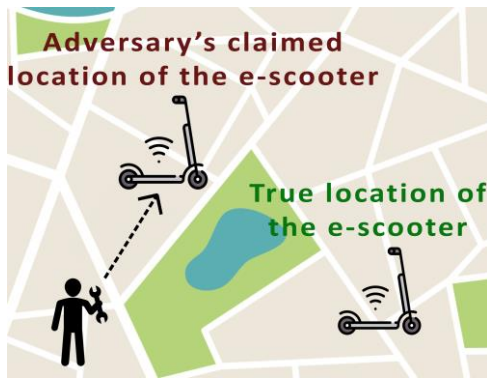


Fig. 9 Preventing Companies from Making E-scooters Maintenance.

In Fig. 9, the adversary uses location spoofing to prevent maintenance crew to find the e-scooter so that they could not be charged nor taken to service.

In Fig. 10, maintenance crew checks the e-scooter’s position and goes location  $l + \delta$  to repair and charge the device. But the location crew is going is the spoofed location of the adversary, the true location of e-scooter is location  $l$ .

5) *Theft or sabotage on e-scooters*: Adversary may spoof location of the e-scooters to show them up in irrelevant places. So, the adversary could physically steal or sabotage the e-scooter without being noticed.

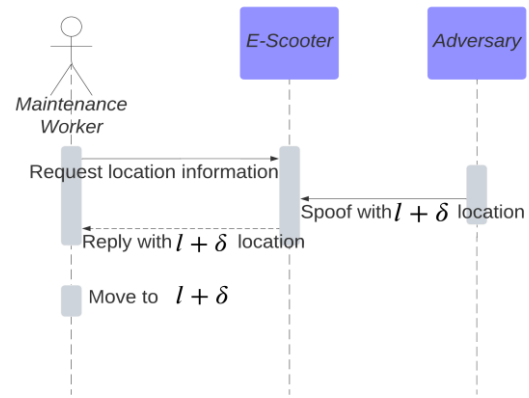


Fig. 10 Preventing Companies from Making E-scooters Maintenance-Sequence.

VI. APPLICABILITY OF THREATS

Adversaries can use different location spoofing techniques for different scenarios depending on the circumstances. For instance, in WPS spoofing techniques fake access points cannot be dynamically created in the direction of movement of the device hence this kind of spoofing technique will fail against non-stationary e-scooters. Applicability of spoofing techniques could be analyzed based on 2 scenarios which are stationary and non-stationary e-scooters.

A. *Spoofing the location of a stationary e-scooter*

When the goal is to harm the user and the company by falsifying the location information, as in threats 1,4 and 5, adversaries could use both WPS and GPS location spoofing techniques.

- Step 1 (Choosing the ideal target location): Most of the time geo-location capable devices use GPS based location calculation prior to Wi-Fi Positioning System. Having a healthy signal coming from various GPS signals makes Wi-Fi spoofing almost impossible. In order to have a successful spoofing attack, the target location must be somewhere surrounded by buildings having low GPS signals.
- Step 2 (Collecting falsified position information): Then, the adversary needs to generate Wi-Fi Hotspot Tags. To do that adversary may use Wigle.net or Skyhook to collect Wi-Fi Hotspot tags in use around the world.
- Step 3 (Broadcasting Wi-Fi Access Points): After fake BSSIDs are generated having target position, beacons are generated and broadcasted. The SkyLift project can be used for this purpose. SkyLift is an open source tool for broadcasting Wi-Fi beacon frames with a low cost Wi-Fi microcontroller.

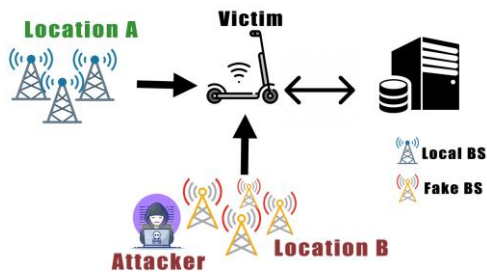


Fig. 11 Wi-Fi Spoofing Attack.

In Fig. 11, an adversary generates fake Wi-Fi access points to falsify the victim’s position making it appear at "Location B" instead of "Location A".

To spoof GPS, an adversary needs to transmit GPS signals via antenna to falsify the target e-scooter’s position. GPS Signal Simulator or Receiver Based Spoofers can be used as explained in section 4.

There are various GPS Signal Simulators available on the market, such as Hack-RF, bladeRF, NESDR Nano 2+, USRP or NESDR SMARt HF Bundle. They all serve the same purpose, with different capabilities. For instance, USRP has the greatest bandwidth and speed amongst all. NESDR Nano 2+ is the smallest SDR which is even smaller than a usb flash drive, if the adversary’s priority is portability then it is the right choice. Hack-RF is an open source and cost friendly SDR platform. Here is a sample command line argument to transmit generated signals using HackRF:

```
hackrf_transfer -t gpssim.bin -f 1575420000 -s 2600000 -a 1 -x 0 -R
```

To execute the command successfully, -t file name, -f frequency, -s sample rate and -a amp enable (1 or 0) needs to be given as an input. After the command, HackRF transmits those signals using an antenna. The target device calculates its position based on the falsified GPS signals and GPS position spoofing occurs.

When an adversary wants to spoof an e-scooter that is checking time synchronization of GPS signals as a security mechanism then Receiver Based GPS Spoofer could be used. Adversary needs to place the spoofer close to the e-scooter where it could gather GPS signals and puzzle out the time information of authentic GPS signals. After this step, time synchronized GPS signals could be transmitted to the e-scooter.

In addition to spoofing, jamming attacks also pose a significant threat against WPS and GPS. Jammer could possibly block Wi-Fi enabled devices from successfully getting signals to calculate the position. On the other hand, it can also hide or change GPS signals before it is being received by the device. In [14], authors successfully execute a GPS jamming attack using Software Defined Radio and prevent devices from receiving GPS signals.

**B. Spoofing the location of a non-stationary e-scooter**

When the adversary’s goal is to falsify the location of a nonstationary e-scooter as in threats 2 and 3, the GPS location spoofing techniques can be used. In WPS spoofing technique, fake access points cannot be dynamically created in the direction of movement of the device hence this kind of

spoofing technique will fail. In addition, if an e-scooter starts to get healthy GPS signals while moving from one location to another, WPS spoofing would not work properly. On the other hand, GPS spoofing techniques could be used against a non-stationary e-scooter.

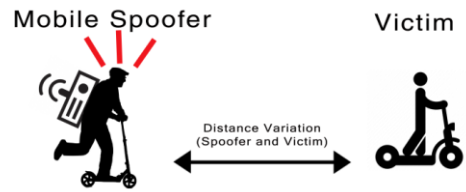


Fig. 12 Mobile Wi-Fi spoofing attack.

Both Receiver Based Spoofer and GPS Signal Simulator should be attached to a mobile vehicle and follow the e-scooter. In Fig. 12, the vehicle carries a spoofer and location spoofs the target device. For the sake of portability, adversaries may use NESDR Nano 2+ as an SDR.

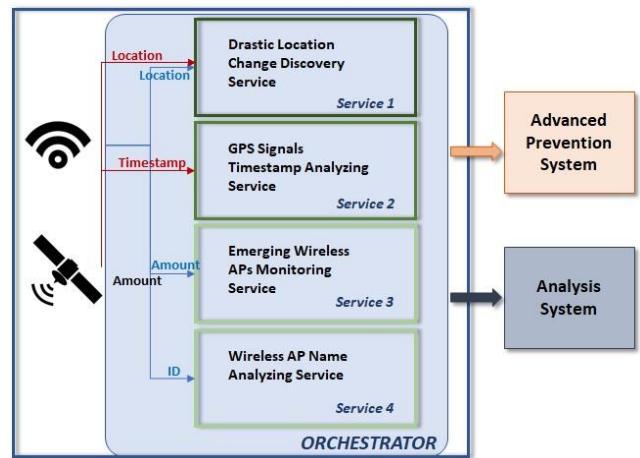


Fig. 13 Conceptual Model of SPFES.

**VII. SPOOFING PREVENTION FRAMEWORK FOR E-SCOOTERS (SPFES)**

E-scooter geo-location spoofing poses significant threats to user security, regulations and e-scooter sharing companies’ costs and prestige. Spoofing Prevention Framework for E-Scooters (SPFES) is a framework designed to detect the location of spoofing attacks on e-scooters, thus enabling taking preventive measures. A properly configured framework can prevent e-scooters from misleading both users and companies.

The proposed framework has a service-based architecture. It relies on GPS and WPS receivers in e-scooters. E-scooter having both capabilities are able to employ more services. The framework collects and processes rich information in GPS signals.

The proposed location spoofing detection and prevention framework provides four different services. (i) Discovering the drastic differences in the location, (ii) checking the new number of Wireless Access Points that have appeared, (iii) analyzing GPS signals timestamps, and (iv) analyzing the



names of the Wireless Access Points. The working logic of the services is explained below, and the conceptual model of the SPFES framework is given at 13. Each service can be configured to work as a standalone service or a part of a multiple-service combination. Advanced prevention system module implements recent prevention techniques at the edge. Using edge computing, the detected spoofing attacks are prevented at the vicinity without latency.

#### A. Service 1: Drastic Location Change Discovery Service

The first service compares perceived location shift and error threshold to detect a location spoofing between two consecutive timestamps. Perceived location shift is calculated by using the latitude and longitude data taken from both GPS and Wi-Fi sensors. Most of the e-scooters of the e-scooter sharing companies have a very limited top speed which is around 20 kilometers per hour due to safety concerns [15]. Hence, they are not able to make drastic-location changes. Service 1 focuses on drastic location change in a limited time span. The applied speed limitation and e-scooters' physical speed limits are the rationale of this service. It checks whether or not there is a drastic difference between two consecutive timestamps by comparing error threshold and perceived location shift. It is considered to be a spoofing attack, if this service notices such a shift.

#### B. Service 2: GPS Signals Timestamp Analyzing Service

As stated in section 4, GPS Signal Simulator attack could be used to spoof GPS signals. It generates strong GPS signals to override the genuine ones. This type of attack is easy and effective but does not synchronizes the timestamp. Because it is easy and low-cost, usually spoofing attacks take place by using GPS Signal Simulator. Service 2 focuses on the weakness of this type of attack by analyzing GPS signals timestamps. If the timestamps are not coherent with the latest stored timestamps, then it is considered to be a spoofing attack.

This service would fail if the adversary uses receiver-based spoofers, which synchronize the signals with target GPS signals. Although this type of attack is complex and expensive to execute, and it is not expected to be used too often, the SPFES relies on more than one service to overcome this threat.

#### C. Service 3: Emerging Wireless APs Monitoring Service

Service 3 monitors emerging Wireless Access Points. In WPS spoofing attack, the adversary creates fake Wi-Fi Access Points to falsify the Wi-Fi Positioning system. While doing so, the adversary needs to create necessary fake access points to successfully spoof the location. The number of signals coming from fake access points would be strong enough to have an effect on the trilateration method. This situation requires a significant increase in the number of Wi-Fi Access Points. Service 3 checks the newly emerged access points to determinate location spoofing attacks.

#### D. Service 4: Wireless AP Name Analyzing Service

Service 4 focuses on analyzing the names of the Wireless Access Points. The fake access points created by spoofing

tools (i.e. "SkyLift") has similar names by default. Most of the time, attackers do not pay attention to changing names of the access points. Too much similarity between the names of surrounding access points is considered to be a possible spoofing attack. Service 4 analyzes the names of surrounding Wi-Fi Access Points and detects location spoofing attacks.

These services can be configured to run per se or collectively in different combinations, according to the requirements of the provider. E-scooter systems are different from other smart systems (i.e. smartphones, automobiles) in that they do not have immense computation capability, and implementing more complex detection systems are not always cost efficient in terms of running hardware and computation.

#### E. Orchestrator

The Orchestrator is the entity that runs on top of all services. The Orchestrator has multiple tasks: (i) It enables the dynamical configuration of effective services. The service provider can access devices, on-premise or online, to configure the services properly. (ii) It coordinates the smooth running of services together. It checks the running services, ensures the elimination of deadlocks, and keeps a log of the service activities. (iii) When the services detect any malicious activity, the orchestrator either communicates with the edge computing device for advanced prevention, or takes predefined prevention steps to keep services running. Necessary computation is done at the edge and rapid decision and reaction are taken without latency. Advanced prevention capability using edge computing is not planned as a mandatory service.

#### F. Limitations & Discussion

Since SPFES is a service-based framework, it is able to accommodate state of the art spoofing detection techniques. It is also able to cooperate with advanced prevention systems, or advanced analyzing systems. SPFES enhances the security and reliability of the e-scooter system. On the other hand, this framework brings power consumption cost to the e-scooter system.

SPFES is devised to prevent the advanced spoofing attacks, but it also employs basic defense steps such as discarding the possible spoofing threat signals. It is a simple but effective solution to implement. These predefined steps can be adjusted to implement a certain amount of time and then the advanced system can be called. These adjustments affect the performance of the SPFES.

Another limitation of the framework is that SPFES assumes that the legitimate geo-location signals are available prior to location spoofing attacks. SPFES will not detect spoofing attacks if the spoofing attack starts from the beginning of the journey and keeps going along.

## VIII. CONCLUSION

Geo-location services become more popular because of the expanding scale of location-based applications. Global Positioning System and Wi-Fi Positioning System are most commonly used techniques for positioning. E-Scooters have joined geo-location enabled devices with the growing industry of e-scooter sharing. An adversary could use location spoofing



techniques to prevent users and maintenance crew from finding an e-scooter, lower the quality of the ride, prevent users from ending the ride or limit the e-scooters speed, and show them in unallowed areas. There could be serious harm done to E-Scooter Companies and users in terms of time, money and prestige. In this study, we explained working principles of positioning systems, analyzed the location spoofing threats on e-scooter sharing environments, and demonstrated unique implementations of different spoofing attacks. We also proposed a spoofing attack prevention framework for e-scooters (SPFES) that is able to detect any malicious activity via monitoring location information and tracking significant changes around. The proposed service-based framework is able to provide different levels of awareness, and enabled capabilities.

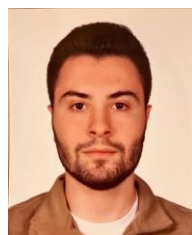
Implementation of the SPFES framework will be the next phase of our studies. In future, SPFES implementation can be improved using artificial intelligence (AI) and machine learning algorithms. AI enabled edge computing will also help to detect anomalies with better accuracy.

## REFERENCES

- [1] Jindal, K.; Dalal, S.; Sharma, K. K. "Analyzing Spoofing Attacks in Wireless Networks". 2014 Fourth International Conference on Advanced Computing Communication Technologies: (February 2014), 398–402. doi:10.1109/ACCT.2014.46. ISBN 978-1-4799-4910- 6. S2CID 15611849.
- [2] Getting the User's Location, Apple Developer Documentation. [Online]. [Accessed: 10-Dec-2022]. Available: <https://developer.apple.com/documentation/corelocation/gettingtheuserslocation>.
- [3] Geolocation API, Google. [Online]. [Accessed: 10-Dec-2022]. Available: <https://developers.google.com/maps/documentation/geolocation/overview>
- [4] EUROPEAN GNSS (GALILEO) OPEN SERVICE Issue 2.0, January 2021 NAVIGATION SOLUTIONS POWERED BY EUROPE SIGNALIN-SPACE INTERFACE CONTROL DOCUMENT
- [5] N. Vinayaga-Sureshkanth, R. Wijewickrama, A. Maiti, and M. Jadhwal, "Security and privacy challenges in upcoming Intelligent Urban Micromobility Transportation Systems," Proceedings of the Second ACM Workshop on Automotive and Aerial Vehicle Security, (2020). doi:10.1145/3375706.3380559
- [6] A. S. Yilmaz, H. Çukurtepe and E. Kuğu, "Analysis of Location Spoofing Threats on E-Scooter Sharing," 2022 30th Signal Processing and Communications Applications Conference (SIU), Safranbolu, Turkey, 2022, pp. 1-4, doi: 10.1109/SIU55565.2022.9864946.
- [7] Jovanovic, A., et al.: Multi-test detection and protection algorithm against spoofing attacks on GNSS receivers. 2014 IEEE/ION Position, Location and Navigation Symposium (PLANS), (May 2014), pp. 43–49. doi:10.1109/PLANS.2014.6851501
- [8] F. Formaggio, S. Ceccato, F. Basana, N. Laurenti, S. Tomasin, GNSS Spoofing Detection Techniques by Cellular Network Cross-check in Smartphones, in: Proceedings of the 32nd International Technical Meeting of the Satellite Division of the Institute of Navigation, ION GNSS+ 2019. doi:10.33012/2019.17076.
- [9] Khan, F.; Al-Atawi, A.A.; Alomari, A.; Alsirhani, A.; Alshahrani, M.M.; Khan, J.; Lee, Y. Development of a Model for Spoofing Attacks in Internet of Things. *Mathematics* (2022), 10, 3686. <https://doi.org/10.3390/math10193686>
- [10] Tzeng C.L. Global Wi-Fi Enabled Devices Shipment Forecast, 2020–2024. Market Intelligence Consulting Institute (MIC); Taiwan, China: (2020)
- [11] D. -K. Lee et al., "Detection of GNSS Spoofing using NMEA Messages," 2020 European Navigation Conference (ENC), (2020), pp. 1-10, doi: 10.23919/ENC48637.2020.9317470.
- [12] Kexiong Curtis Zeng, Yuanhao Shu, Shinan Liu, Yanzhi Dou, and Yaling Yang. A Practical GPS Location Spoofing Attack in Road Navigation Scenario. In Proceedings of the 18th International Workshop on Mobile Computing Systems and Applications (HotMobile '17). Association for Computing Machinery, New York, NY, USA, (2017) 85–90. <https://doi.org/10.1145/3032970.3032983>

- [13] Celestin M, Achara JP, Cunche M, "Short: Device-to-Identity Linking Attack Using Targeted Wi-Fi Geolocation Spoofing," Acm Conference on Security Privacy in Wireless Mobile Networks, (2015). doi:10.1145/2766498.2766521
- [14] R. V. Karpe and S. Kulkarni, "Software defined radio based Global Positioning System jamming and spoofing for vulnerability analysis," 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), (2020). doi:10.1109/ICESC48915.2020.9155565
- [15] Ma, Q.; Yang, H.; Mayhue, A.; Sun, Y.; Huang, Z.; Ma, Y. E-Scooter Safety: The Riding Risk Analysis Based on Mobile Sensing Data. *Accid. Anal. Prev.* (2021), 151, 105954. <https://doi.org/10.1016/j.aap.2020.105954>
- [16] Yu, D. Y., Ranganathan, A., Locher, T., Capkun, S., Basin, D. Short paper: Detection of GPS spoofing attacks in power grids. *WiSec 2014 - Proceedings of the 7th ACM Conference on Security and Privacy in Wireless and Mobile Networks*, (2014), 99–104. <https://doi.org/10.1145/2627393.2627398>
- [17] Wang, K. Time and Position Spoofing with Open Source Projects. (2015).
- [18] Meng, L., Yang, L., Yang, W., amp; Zhang, L. A survey of GNSS spoofing and anti-spoofing technology. *Remote Sensing*, 14(19), 4826, (2022). <https://doi.org/10.3390/rs14194826>
- [19] Huang, J., Lo Presti, L., Motella, B., amp; Pini, M. GNSS spoofing detection: Theoretical Analysis and performance of the ratio test metric in open sky. *ICT Express*, 2(1), (2016), 37–40. <https://doi.org/10.1016/j.ict.2016.02.006>.
- [20] Ye, A., Li, Q., Zhang, Q., Cheng, B. Detection of Spoofing Attacks in WLAN-Based Positioning Systems Using WiFi Hotspot Tags. *IEEE Access*, (2020), 8, 39768–39780. <https://doi.org/10.1109/ACCESS.2020.2976189>

## BIOGRAPHIES



**AHMET SAIM YILMAZ** received the B.S. degree from Ted University, Ankara, Turkey, in 2022. He is currently a researcher in TÜBİTAK-MAM working on an Energy Management System in Turkish Power Grid called as Load Dispatcher Information System (YTBS).



missions. Currently, he is Asst. Professor at Valparaiso University/IN, USA.

**HAYDAR ÇUKURTEPE** received the B.S. degree from Turkish Military Academy, master's degree from Air Force Academy in Software Engineering, and PhD from Istanbul Technical University, Computer Engineering Program in 2014. He served as Information System Staff in the Turkish Armed Forces, and NATO



He served as Information System Staff in the Turkish Armed Forces, and NATO missions. Currently, he is Asst. Professor at Valparaiso University/IN, USA.

**EMIN KUĞU** received the B.S. degree from Istanbul University Computer Science Engineering Department, his master's degree from Air Force Academy Software Engineering and his doctoral degree from Old Dominion University Electrical and Computer Engineering program. He worked as a part-time lecturer at different universities. He worked as a project manager in the Air Force Command Information System (HvBS) project, and took part in the software development processes of international defense projects.

# A Modular Multilevel Converter-Based Pulsed Electric Field Generator Design for Electroporation Applications

Ovul Eski, Kemal Sahin and Sevilay Cetin


**Abstract**—In this work, a modular multilevel converter-based pulsed electric field generator design is presented. In the proposed design, operation frequency is selected higher level compared to the similar works to provide new possibilities for better and healthy life. The pulse generation (PG) and the adapted modular multilevel converter (MMC) topology that is to be covered are introduced including the operating principles and very basic analysis. Finally, performance of the proposed MMC with four half bridge submodules is tested on a prototype built in laboratory. A bidirectional voltage is produced with 80 V amplitude pulsating at 100 kHz operation frequency.

**Index Terms**— Electroporation (EP), Modular multilevel converter (MMC), Pulsed Electric Fields (PEFs), pulse generation (PG).


## I. INTRODUCTION

FROM ELECTRIC fish to Lichtenberg figures the concept of electricity and life sciences had a certain space in the lives of humanity throughout history yet the knowledge of electricity and its interaction with tissues and cells was rather limited [1] [2]. The use of different kinds of electric fish, for instance, electric eel, catfish, and electric rays as therapeutic agents were reported in many historical records from different parts of the world from ancient Greece or Ancient Egypt to South American Indians [2]. For facial palsy and ptosis, a traditional Chinese prescription is stated in [2] suggesting an electric catfish tail is applied on the spot. According to [2], a torpedo, which is most probably prescribed by Hippocrates to Gallen, can deliver shocks of 45 volts of electricity whereas an electric eel from Amazon, the application of which is reported to treat gout by native South Americans by early explorers,


OVUL ESKI, is with Department of Biomedical Engineering University of Pamukkale, Denizli, Turkey, (e-mail: [ovuleski@gmail.com](mailto:ovuleski@gmail.com)).

 <https://orcid.org/0000-0002-4327-3100>

KEMAL SAHIN, is with Department of Biomedical Engineering University of Pamukkale, Denizli, Turkey, (e-mail: [sahinkemal.tech@gmail.com](mailto:sahinkemal.tech@gmail.com)).

 <https://orcid.org/0000-0002-8079-9224>

CETIN SEVILAY, is with Department of Biomedical Engineering University of Pamukkale, Denizli, Turkey, (e-mail: [bajece@hotmail.com](mailto:bajece@hotmail.com)).

 <https://orcid.org/0000-0002-9747-4821>

Manuscript received May 24, 2023; accepted Aug 21, 2023.  
DOI: [10.17694/bajece.1301685](https://doi.org/10.17694/bajece.1301685)

can deliver up to 600 volts.

However, during those long centuries and broad application all around the world throughout a vast number of cultures from Islamic tradition to Roman medicine in the western part and China to South America to Eastern, our understanding of the concept is almost metaphysical and far from today's theories and thereby the control of the electricity is out of scope [1] [2] [3].

In the mid-18<sup>th</sup> century, the invention of the Leyden jar by Musschenbroek can be considered a turning point from this stagnant state to a fast transition of electricity to settle every aspect of our daily life, especially the controlled usage of electricity to live tissue [1] [2] [3]. However, the curious beginning of these applications, one of the first attempts includes some “party tricks” like Nollet’s made of King’s guards to leap simultaneously [3], with the observation of red spots by the application of discharge on the skin [3], this path leads to the way to Galvani’s animal electricity [1] and finally the modern theories. In 1898, it is reported the application of electric fields as bactericidal to process water by G.W. Fuller [1]. At that point, neither the term electroporation nor the reversible and irreversible impacts resulting from the application of exogenous pulsed electric fields (PEFs) had been coined yet, this had to wait until the mid of the next century [1]. In 1948, Tiselius won Nobel Prize in chemistry due to his work on electrophoresis, the phenomenon, first described in 1807, basically points out that particles can be moved with the application of an external electric field [4]. Currently, there many techniques and modalities can be seen in different industries such as the food industry, environmental processes, as well as biotechnological and clinical applications like electrochemotherapy (ECT), gene electro transfer (GET), and nonthermal irreversible electroporation (NTIRE) for ablation of solid tumors [5].

Although there are some modalities like TTF (tumor treating fields) in which bio electrorheological models are applied to explain [6] [7], usually electro permeabilization, reversible electroporation (RE) and irreversible electroporation (IRE) are three key terms used to describe the main mechanisms practically employed [8]. Although there are still gaps in our understanding of the phenomena, there are major approaches that will be explained in a more detailed manner in this paper. In addition to theoretical explanations and mathematical models, there are different approaches to

simulating the mechanism, like molecular dynamics.

From the electrogenesis process of electric fish to the static electricity of Leyden Jar there had not been much progress for long centuries on the other hand since the 18<sup>th</sup> century there is a progressive pace in the technical aspects of pulse generation. Conventional approaches like Marx or Blumlein generators [9] are still being applied in accord with the technological progress in power electronics as well as the manufacturing process in solid-state switches. In order to provide flexible and controllable pulse generation, modular multilevel converter (MMC) based power electronics is a good option to meet different application requirements. Each application can require different voltage level. Therefore, this work is focused on MMC based pulse generation.

In this paper, the first section shortly describes the basic electroporation mechanism. Later, pulse generation and pulse generators (PGs) are briefly explained. Finally, a prototype is built to validate the performance of the proposed MMC with four half bridge submodules adapted from [9] and [20]. In [9], pulse width modulation carrier repetition time is only 40 μs in the experimental pulse generator setup with three modules. Increasing operating frequency opens a new gate to answer the question of vitality as well as new possibilities for a better and healthy life yet. Therefore, in this work, the operation frequency is selected as 100 kHz. According to the measured results, bidirectional output voltage pulsating at 100 kHz is validated on the built prototype.

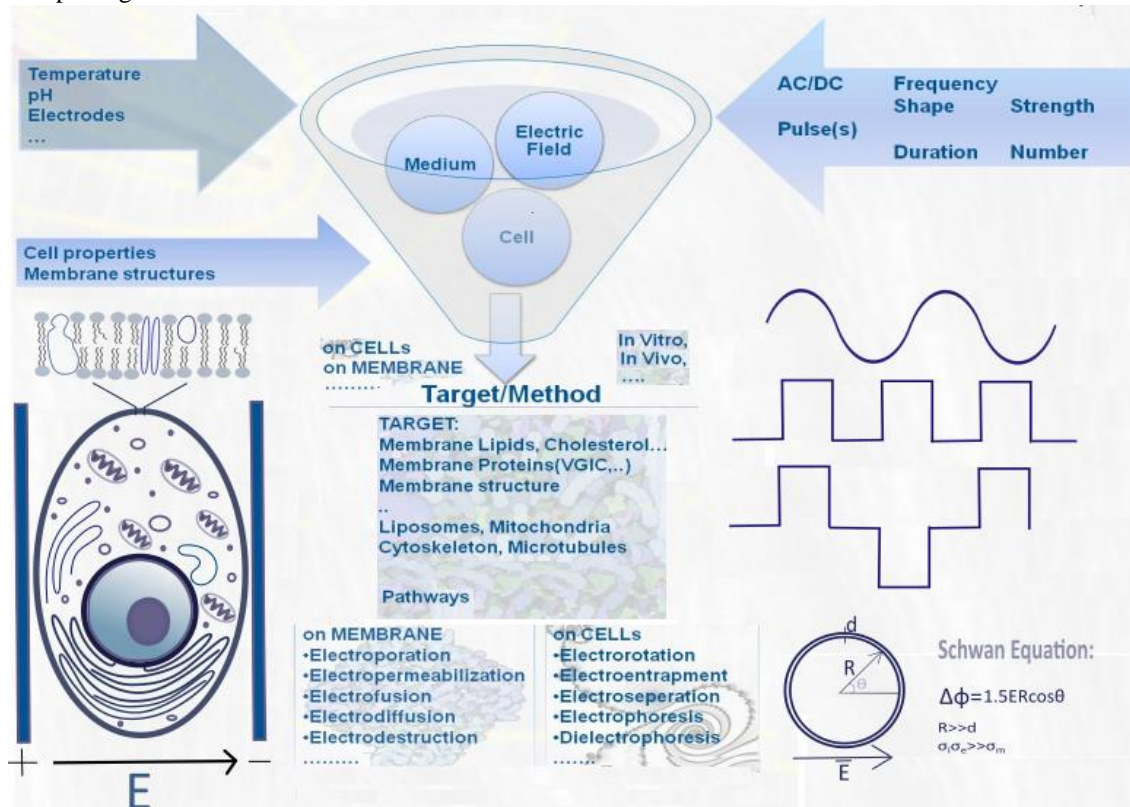


Fig. 1. Electric field applied on a single cell, Basic parameters, and Schwann equation.

## II. BASIC ELECTROPORATION MECHANISM

As it is mentioned in the previous section there are vast numbers of applications covering life sciences and growing at a rapid rate since the second half of the 20<sup>th</sup> century. However, although it is widely applied there are still unknowns about the mechanism that explains how the phenomenon does work. Fortunately, the basic models are mostly able to describe the fundamental behavior. Certainly, it is not yet easy to observe the changes in and next to the cell membrane directly, the developments in new techniques and utilizing computational modeling applications we, humanity, understand the process better every day, with every new question that is raised due to new experiments there is a new answer is waiting for us to open a new window to comprehend the behavior of the cell, as well as the vitality as a concept.

Before diving into the mechanisms behind the pore formation on the plasma membrane of the cell, a quick review of a single cell would be a good starting point. Essentially, there are two distinct types of cells, eukaryotic and prokaryotic due to the organelles they own. Eukaryotic cells as in the case of human beings or plants, have organelles having their own membranes inside the cell whereas inside prokaryotic cells there are no membrane-bound organelles, like in bacteria. There can also be a cell wall in some bacteria species and plant cells but for simplicity, this would be out of our scope in this study. The plasma membrane is consisting of a phospholipid bilayer with hydrophilic polar heads pointing interior side of the cell (cytoplasm) and extracellular environment (suspension fluid, intracellular fluid, etc.) and hydrophobic tails between the polar heads as it is shown in Figure 1 [10]. There are protein-based structures forming



channels, pumps, and gates as well as some other molecules like cholesterol or glycolipids with various functions [10]. In this study, the model that would be studied assumes only the sandwich-like phospholipid bilayer with hydrocarbon tails between the polar heads [10] [11]. In essence, commonly accepted models defining the cell membrane behavior under an external electric field define a critical voltage over which pore formation models refer formation of water fingers through this bilayer first in hydrophobic character like a hole then transforming into hydrophilic character by changing electroporation or electro permeabilization as it is represented in [8]. The shift in the membrane potential difference exceeds the critical value in a very short time, in a range changing from nanoseconds to microseconds, before permeability is increased to a detectable level and during the pulse width, if the amplitude of the pulse remains above the critical value up to milliseconds permeability remains even intensify. After those initiation and expansion stages partial recovery (micro to milliseconds) and resealing (seconds to minutes) stages occur, and memory may hold for hours before the membrane turns to its normal state. Along with the mathematical models describe mainly the induced transmembrane voltage. However, other studies are taking into consideration the other kind of effects regarding the signal molecules, voltage sensors of membrane [12], mechanical effects [13], and remodeling the cytoskeleton [14] can be found throughout the literature on pulsed electrical field studies.

Fig. 1 visually summarizes this section as well as the common notion of the use of external electrical fields on living tissues and cells. In the top and the middle part, some basic factors that affect the procedure are reviewed quickly. There are mainly three elements describing the application: Electric fields, medium, and cell. Certainly, those can be extended considering the nature of the application. For example, as an applicator, a PEF (pulsed electric field) chamber can be used or electrodes with different structures that are made up of different materials, or the suspension medium for in vitro applications may change and affects the results. For in vivo tumor ablation, the target area would be much larger compared with a microfluidic or single-cell experiment, and so does the requirements. Depending on the frequency and the strength of the pulses applied, the target may change from cell membrane to cell organelles, or bio electrorheological effects on the cytoskeleton would become assertive. On the bottom left is a stylized single cell with its nucleus and some organelles, the electrical field is symbolically shown from the positive to the negative direction between the two electrodes, and the lipid bilayer with embedded molecules like proteins depicting cell membrane can be seen. On the right-hand side, one can see basic voltage applications on the electrodes that are affecting the nature of electrical fields formed between and sometimes among the electrodes. DC or AC form can be applied, and pulse shapes, widths, strength, polarity even pulse intervals between two successive pulses are reported to have resulted in significant differences depending on the application.

### III. PULSED ELECTRIC FIELD (PEF) GENERATORS

In [15], pulse generation stages are listed under the name of pulsed power train diagram as follows: high voltage supply, energy storage, pulse compression stages, impedance matching, switch, and load. Although the fundamental mechanism, that is the application of the stored energy in a short time is the same, different approaches had been developed since the Leyden Jars of the 18<sup>th</sup> century. Marx generators, similar to Leyden Jars, were using capacitive storage and spark gaps when they are first invented [9][15]. Whereas transmission lines or Blumlein generators, although they are also among the conventional types, they take the advantage of inductive storage [15]. Today, soft switching techniques and/or resonant generators [16], solid-state pulsers [17], and wide bandgap-based designs [18][19] are joined in the methods and components. Moreover, different topologies originally designed for different purposes like modular multilevel converters used in high voltage direct current transmission lines have been able to be adapted to generate pulses [9].

#### A. Modular multilevel converter topology with half bridge submodules for bipolar pulse generation

In a typical half-bridge submodule (HB-SM), there are two switches and one capacitor as shown in Fig.2. Four HB-SMs has been interconnected to obtain bipolar waveforms in this study and that is adapted from [9], [20].

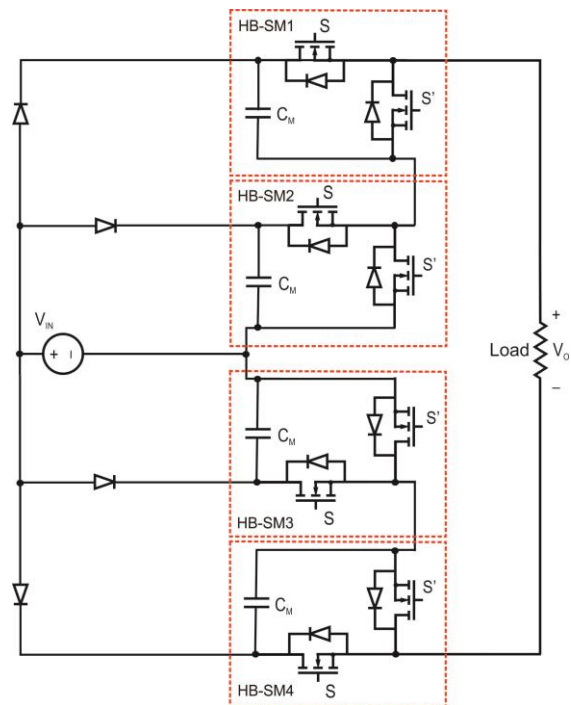


Fig. 2. Pulse generator with MMC topology in bipolar configuration with HB-SMs.

#### B. Operating principles and brief analysis

As mentioned before, each module is essentially composed of two switches  $S$ ,  $\bar{S}$  and a capacitor,  $C_M$ , that is pre-charged with voltage,  $V_{CM}$ , that is equal to the input voltage  $V_{in}$ . As it is shown in Fig. 2, employing interconnected four HB-SMs



rectangular narrow pulses with amplitudes up to  $2 \cdot V_{CM}$  to obtain bipolar outputs.

Pulse widths, and frequencies as well as the pulse shapes can be altered using different switching schemes. That is the reason the speed of the components and the technological advancements regarding those are critical to obtaining faster circuits and consequently narrower pulses.

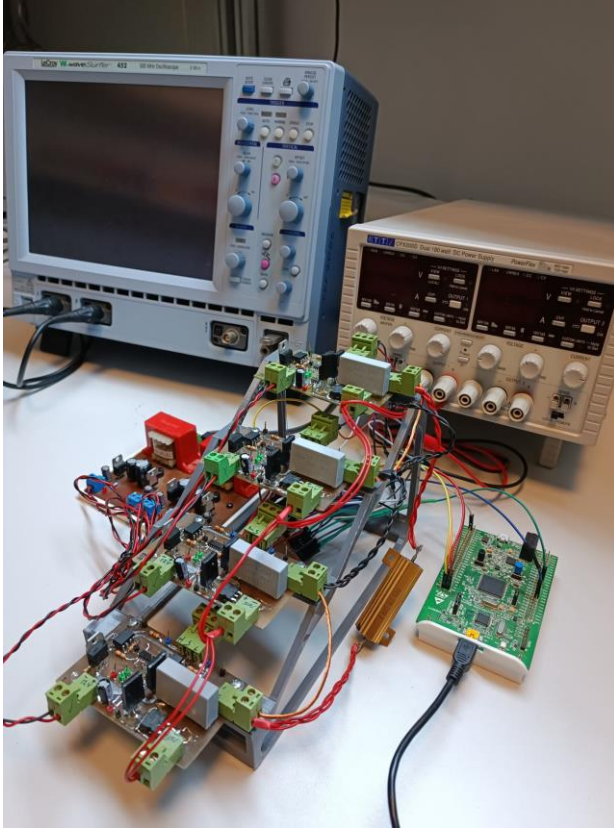


Fig. 3 The picture of the built MMC prototype with four submodules.

For each submodule, a diode is used to prevent the current flow back from the capacitors, these diodes are also known as directing/blocking diodes according to [9]. Using different switching schemes controlling  $S$  and  $\bar{S}$ , submodules can be activated or bypassed to obtain different voltage levels at the output. Activation is accomplished by the conduction of  $S$  while  $\bar{S}$  is at the off state while for bypass of the submodule  $S$  is off and  $\bar{S}$  is on. It is important to control the switches of the submodules that can be off at the same time but should not be on simultaneously to avoid short-circuiting the capacitor.

In order to produce positive pulse voltage at the load,  $S$  switches of HB-SM1 and HB-SM2 are on and their  $S'$  switches are off condition while  $S$  switches are off and  $S'$

switches are on in HB-SM3 and HB-SM4. Inversely,  $S'$  switches in HB-SM1 and HB-SM2 are on while  $S$  switches are off in HB-SM3 and HB-SM4 to produce negative pulse voltage at the load. However, it offers great flexibility the time constant of the circuit can become an issue, especially for fast-switching circuits for both charging and discharging cases.

The equivalent capacitor,  $C_{eq-ch}$ , for the general circuit in case of charging, can be approximated, considering the ideal conditions, as follows:

$$C_{eq-ch} = (2C_M) \quad (1)$$

The equivalent capacitor,  $C_{eq-dis}$  for the positive or negative peak voltage case, i.e, for the discharge of all submodules simultaneously case, is:

$$C_{eq-dis} = C_M/2 \quad (2)$$

For a resistive load, an RC circuit would be formed and the capacitor voltage can be defined, having a well-known mathematical model with a time constant of  $\tau$  equal to  $RC_{eq}$ .

#### IV. EXPERIMENTAL RESULTS

The proposed MMC which has four half bridge modules and produces bidirectional voltage at the output is validated by a prototype built in the laboratory. A load resistance with  $330 \Omega$  is implemented at the output of the MMC prototype. STF19NF20 power MOSFETs are used for switching process and STPS30SM100S diodes are used to prevent reverse current from the capacitors. The operation frequency is selected high enough as 100 kHz. The input voltage is selected as 40 V, as scaled-down compared to real application requiring kVs, to provide safe operation in laboratory. The value of the capacitor used in each sub module is  $1 \mu F$ .

In the control circuit design, STM32F407VET6 microcontroller is used to produce required switching control signals. The control signals are isolated with optocoupler to prevent short-circuit of the capacitors. UCC27200 gate drivers are used to drive power MOSFETs. The picture of the built MMC prototype with four submodules is shown in Figure 3.

Fig. 4 shows gate driver signals,  $v_s$ ,  $v_{s'}$  applying to the  $S$  and  $S'$  in a half bridge module of built MMC. Fig. 5 is shows gate driver signals of three HB-SMs.  $v_{S-HB-SM1}$  and  $v_{S'-HB-SM2}$  are the control signals of HB-SM1. Each signal of HB-SMs is separately isolated from each other. The input signals of the drivers are isolated by optocouplers while the outputs are isolated by the using of a separate power supply for each driver. Fig. 6 shows the output voltage measured across the load. Bidirectional voltage at the output with 80 V magnitude is obtained. The capacitor voltage waveforms of four HB-SMs are given in Fig. 7. The charge and discharge of the capacitor are achieved for the required output voltage. Thus, bidirectional output voltage is produced with 80 V amplitude pulsating at 100 kHz operation frequency.

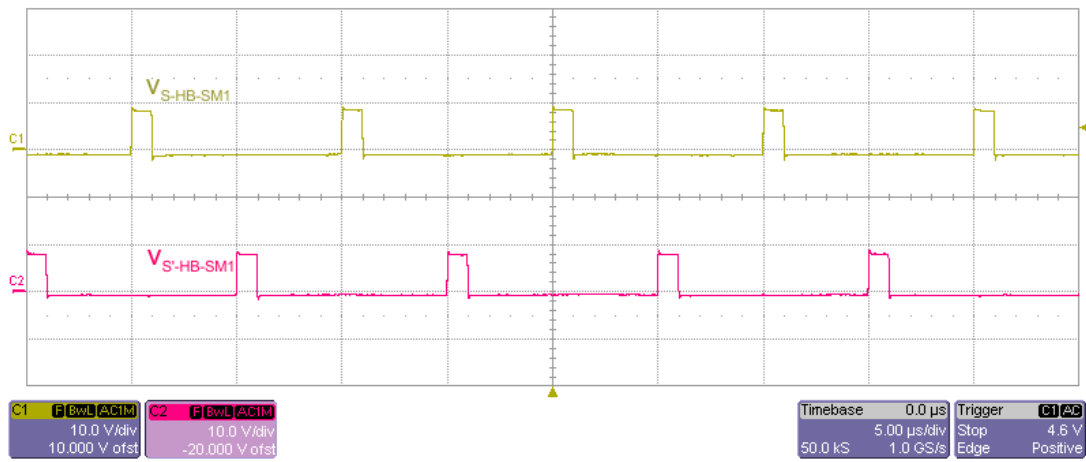


Fig. 4 The gate driver signals applying S and S' MOSFETs in a HB-SM.

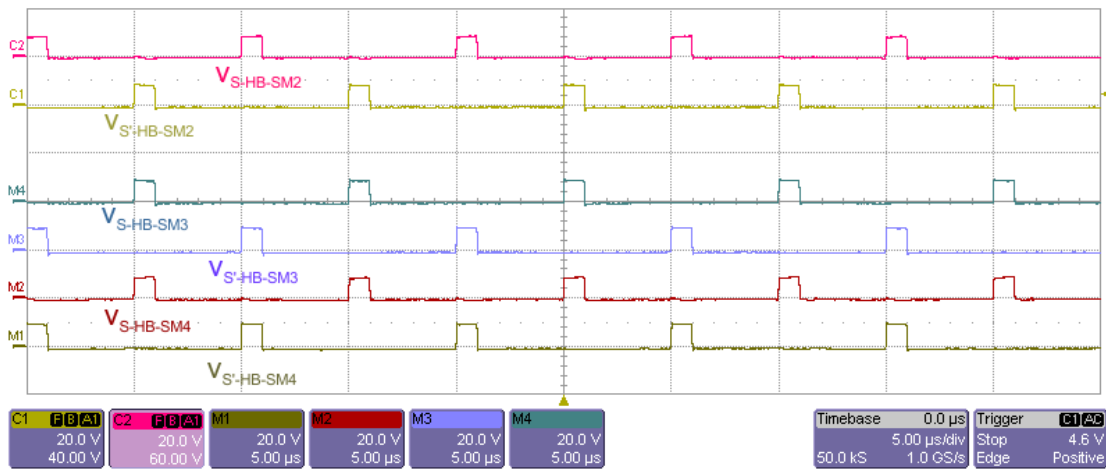


Fig. 5 The gate driver signals applying S and S' MOSFETs of three HB-SMs.

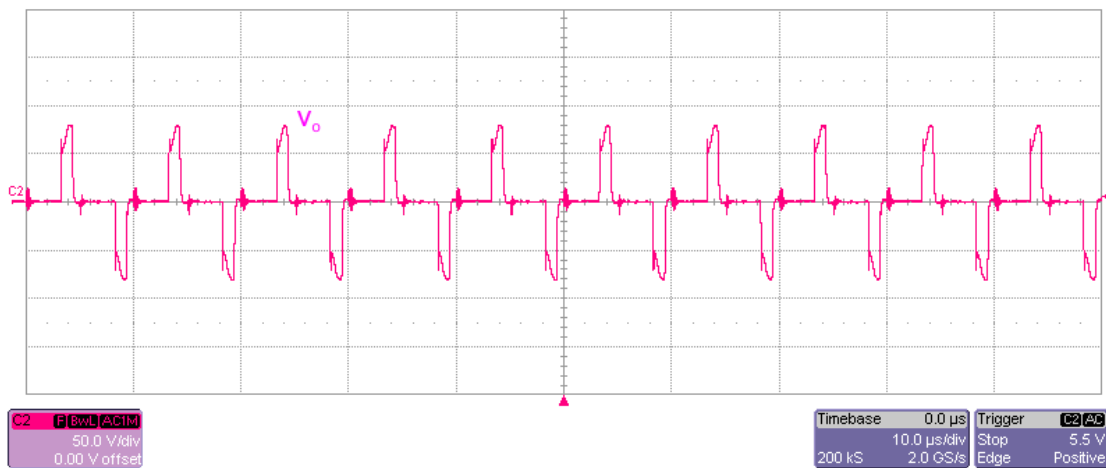


Fig. 6 Bidirectional output voltage measured across the load of proposed MMC with four HB-SMs.

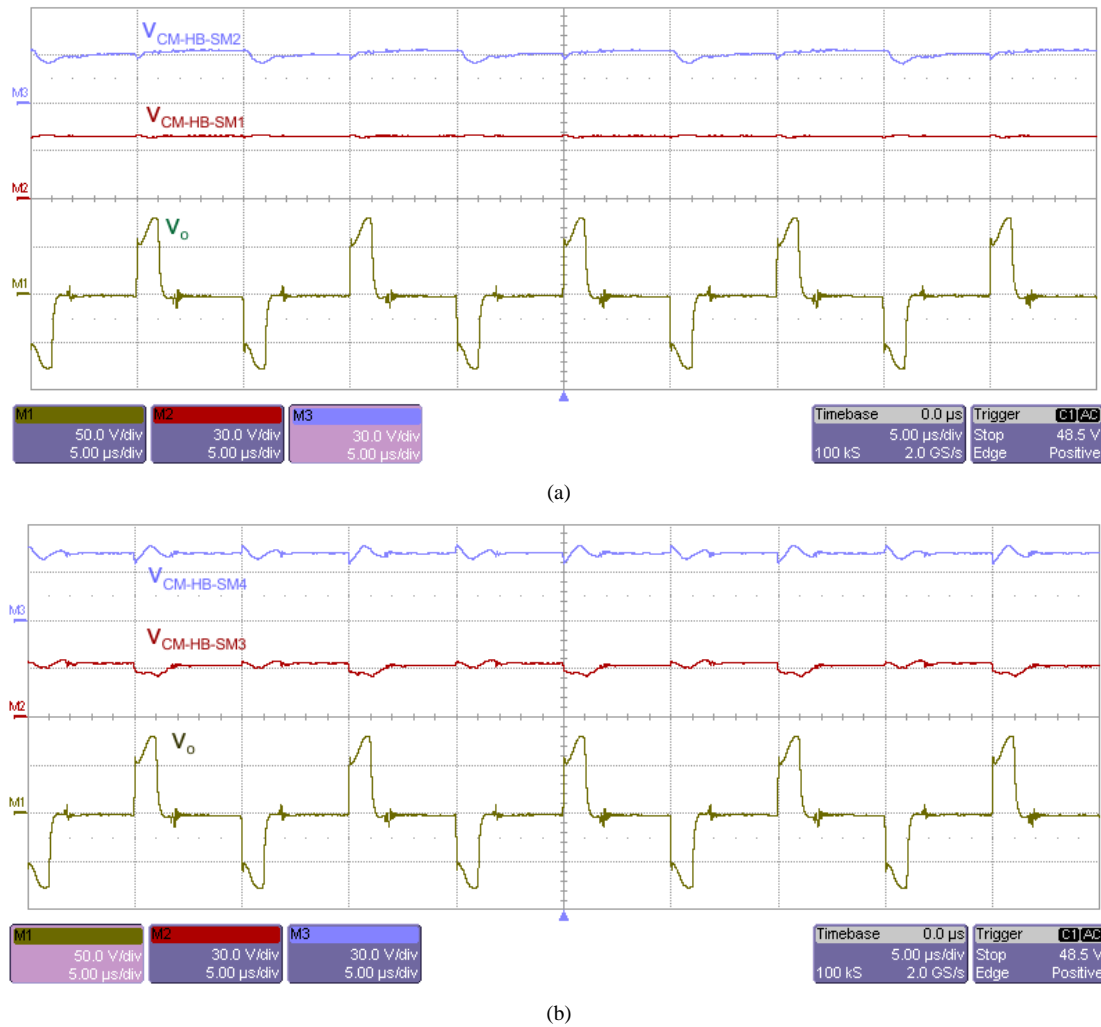


Fig. 7. The capacitor and load voltage waveforms of the built MMC: (a) The load and capacitor voltage of HB-SM1 and HB-SM2,  $V_{CM-HB-SM1}$  and  $V_{CM-HB-SM2}$  (b) The load and capacitor voltage of HB-SM3 and HB-SM4,  $V_{CM-HB-SM3}$  and  $V_{CM-HB-SM4}$ .

## V. CONCLUSION

After a short opening about the use of electric fields throughout history and all around the cultures on the earth a brief introduction on the mechanism of cell membrane manipulation with exogenous electric fields. Later, the concepts about pulse generation and the topology to be discussed were quickly explained. Finally, a scaled-down prototype was built to validate the operation of the MMC with four half modules. According to obtained results, designed MMC provides bidirectional output voltage with 80 V magnitude, pulsating at 100 kHz operation frequency. Thus, scaled-down bidirectional pulse waveforms required for electroporation application was experimentally produced. When the proposed design is used in electroporation applications, it is expected that the operation at high frequency will provide new possibilities for better and healthy life. With the modular structure, MMC has flexible design advantages in case of any requirement in electroporation applications. The voltage sharing, scale and control of the proposed MMC can be easily modified with the modular structure.

Every discrepancy with models and practical experiments and every new question that arises from this work opens a new window even a gate deep inside the very foundation of our lives. The cell and organelles membranes play a key role to understand the whole features of the phenomena related to electrical fields and cell interaction. It is expected that the proposed MMC will be potential source for the future works of the electroporation pulse generator designers.

## ACKNOWLEDGMENT

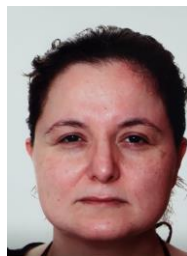
This work is supported by Pamukkale University under grant number 2022FEBE004.

Author contribution statements; Ovul Eski put contribution in literature searching, generating main ideas of this paper, analyzing of the proposed MMC with four half modules. Kemal Sahin put contribution in the building prototype and obtaining measurement results. Sevilay Cetin put contribution in the managing all stages in the paper.

## REFERENCES

- [1] A. Rolong, R.V. Davalos, B. Rubinsky, "History of electroporation." In *Irreversible Electroporation in Clinical Practice*, 2018, pp. 13-37. Springer, Cham.
- [2] C. H. Wu, "Electric fish and the discovery of animal electricity: the mystery of the electric fish motivated research into electricity and was instrumental in the emergence of electrophysiology." *American Scientist*, vol. 72, no:6, 1984, pp. 598-607.
- [3] C. D. McCaig, A.M. Rajnicek, B. Song, M. Zhao, "Controlling cell behavior electrically: current views and future potential." *Physiological Reviews*, 2005, pp. 943-978.
- [4] F. Carpi, S. Maddio, B. Pelosi, "An unexpected mossotti: His formula at the basis of dielectrophoresis in modern molecular biology." *URSI Radio Science Bulletin* 2020, no: 373, 2020, pp. 83-85.
- [5] M.L. Yarmush, A. Golberg, G. Serša, T. Kotnik, D. Miklavčič, "Electroporation-based technologies for medicine: principles, applications, and challenges." *Annu Rev Biomed Eng.*, vol. 16, no. 1, 2014, pp.295-320.
- [6] A.A. Aguilar, M.C. Ho, E. Chang, K.W. Carlson, A. Natarajan, T. Marciano, Z. Bomzon, C.B. Patel, "Permeabilizing cell membranes with electric fields." *Cancers*, vol. 13, no. 9, 2021, 2283.
- [7] X. Li, F. Yang, B. Rubinsky, "A theoretical study on the biophysical mechanisms by which tumor treating fields affect tumor cells during mitosis." *IEEE Transactions on Biomedical Engineering*, vol. 67, no: 9, 2020, pp. 2594-2602.
- [8] T. Kotnik, L. Rems, M. Tarek, D. Miklavčič, "Membrane electroporation and electropermeabilization: mechanisms and models." *Annual review of biophysics*, vol. 48, 2019, pp. 63-91.
- [9] M.A. Elgenedy, A.M. Massoud, S. Ahmed, B.W. Williams, R. Jim McDonald, "A modular multilevel voltage-boosting Marx pulse-waveform generator for electroporation applications." *IEEE Transactions on Power Electronics*, vol. 34, no. 11, 2019, pp. 10575-10589.
- [10] T. Kotnik, P. Kramar, G. Pucihar, D. Miklavcic, M. Tarek, "Cell membrane electroporation-Part 1: The phenomenon." *IEEE Electrical Insulation Magazine*, vol. 28, no. 5, 2012, pp.14-23.
- [11] K.N. Aycok, R.V. Davalos, "Irreversible electroporation: background, theory, and review of recent developments in clinical oncology." *Bioelectricity*, vol. 1, no. 4, 2019, pp. 214-234.
- [12] L. Rems, M.A. Kasimova, I. Testa, L. Delemotte, "Pulsed electric fields can create pores in the voltage sensors of voltage-gated ion channels." *Biophysical Journal*, vol. 119, no. 1, 2020, pp. 190-205.
- [13] E.P.W Jenkins, A. Finch, M. Gerigk, I.F. Triantis, C. Watts, G.G. Malliaras, "Electrotherapies for Glioblastoma." *Advanced Science*, vol. 8, no. 18, 2021, 2100978.
- [14] D.E. Chafai, F. Vostárek, E. Dráberová, D. Havelka, D. Arnaud-Cormos, D.P. Leveque, J. Janáček, L. Kubínová, M. Cifra, P. Dráber, "Microtubule cytoskeleton remodeling by nanosecond pulsed electric fields." *Advanced Biosystems*, vol. 4, no. 7, 2020, 2000070.
- [15] J. Mankowski, M. Kristiansen, "A review of short pulse generator technology." *IEEE Transactions on Plasma Science*, vol. 28, no. 1, 2000, pp. 102-108.
- [16] S.-Y. Tseng, T.-F. Wu, H.-R. Yang, J.-C. Guo, J.-C. Hung "Soft-switching series-resonant converter to generate high output voltage for processing microbes." In *Nineteenth Annual IEEE Applied Power Electronics Conference and Exposition APEC'04*, vol. 2, 2004, pp. 905-911.
- [17] R. Sundararajan, J. Shao, E. Soundarajan, J. Gonzales, A. Chaney, "Performance of solid-state high-voltage pulsers for biological applications-a preliminary study." *IEEE Transactions on Plasma Science*, vol. 32, no. 5, 2004, pp. 2017-2025.
- [18] E. Pirc, D. Miklavčič, M. Reberšek, "Nanosecond pulse electroporator with silicon carbide MOSFET s: Development and evaluation." *IEEE Transactions on Biomedical Engineering*, vol. 66, no. 12, 2019, pp. 3526-3533.
- [19] H. Sarnago, J.M. Burdio, T. Garcia-Sanchez, L.M. Mir, I. Alvarez-Gariburo, O. Lucia. "GaN-Based versatile waveform generator for biomedical applications of electroporation." *IEEE Access*, vol. 8, 2020, pp. 97196-97203.
- [20] O. Eski, S. Cetin, "Modular pulsed electric field generator based on modular multilevel converter topology with four half bridge submodules to experience biological loads", *IEEE International Conference on Power Electronics and Motion Control*, pp. 1-6, Brasov/Romania, 2022.

## BIOGRAPHIES



**OVUL ESKI** was born and raised in Karabük, Turkey. She received her Bachelor of Science, B.Sc. degree in Electrical and Electronics Engineering from the Middle East Technical University (METU), in Ankara, Turkey, in 2005. She had industry experience in several sectors including dwelling and hydropower construction as well as switchyard equipment and cable manufacturing. She had her Master of Science, M.Sc. degree in Biomedical Engineering at Pamukkale University, in Denizli, Turkey, in 2022. Her research interests cover power electronics in biomedical sciences, manipulation of cells and tissues with pulsed electromagnetic stimulations, bioelectricity, and excitable cells, especially neurons and neuroscience. She is a member of The Chamber of Electrical Engineers of Turkey, EMO.



**KEMAL SAHIN** was born in Karaman, Turkey. Between 2017 and 2021, he obtained his bachelor's degree from the Biomedical Engineering Department at Denizli Pamukkale University and is currently pursuing his master's degree. He has industry experience in various fields, including biomedical life support and therapeutic devices, 3D machine and system modeling in the mechanical field, as well as electronics. He is currently conducting research in the fields of power electronics and wireless energy transfer.



**SEVILAY CETIN** was born in Denizli, Turkey. She received her B.S. and M.S. in Electrical-Electronics Engineering from Pamukkale University, Denizli, Turkey, in 2001, 2005, respectively. She received her Ph.D in Electrical Engineering, from Yildiz Technical University, Istanbul, Turkey, in 2011.

She was a Post-Doctoral Research Associate in NY State Center for Future Energy Systems, Rensselaer Polytechnic Institute, Troy, New York, from 2013 to 2014. She worked as Assistant Professor from 2012 to 2017 and as Associate Professor from 2017 to 2022 in the Department of Biomedical Engineering, Pamukkale University, Denizli, Turkey. Currently, she works as Professor in the Department of Biomedical Engineering, Pamukkale University, Denizli, Turkey. Her main research interests include DC-DC, AC-DC converter topologies, soft switching techniques, high efficiency and high-power density power conversion for automotive, industrial and biomedical applications.



# Different Induction Motor Faults by New Proposed Random Forest Method

Cigdem Bakir

**Abstract**—Induction motors (IM) are widely used in industry. Failures in asynchronous motors cause disruptions and interruptions in production processes. Due to this situation, economic losses are experienced. Monitoring the induction motor status and monitoring the symptoms before the failure occurs is a matter of great importance in the industry. In this study, 8 different situations that may occur in the motor were monitored through the acceleration and sound data obtained from the induction motor. The feature vector was created with the Short-Term Fourier Transform (STFT) method on the acceleration and sound data obtained from the engine. The feature vectors were classified using the Random Forest (RF) method. The feature vectors created from the acceleration and sound data were also analyzed separately and the classification performance was examined. In addition, a new RF algorithm based on weight values using the Gini algorithm has been proposed. With this algorithm, the traditional RF algorithm has been developed and the success rates have been increased. In classical RF classification based on acceleration and sound data, 89.9% accuracy was achieved. The success rate of the proposed model was 95.7%. This shows that the proposed model successfully detects all types of faults in asynchronous motors. In addition, when we compared in terms of time, it was observed that the proposed model produced faster and more accurate results both in fault detection and in the production maintenance phase.

**Index Terms**—Random Forest, the Short-Term Fourier Transform, Induction motors, feature vector,

## I. INTRODUCTION

Asynchronous motors are more widely used among electric motors, especially in industry. Simple working principles, low production costs, maintenance and repair conveniences of induction motors increase their preference. At the same time, the ease of speed control provides an advantage for different production process needs.

Asynchronous motors basically consist of 4 components. These are the stator, rotor and shaft components. The general structure of the induction motor is given in our study. In these motors, the stator copper cable windings work on the principle

that the rotor placed in the stator rotates and the shaft connected to the rotor rotates by creating a magnetic field at AC voltage.

In asynchronous motors, different malfunctions may occur due to working principles, usage conditions, electrical network conditions and environmental conditions. When the causes of these failures were examined by conducting a research on 7500 engines by IEEE and EPRI, it was seen that they were gathered under 4 main headings. The proportional distributions of failure types are shown in Figure 1 [1]. The schematic variety of faults is shown in Figure 2.

Various data obtained from the engine are used to detect faults [2]. These are basically voltage, current, acceleration, vibration, temperature, sound, air flow and thermal image data. The electrical faults that occur can cause significant changes in the mechanical behavior and warming up due to the working principle of the asynchronous motor, especially in the current. Similarly, changes in current values may occur in mechanical faults.

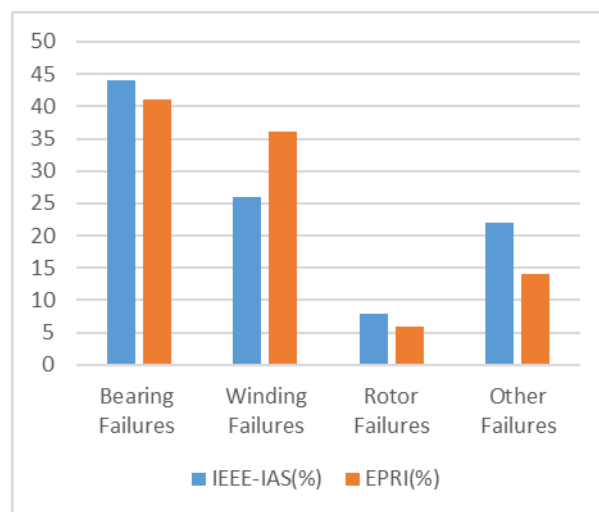



Fig.1. Proportional distribution of asynchronous motor failures

An AC powered induction motor with a single phase AC voltage of 230 V 50 Hz is generated by obtaining measurements to start and run the motor using a motor capacitor on a condition monitor screen. The motor has a removable fan guard. The original fan is replaced with modified 3D printed fans that look like the original fan but have missing fan blades. The motor is connected to the air

ÇİĞDEM BAKIR, is with Department of Software Engineering University of Dumlupınar University, Kütahya, Turkey, ( e-mail: [cigdem.bakir@dpu.edu.tr](mailto:cigdem.bakir@dpu.edu.tr)).

 <https://orcid.org/0000-0001-8482-2412>

Manuscript received Apr 14, 2023; accepted Aug 11, 2023.

DOI: [10.17694/bajece.1283336](https://doi.org/10.17694/bajece.1283336)

compressor using a metal shaft with an attached 3D-printed apparatus that allows a screw to be tightened to create a state of unbalance. The aluminum profiles holding the motor and compressor are separated from each other and the main body by polymer springs. An LSM9DS1 sensor is mounted in the motor profile. The sensor is used to measure 3-axis accelerations with 400 Hz sampling. The sensor is connected to an ESP32 microcontroller that reads measurements at a sampling time of  $\leq 3$  ms.

8 different conditions monitored for the engine are given in Table 1. Each condition is labeled with an ID and an abbreviated tag and a short description is given. Acceleration data were collected in 10 seconds with the LSM9DS1 sensor for each condition. The data is presented as raw data in the form of 'time series data'. Acceleration values are represented in mg ( $=10^{-3}$  g). In addition to the acceleration data, audio data can be obtained with the microphone. Microphone data can be generated at 48000 Hz and 16-bit resolution [3].

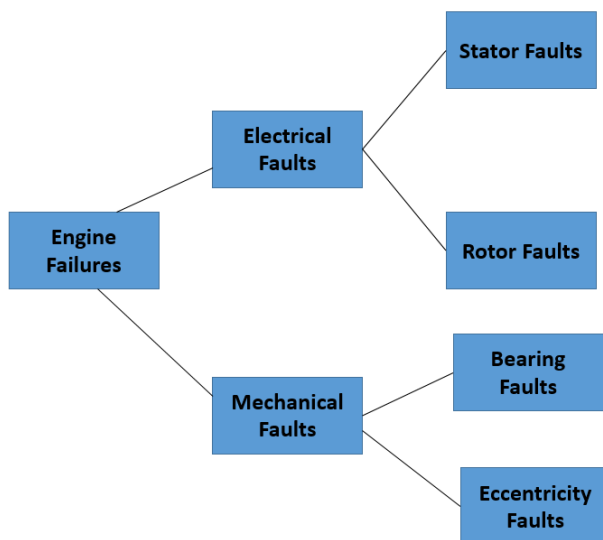


Fig.2. Types of asynchronous motor faults

TABLE I  
CONDITIONS OF THE ENGINE

ID	Label	Explanation
1	off	The engine will not start.
2	on	The engine is running.
3	cap	Motor capacitor is disabled.
4	out	Compressor outlet valve is compressed.
5	unb	A grub screw is placed on one side of the shaft to create unbalance.
6	c25	By attaching a cover to the fan housing, the air passage is reduced by 25%.
7	c75	By attaching a cover to the fan housing, the air passage is reduced by 75%.
8	vnt	The fan was replaced with a defective fan with 3 fan blades missing.

It is recommended to use the frequency attributes in the 'Frequency Attributes' folder to train a condition monitoring classifier. These features are obtained by Short-Time Fourier Transform using a 200 ms rectangular window on both the acceleration and sound data. The windows were overlapped by 80% to obtain more observations. Acceleration data is converted to frequency amplitude values of 10, 15, ..., 120 Hz (23 units) and audio data to 25, 50, ..., 2500 Hz (100 units). These values are the amplitude values corresponding to the relevant frequency band [4].

In this way, a total of 169 frequency attributes are obtained, each of which is  $3 \times 23 = 69$  acceleration and 100 sound sources. For a recording time of 10 seconds, 250 observations are formed with a 200 ms window and 80% overlap. The resulting attribute dataset, consisting of  $8 \times 250 = 2000$  observations with 250 observations per condition, is labeled with corresponding IDs and tags. In addition, the timestamp when the STFT window started is also given in the data set. While creating the attribute names, the first character axis (x,y,z), the next 3 characters are the data type (Acc, snd), the next 4 characters are the STFT frequency (Acc 10-120, snd 25-2500), and the last two characters are fixed " Hz". It means "Hz". Examples are xAcc0085Hz, zAcc0015Hz, snd0075Hz, snd1225Hz. All attributes are of floating point data type.

WEKA software was used to classify the feature vector with RF method based on machine learning. WEKA is the abbreviated name of the software platform developed at the University of Waikato for use in machine learning processes and formed from the initials of the words "Waikato Environment for Knowledge Analysis". It offers many currently widely used machine learning algorithms and methods to users [5].

One of the research subjects studied is the classification of faults through engine data, which contains important clues about faults. Measurement instruments, microphones and thermal cameras are used together with sensors to collect data. Working on raw data is both difficult and requires high microprocessor power. Therefore, size reduction is preferred by processing with signal processing methods rather than using raw data directly. Signal processing methods are effective in obtaining meaningful results by transforming the data obtained in the time-amplitude axis into different planes. This process is commonly referred to as feature extraction. Among the signal processing methods, Fourier Transform and Wavelet Transform are frequently preferred methods [6].

The feature vector obtained by feature extraction is classified using machine learning methods. Machine learning methods are evaluated in two groups as supervised and unsupervised. While class value is assigned to feature vector columns in supervised learning, class value is not assigned in unsupervised learning. Among the frequently preferred methods in machine learning and classification applications are decision tree, support vector machine, artificial neural networks, random forest method.

## II. RELATED WORKS

Some of the previous studies on the classification of asynchronous motor failures are given below.

In the study, asynchronous motor condition monitoring data set, which was collected and presented by the data collection mechanism created by Matzka et al., was tried to be classified by random forest method [7].

Yang et al., proposed a hybrid method combined with genetic algorithm to explore the possibilities of applying RF method in machine fault diagnosis and to improve classification accuracy [8]. The proposed method is illustrated by a case study on diagnosing asynchronous motor failures. Experimental results show that RF method has higher overall performance than ART-KNN, SVM and CART methods. In addition, it is seen that the proposed hybrid method (RFOGA) reaches 98.89% success.

Dos Santos et al., proposed an approach based on Random Forest and Park Vector to detect stator winding short-circuit faults in squirrel cage induction motors [9]. It is accomplished by scoring the Park Vector for both current and voltage as well as imbalance in current and voltage waveforms. The proposed strategy was tested experimentally on a custom 400-V, 50-Hz, 4-pole, 2.2-kW asynchronous motor and it was possible to simulate reconfigurable stator windings with different types of inter-turn short circuits. Even when only 1 kHz sampling frequency is used to obtain current and voltage waveforms in three phases and the use of Fast Fourier Transform is avoided, the results are quite promising. The developed solution can be used to monitor engine condition and implement advanced predictive maintenance strategies.

Vamsi et al., used an RF classifier to perform real-time fault monitoring for a squirrel cage induction motor [10]. The advantage of the proposed method is shown in the fact that it has a graphical user interface with all its functions running in the background to ultimately display the accuracy of the estimation and the status of the machine.

Sonje et al., proposed a newly developed classification model for multi-class diagnostics in an induction motor [11]. The basis of this model is RF. The stator currents were obtained using different criteria from the induction motor. Fourteen time features were determined for each current signal within the preprocessing steps on the dataset they used. These current signals form the input of the model. In addition, rotor, stator and different faults form the classes of the dataset. The obtained results were compared with the Multilayer Perceptual Neural Network (MLPNN) with various performance measures. The results are more successful than the MLPNN classification.

Saberi et al., presented a diagnostic scheme for asynchronous motors using SVM and RF [12]. First, a set of time-domain and frequency-domain characteristics are extracted from vibration and current signals under different operating conditions of the motor. Then, these features are combined and accepted as the input of the SVM-based classification model. To avoid overfitting, RF was used to identify the most dominant features that contributed to correct classification. It has been proven that the proposed method can obtain highly accurate diagnostic results for broken rotor bar and eccentricity faults and can properly handle high dimensionality of aggregated data.

Quiroz et al., proposed a new analysis-based approach for broken rotor bar failure in the engine starting state [13]. For

the detection of engine failure, data were obtained using intact and broken engines. 13 different time features were created from the current signals and the dataset was trained with RF. In addition, the 2 most important features were determined by feature selection. 98.8% success was achieved with the proposed model. The results were compared with different classification algorithms such as regression and SVM.

Harach et al., proposed a model for the EU25 inspection station to detect emission measurements [14]. In this model, different vehicle samples were tried to be determined with 9 static test points, constant engine load and engine speed using machine learning algorithms. In addition, a different platform was used for the engine condition coefficient and function coefficient. Detection of motor defects was calculated by regression method.

Elhaija et al., tried to detect broken motor rods failure and correct functioning of induction motor in asynchronous motors with optimizable neural networks [15]. Different approaches have been developed to improve existing signals. The simulation-based approach has achieved considerable success in detecting broken engine failure. In addition, they made error detection with the developed method. With the error detection based on the deep learning approach, both the error rate was tried to be reduced and the time for error detection was determined in time.

Reyes-Malanche et al., performed electric current phasor analysis for short-circuit fault diagnosis in asynchronous motors [16]. By using fuzzy logic, they tried to detect faults in asynchronous motors during the production process and to prevent possible interruptions. Online monitoring of induction motors is very important nowadays. They have developed new different approaches for this online monitoring in asynchronous motors. They have developed a non-invasive method that will reduce the timeout of induction motors and detect short-circuit faults in the stator windings of the induction motor. This method relies on phasor analysis and RMS values of line currents, followed by a small simple if-then rule to perform diagnosis and identification of stator winding faults. Results from different experimental tests on a rewound induction motor stator to induce short-circuit faults show that the proposed approach is capable of detecting and finding with high efficiency the initial and forward deficiencies in the insulation of the windings.

### III. PROPOSED RF MODEL

The Random Forest method is an ensemble learning method built on decision trees (DT). Community learning methods are popular in the field of machine learning. It can be used for classification and regression operations. As a basic principle, ensemble learning methods decide on the solution of a classification problem depending on the decisions of more than one classifier. In decision making, the highest vote (majority) or the lowest error (minimized error) approaches can be preferred.

In the RF method, there is a forest structure consisting of individual decision trees for the same problem [17]. The decision of each tree in the forest is independent of the decisions of other trees. The randomness in the method is due to the selection of variables in the creation of a decision tree.

The traditional decision tree consists of root, twig (branch) and leaves (leaf). While the root and branch structures are determined according to the variables that will determine the classification result, the leaves show the class decisions. The information gain of the variable is calculated in the selection of the variable for the roots and branches in the traditional decision tree. One of the most used calculation methods is the GINI index. Information gain is the effect of the variable on the class decision. The higher this effect, the higher the effect of the variable on the outcome. The root of the decision tree is established with the variable with the highest information gain and the branches are completed according to the gain order of the variables. In the RF method, variable selections are made randomly while creating each tree. This randomness is useful in preventing overfitting. At the same time, the samples to be selected for each tree created are selected as a subset of the whole sample with the bootstrap technique with a random approach.

The RF method was developed by Breiman [18]. Breiman, who previously developed the CART (Classification and Regression Trees) method, developed the RF technique by first developing Bagging- Bootstrap Aggregating and then Random Subspace methods and combining these methods because the CART method, which has a very good learning ability, is prone to excessive learning.

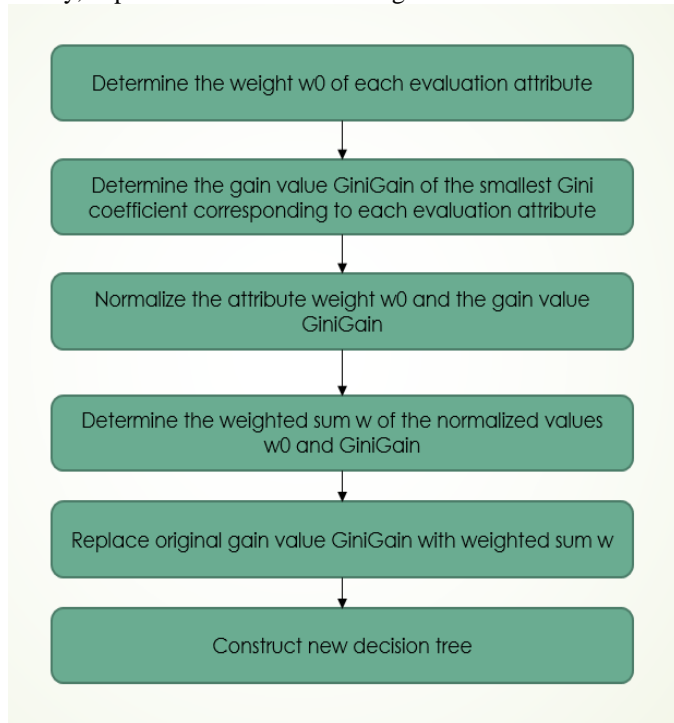


Fig.3. Proposed RF flow chart

In the RF method, 2/3 of the data set is used as training data and 1/3 is used as test data. The bagging method can be applied to many different tree methods. In a data set containing m samples, n samples (n<m) are selected with the bootstrap sampling method, and new models are created with K trees. K prediction values produced by K decision tree models are combined.

In the Random Subspace method, if there are P variables in the data set, less than P random variables are selected. For each tree created in the Bagging method, it is ensured that the

branches are made over these selected random variables. As a result, randomness is obtained through the variables.

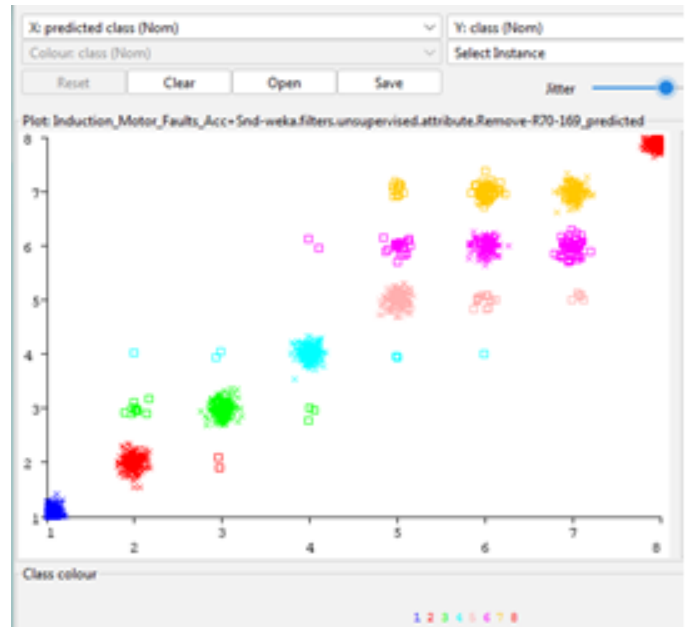


Fig.4. Graphical view of the traditional RF classification model

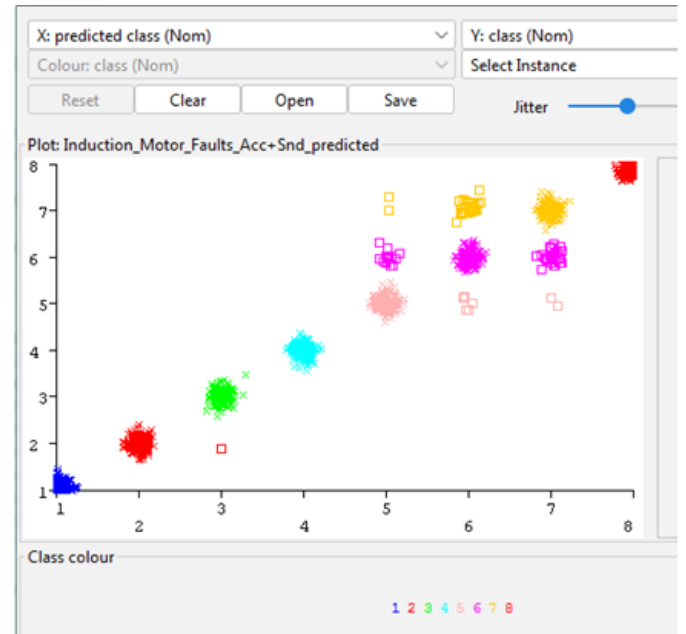


Fig.5. Graphical view of the proposed RF classification model

In the proposed RF method, the Gini algorithm is used to obtain the weights of the features and the value of the smallest Gini coefficient corresponding to the feature is taken and all the resulting values are added. The optimum weight value is calculated by constantly changing the weighted total with the original weight value. Thus, a new decision tree is created with the CART node split criterion. All weight values are normalized. The normalized value is constantly replaced by the original weight value. The flow chart of the proposed RF method is given in Figure 3 [19].

In our proposed model, unlike the traditional RF classification model, CART (trees with a weight determination



algorithm based on decision making, trial and evaluation) is used to obtain the weights of the features based on regression. The gain value of the Gini coefficient is calculated with the Gini algorithm, which determines the feature weight. The gain value of the smallest Gini coefficient is weighted by the same attribute and summed. The weighted sum is then used to replace the original gain value. This value rule is used as a new CART node split criterion to create a new decision tree, thus creating a new random forest, C-RF.

IV. EXPERIMENTAL STUDY

The performance criteria reached as a result of the traditional RF classification process by using all the features created from the acceleration and sound data are given in Table-2. Table 3 shows the confusion matrix. Figure 4 shows the graphical view of correct and incorrect classifications. When the results are examined, it is seen that an accuracy of 89.9% was obtained for all classes.

The performance criteria reached as a result of the traditional RF classification process by using all the features created from the acceleration and sound data are given in Table-4. Table 5 shows the confusion matrix. Figure 5 gives a graphical view of the correct and incorrect classifications.

When the results are examined, it is seen that 95.7% accuracy was obtained for all classes.

When the classification results are evaluated in general, it is concluded that when the features obtained from the acceleration and sound data are used together, a higher accuracy classification performance is achieved when they are used separately. It is seen that class 1 of the classes is classified exactly correctly in each approach. It is seen that classes 6 and 7 are classified correctly with the lowest accuracy in each approach. In addition, it has been observed that the proposed model gives more successful results in all types of failures compared to the traditional RF model.

For the induction motor monitoring system, it is considered that the features extracted from the acceleration and sound data with STFT and the classification performed using the RF classifier have achieved acceptable results, but it would be appropriate to obtain the features with different methods and to conduct a comparative study with different machine learning algorithms. In addition, it should be taken into account that there is a need for closer studies to real-world applications where engine failures can coexist in different combinations.

TABLE II  
TRADITIONAL RF RESULTS USING ALL DATA FEATURE

Class	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area
1	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000
2	0.992	0.006	0.958	0.992	0.974	0.971	1.000	0.998
3	0.952	0.002	0.983	0.952	0.967	0.963	1.000	0.998
4	0.972	0.003	0.980	0.972	0.976	0.972	0.999	0.996
5	0.928	0.021	0.862	0.928	0.894	0.879	0.994	0.949
6	0.664	0.046	0.672	0.664	0.668	0.621	0.954	0.723
7	0.684	0.037	0.728	0.684	0.705	0.665	0.965	0.732
8	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000
Weighted Avg.	0.899	0.014	0.898	0.899	0.898	0.884	0.989	0.925

TABLE III  
CONFUSION MATRIX RESULTS FOR TRADITIONAL RF

classified as -->	a	b	c	d	E	f	g	h
a = 1	250	0	0	0	0	0	0	0
b = 2	0	248	2	0	0	0	0	0
c = 3	0	9	238	3	0	0	0	0
d = 4	0	2	2	243	2	1	0	0
e = 5	0	0	0	0	232	13	5	0
f = 6	0	0	0	2	23	166	59	0
g = 7	0	0	0	0	12	67	171	0
h = 8	0	0	0	0	0	0	0	250

TABLE IV  
PROPOSED RF RESULTS USING ALL DATA FEATURE

Class	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area
1	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000
2	0.996	0.000	1.000	0.996	0.998	0.998	1.000	1.000
3	1.000	0.001	0.996	1.000	0.998	0.998	1.000	1.000
4	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000
5	0.972	0.010	0.935	0.972	0.953	0.946	0.998	0.985
6	0.812	0.020	0.853	0.812	0.832	0.809	0.982	0.895
7	0.872	0.019	0.865	0.872	0.869	0.850	0.987	0.912
8	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000
Weighted Avg.	0.957	0.006	0.956	0.957	0.956	0.950	0.996	0.974

TABLE V  
CONFUSION MATRIX RESULTS FOR PROPOSED RF

classified as -->	a	b	c	d	e	f	g	h
a = 1	250	0	0	0	0	0	0	0
b = 2	0	249	1	0	0	0	0	0
c = 3	0	0	250	0	0	0	0	0
d = 4	0	0	0	250	0	0	0	0
e = 5	0	0	0	0	243	5	2	0
f = 6	0	0	0	0	15	203	32	0
g = 7	0	0	0	0	2	30	218	0
h = 8	0	0	0	0	0	0	0	250

## V. CONCLUSION

Asynchronous motors have a very important place in industry and industry. However, the use of asynchronous motors brings with it many risks in daily life. At the beginning of these risks, many failures such as broken motor and bearing, maintenance of asynchronous motors, and short-circuit isolation are the most common types. In order to solve these problems, it is necessary to determine the malfunctions and other problems that may occur at the beginning. In our study, a classification method has been developed that can detect these problems in the first place and automatically detect the faults. With this model, it has given very successful results in detecting all the errors that may occur in asynchronous motors. In addition, our work is important in terms of determining the industrial maintenance phase and the detection of errors in the shortest possible time in terms of time. In this way, all major industrial failures can be detected without creating future problems and financial losses that may occur are tried to be prevented. Errors are minimized by automatic diagnosis. Our work may be a step forward for industrial applications in the future.

Our proposed RF model offers the advantages of the integrated learning method in the processing and classification of unbalanced data. Also, performing the features with CART to find the feature weight helps to include subjectivity and other qualitative assessment information to those performing fault detection in asynchronous motors. The feature weight and the gain value of the smallest Gini coefficient corresponding to the same attribute can now be weighted and summed proportionally, and the weighted sum is used to replace the original gain value as the new CART node split criterion to create the new RF. With this model, it alleviates the shortcomings of the traditional RF model's original division criterion of classification and reduces the preference for multi-valued features. Thus, the traditional RF model was developed.

Although the RF model we propose successfully detects the faults of all asynchronous motors, there are also some shortcomings. In the future, the study will be expanded by considering some deficiencies such as data instability, scarcity or unknown malfunctions.

## REFERENCES

[1] Koca, Y. B., & Ünsal, A. (2017). Asenkron motor arızalarının değerlendirilmesi. *Teknik Bilimler Dergisi*, 7(2), 37-46.

- [2] Yi, L., Sun, T., Yu, W., Xu, X., Zhang, G., & Jiang, G. (2022). Induction motor fault detection by a new sliding mode observer based on backstepping. *Journal of Ambient Intelligence and Humanized Computing*, 1-14.
- [3] Miljković, D. (2011, May). Fault detection methods: A literature survey. In 2011 Proceedings of the 34th international convention MIPRO (pp. 750-755). IEEE.
- [4] Liu, W., Chen, Z., & Zheng, M. (2020, July). An audio-based fault diagnosis method for quadrotors using convolutional neural network and transfer learning. In 2020 American Control Conference (ACC) (pp. 1367-1372). IEEE.
- [5] Eibe Frank, Mark A. Hall, and Ian H. Witten (2016). *The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques"*, Morgan Kaufmann, Fourth Edition, 2016.
- [6] Boltežar, M., Simonovski, I., & Furlan, M. (2003). Fault detection in DC electro motors using the continuous wavelet transform. *Meccanica*, 38, 251-264.
- [7] Matzka, S., Pilz, J., & Franke, A. (2021, September). Structure-borne and Air-borne Sound Data for Condition Monitoring Applications. In 2021 4th International Conference on Artificial Intelligence for Industries (AI4I) (pp. 1-4). IEEE.
- [8] Yang, B. S., Di, X., & Han, T. (2008). Random forests classifier for machine fault diagnosis. *Journal of mechanical science and technology*, 22(9), 1716-1725.
- [9] Dos Santos, T., Ferreira, F. J., Pires, J. M., & Damásio, C. (2017, May). Stator winding short-circuit fault diagnosis in induction motors using random forest. In 2017 IEEE International Electric Machines and Drives Conference (IEMDC) (pp. 1-8). IEEE.
- [10] Vamsi, I. V., Abhinav, N., Verma, A. K., & Radhika, S. (2018, December). Random forest based real time fault monitoring system for industries. In 2018 4th International Conference on Computing Communication and Automation (ICCCA) (pp. 1-6). IEEE.
- [11] Sonje, M. D., Kundu, P., & Chowdhury, A. (2017, August). A novel approach for multi class fault diagnosis in induction machine based on statistical time features and random forest classifier. In *IOP Conference Series: Materials Science and Engineering* (Vol. 225, No. 1, p. 012141). IOP Publishing.
- [12] Saberi, A. N., Sandirasegaram, S., Belahcen, A., Vaimann, T., & Sobra, J. (2020, August). Multi-Sensor fault diagnosis of induction motors using random forests and support vector machine. In 2020 International Conference on Electrical Machines (ICEM) (Vol. 1, pp. 1404-1410). IEEE.
- [13] Quiroz, J. C., Mariun, N., Mehrjou, M. R., Izadi, M., Mison, N., & Radzi, M. A. M. (2018). Fault detection of broken rotor bar in LS-PMSM using random forests. *Measurement*, 116, 273-280.
- [14] Harach, T., Simonik, P., Vrtkova, A., Mrovec, T., Klein, T., Ligor, J. J., & Koreny, M. (2023). Novel Method for Determining Internal Combustion Engine Dysfunctions on Platform as a Service. *Sensors*, 23(1), 477.
- [15] Elhajja, W. A., & Al-Hajja, Q. A. (2023). A novel dataset and lightweight detection system for broken bars induction motors using optimizable neural networks. *Intelligent Systems with Applications*, 17, 200167.
- [16] Reyes-Malanche, J. A., Villalobos-Pina, F. J., Ramirez-Velasco, E., Cabal-Yepez, E., Hernandez-Gomez, G., & Lopez-Ramirez, M. (2023).

Short-Circuit Fault Diagnosis on Induction Motors through Electric Current Phasor Analysis and Fuzzy Logic. *Energies*, 16(1), 516.

- [17] Roy, S. S., Dey, S., & Chatterjee, S. (2020). Autocorrelation aided random forest classifier-based bearing fault detection framework. *IEEE Sensors Journal*, 20(18), 10792-10800.
- [18] Breiman, L. (2001). Random forests. *Machine learning*, 45, 5-32.
- [19] Wang, J., Rao, C., Goh, M., & Xiao, X. (2023). Risk assessment of coronary heart disease based on cloud-random forest. *Artificial Intelligence Review*, 56(1), 203-232.

#### BIOGRAPHIES



**ÇİĞDEM BAKIR** is a Assistant Professor of Software Engineering at the Engineering Faculty – Kutahya University, Turkey. She received the B.S. degree in computer engineering from the University of Sakarya, in 2010, and the M.S. and Ph.D. degrees in computer engineering from Yildiz Technical University, Istanbul. She is currently pursuing the doctorate degree in computer science with the University of Yildiz Technical, Istanbul. She was a Research Assistant at Yildiz Technical University and Iğdir University. She was an Instructor at Erzincan Binali Yildirim, from 2020 to 2021. She has been an Assistant Professor with the Software Engineering Department, Dumlupınar University, since 2021. Her research interests include information security, distributed database, big data, blockchain technology, cloud computing, and computer networks.

# Three Dimensional Formation Control of Unmanned Aerial Vehicles in Obstacle Environments

Abdulmelik Bekmez, Kadir Aram, Burhanettin Can

**Abstract**—Today, the use of small unmanned aerial vehicles (UAVs) has increased due to technological advances. There has been an interest in using multiple UAVs instead of a single UAV to accomplish a given mission. This is because there are many scenarios where the capabilities of a single UAV are inadequate due to certain constraints (battery capacity, time in the air). For this reason, swarm UAV studies have increased. A swarm UAV consists of a large number of UAVs cooperating to accomplish a specific mission.

This study shows three-dimensional formation control of a swarm UAV system in an obstacle environment. A centralized control architecture is used in this process. All task assignments are made from a centralized system. The Artificial potential fields method creates the formation at the target point by avoiding obstacles. The study was carried out using the Robot Operating System (ROS). The methods were tested in Webots simulation environment. Crazyflie robots were used in the experiments.

In the simulation environment, square, star and v formations were first tested in two dimensions. Then, as an example of three-dimensional formation, the cubic and pyramid formations were created and observed.

**Index Terms**—Robotics, Swarm, UAV, Formation.

## I. INTRODUCTION

ROBOTS are used for many different purposes in daily life. Some of these needs are in the terrestrial environment, but some are also required in aerial operations. Ground or aerial robots can be used depending on the need for use. While a single robot can be used while performing these tasks, it may be more advantageous to use multi-robot systems in some cases.

Swarm robotics constitutes an area of scholarly investigation that examines the utilization of multiple autonomous agents (robots) within a system to achieve collective tasks. These tasks are typically beyond the capability of individual robots in isolation, or they are executed with greater efficiency when performed collaboratively as a cohesive group [1].

Swarm structures are inspired by observations of social insects, which show that large numbers of individuals can

interact to form intelligent systems. Swarm structures in nature do not have a centralized coordination system, but their functioning at the system level is robust, flexible and scalable. These characteristics have also been a motivation for swarm robots[2].

Unmanned aerial vehicles (UAV) are used in both civilian and military fields. These areas include fire detection and response applications [3][4], search and rescue operations [5] and road construction works [6]. In the contemporary era, a plethora of missions demands UAVs to navigate over expansive regions or accomplish tasks within tight time constraints. However, relying on a single UAV might not suffice to meet the performance criteria due to its constrained dimensions and limited battery capacity [7].

The use of multiple UAVs has many advantages. These advantages are time efficiency, cost, simultaneous actions, complementarity, fault tolerance and flexibility [8]. Therefore, it is used in many fields, such as Video surveillance [9], traffic monitoring [10], search and rescue [11], Simultaneous Localization and Mapping [12]. UAVs frequently engage in formation flight, where the relative distances between each pair of UAVs remain constant, leading to the cohesive movement of the entire formation. To preserve the shape of the formation, it suffices to maintain the distances between a specific number of agent pairs, effectively ensuring a constant distance between all pairs in the formation [13].

Brandao and Filho used a centralized approach for formation control in their study. These layers are responsible for creating the desired path for the formation, ensuring the robot's desired posture and generating the control signal for each robot to reach its desired position [14]. Kim et al., designed a framework for swarm systems that uses a decentralized control approach. Each agent uses artificial potential functions to form swarms and control their formation [15]. Li et al. created a game model for UAVs to ensure flight safety by avoiding obstacles in a given formation [16]. Yavuz et al., developed a method for task assignment to multiple UAV systems. First, target locations are clustered according to the number of UAVs. Then, appropriate target assignment was made with the Hungarian algorithm. Finally, the optimal path for the mission location was calculated with the ant colony algorithm [17].

This study provides formation control of a system consisting of swarm UAVs. A centralized system calculates formation points, and proportional control is used for position control. Hungarian algorithm was used to assign the formation points to the robots in the swarm. The experiments tested star, crescent, cube, and v-formation shapes. The artificial potential field method is used for obstacle avoidance during the formation

**Abdulmelik Bekmez** is with the Department of Computer Engineering, Engineering Faculty, Fatih Sultan Mehmet Vakıf University, İstanbul, 34445 TURKEY e-mail: abdulmelik.bekmez@stu.fsm.edu.tr

• <https://orcid.org/0009-0008-4211-941X>

**Kadir Aram** is with the Department of Computer Engineering, Engineering Faculty, Fatih Sultan Mehmet Vakıf University, İstanbul, 34445 TURKEY e-mail: karam@fsm.edu.tr

• <https://orcid.org/0000-0002-5780-6334>

**Burhanettin Can** is with the Department of Computer Engineering, Engineering Faculty, Fatih Sultan Mehmet Vakıf University, İstanbul, 34445 TURKEY e-mail: bcan@fsm.edu.tr

• <https://orcid.org/https://orcid.org/0000-0002-0998-4371>

Manuscript received Aug 18, 2023; accepted Sep 18, 2023.

DOI: 10.17694/bajece.1345915



process. In the study's second part, materials and methods are shown; in the third part, the system's structure explained, in fourth part the results of the tests are shown. In the last section, the results obtained are interpreted.

## II. MATERIALS AND METHODS

Utilizing multiple robots enhances the system's robustness compared to a single robot. A multi-robot system exhibits unparalleled superiority, particularly in scenarios involving environmental detection, survivor search, and other intricate tasks. The critical issue is ensuring that multiple robots can be organized effectively. This requires communication and control according to the desired situation [18].

### A. Formation Control Methods

Some of the main elements that need to be considered when controlling the formation are sharing information within the swarm agents and assigning tasks to these agents. Here the decision maker has to take into account issues such as fault tolerance and careful use of energy. There are two main control schemes in formation control: centralized and distributed control methods. In a centralized control framework, a collective of cooperative robots receives control directives from a potent central unit, a robot within the formation equipped with robust computational capabilities or a control station. The core unit establishes communication with all team members, assimilates information from each individual, optimizes vehicle coordination, monitors mission progress, and handles any individual faults that may arise [19]. The distributed formation control approach operates independently of a central control center, granting equal status to all UAVs within the formation. Within the distributed control structure, the global control challenge is decomposed into several sub problems, each independently tackled by the controller within individual UAVs. Through distributed cooperative control, the formation of the UAV swarm can be successfully achieved [20].

### B. Formation Control Strategies

Three main problems must be considered when performing formation control of multiple UAVs. The first one is the task of directing the agents to form the desired formation structure. The second one is the preservation of the created formation. For this, a control strategy should be created. The third is the reconstruction of the formation. When performing operations in the environment, the formation must be preserved in case the formation encounters obstacles or other disturbing factors. Different control strategies have been proposed for such problems. [7].

1) *Virtual structure*: Within the virtual structure approach, the entirety of the formation is regarded as a unified entity capable of coordinated movement akin to a rigid body in a specified direction and orientation. Moreover, this approach ensures the preservation of the geometric arrangement among multiple vehicles based on a reference point within the virtual structure. [21].

2) *Leader-Follower*: The leader-follower technique involves designating a specific robot as the leader, with the remaining robots acting as followers. The leader communicates with all other UAVs within the group and determines the path for the entire UAV formation. The follower UAVs track the leader's reference state and carefully maintain a prescribed distance and angle to avert collisions and disarray, achieving overall consistency in the formation. [22].

3) *Behavior-based*: The behavior-based UAV formation control strategy represents a distributed approach that governs UAVs based on fundamental behavior rules. A robot's overall behavior results from a fusion of various sub behaviors, such as encompassing going to the goal, obstacle avoidance, speed matching, wall-following and formation maintenance. In practical scenarios, each robot's behavior is expressed as a vector comprising magnitude and direction, with the vector's weight modifiable through parameter adjustments. The robot selects appropriate behaviors to generate a movement command by assimilating information from the surrounding environment. The robot's collective behavior vector emerges as the summation of all individual sub behavior vectors [23].

4) *Artificial Potential Field*: Artificial Potential Field (APF) is a method that emulates repulsion and attraction forces akin to those in a gravitational field. The fundamental concept revolves around utilizing attractive and repulsive forces to generate control actions. These forces correspond to the negative gradients of the attractive and repulsive potential fields, respectively. The attractive force facilitates formation convergence, while the repulsive force ensures collision avoidance within the formation [7].

### C. Robot Operating System (ROS)

ROS (Robot Operating System) is an open-source and freely available robotics software framework suitable for deployment in diverse research and commercial applications [24]. It is a software platform that enables the development of robot applications with the libraries and packages it contains [25]. ROS provides many conveniences to users. Some of them are the support of multiple programming languages, the ability to communicate between processes and the use of the developed program code in other projects. It also includes visualization tools [26]. Communication in ROS is done through messages. These messages are transmitted through topics. The nodes in ROS communicate over these topics. The ROS Master, launched at the start of the ROS server, manages the communication between all nodes. The node sends all its information to the ROS master, including the type of data it sends or receives. Nodes that send data are called publisher nodes, and nodes that receive data are called subscriber nodes. The ROS Master has all the publisher and subscriber information running on computers [24].

### D. Simulation

Webots is an open-source and multi-platform desktop application for simulating robots. It provides a development environment for modeling, programming and simulating robots. Webots was chosen because of its rich documentation, ROS2

integration is effortless to set up and configurable, and it has a ready-made crazyflie model.

### E. Quadrotor Robot

The Crazyflie 2.0 nano quadrotor helicopter, functioning as an open-source experimental platform, holds significance for research and educational endeavors in robotics. Weighing a mere 27g and fitting comfortably within one's hand, this quadcopter boasts a diminutive stature. This compact size renders it amenable to indoor flights conducted in closely packed formations. Additionally, owing to its limited inertia, the Crazyflie can endure high-velocity impacts, minimizing potential hazards to humans [27] [28].

## III. IMPLEMENTATION

The first step is to select a suitable software architecture for the implementation of the application. In order to create the desired formation for the swarm, the locations where the robots should go were determined. Then the most suitable robot is selected for the determined formation points. A user interface was designed to observe all these processes.

### A. Software Architecture

The application was developed on a docker image with Ubuntu 22.04 operating system. ROS2 Humble version was used. This version provides tier 1 support on Ubuntu 22.04. All the requirements for the project to be installed and run are created as a Docker image. In this way, users can control both the physical robots and the simulation environment with a container created using the relevant docker image. There is a position controller node for each agent in the swarm. These position controller nodes were developed using the Rust programming language. A user interface has been developed for the control of the swarm system. While developing the interface, the egui library in the Rust language was used. The egui library uses immediate mode architecture while developing the interface. The whole interface is redrawn in each iteration. This allows the system to be developed more dynamically. The user interface is shown in Figure 1.

The following commands can be given to swarm agents through the interface.

- Departure
- Landing
- Navigation
- Formation

### B. Calculation of Formation Points

Formation points are calculated using a circle formation algorithm. All formations are formed by circles, and these circles are calculated and combined one by one. Most regular polygon geometries can be created using circles. When creating any circle formation, a generic algorithm is generated by taking parameters such as circle center, radius and distance between points.

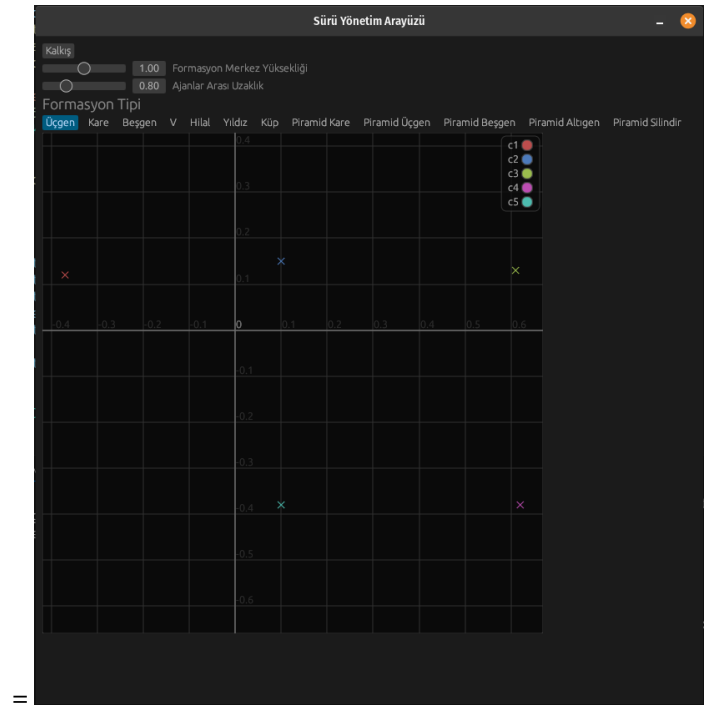


Fig. 1. User Interface

Three-dimensional formations are a combination of two-dimensional ones. Separate two-dimensional formations are created for the desired three-dimensional formation shape. These formations may vary according to the desired shape. For example, two squares are created for a cube formation. For a square pyramid, one is sufficient.

Before calculating the formation points, the radius of the circle is calculated. This is because the radius of the circle will change depending on the distance between the formation points. The radius of the circle is calculated according to Equation 1.

$$r = \frac{d/2}{\sin(\theta/2)} \quad (1)$$

Here :  $d$ =distance between formation positions,  $\theta$  = angle between formation positions.

A formation shape using a circle is shown in Figure 2.

Theta value in Figure 2 indicates the angle between the formation points. This value is calculated in Equation 2.

$$\theta = 2\frac{\pi}{n} \quad (2)$$

$n_i$  indicates the number of position in the formation.  $\theta$  indicates the angle difference between formation points.

Using these equation, the angles of all points are calculated.

$$(0 \ \theta \ 2\theta \ \dots \ 2\pi - \theta)$$

Once all angles are calculated, the required positions are calculated using these angles.

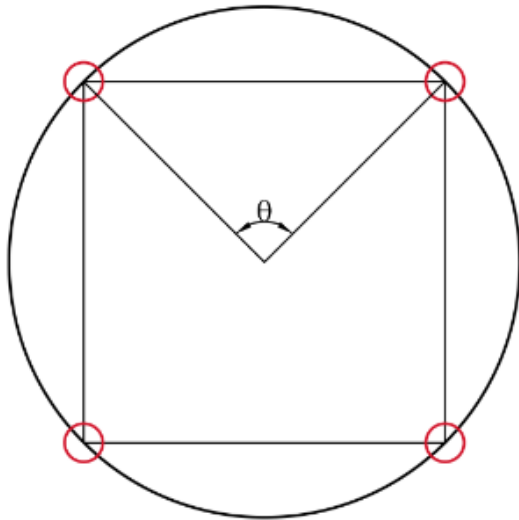


Fig. 2. Formation Circle Points

$$\begin{pmatrix} r\cos(0) + x_0 & r\sin(0) + y_0 & z_0 \\ r\cos(\theta) + x_0 & r\sin(\theta) + y_0 & z_0 \\ r\cos(2\theta) + x_0 & r\sin(2\theta) + y_0 & z_0 \\ \dots & \dots & \dots \\ r\cos(2\pi - \theta) + x_0 & r\sin(2\pi - \theta) + y_0 & z_0 \end{pmatrix}$$

1) *Star Formation:* Two five-element circles are used to calculate the star formation points. The initial angles and radius of these circles are determined to obtain the star formation. It is sufficient that the difference between the starting angles of the circles is  $36^\circ$ . This value is equal to the length of one side of a regular polygon which has 10 points. To calculate their radius, the ratio between the radius of the inner circle and the radius of the outer circle must be known. The circles used for the star formation are shown in Figure 3.

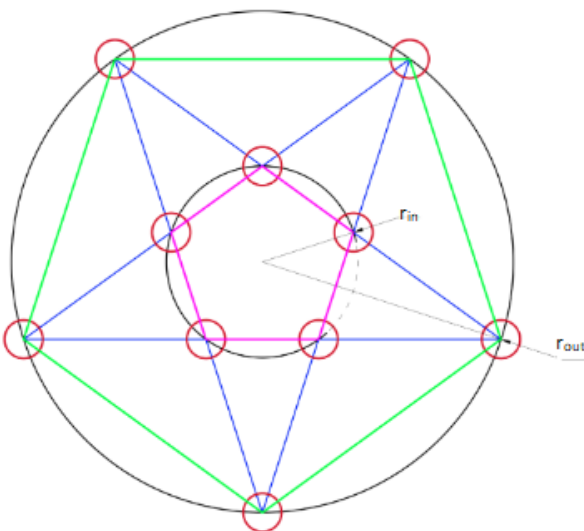


Fig. 3. Circles in Star Formation

In order to find the ratio between  $r_{out}$  and  $r_{in}$ , a corner of the star in Figure 4 is considered. Using the angles and lengths at this corner, the ratio between  $r_{out}$  and  $r_{in}$  is calculated.

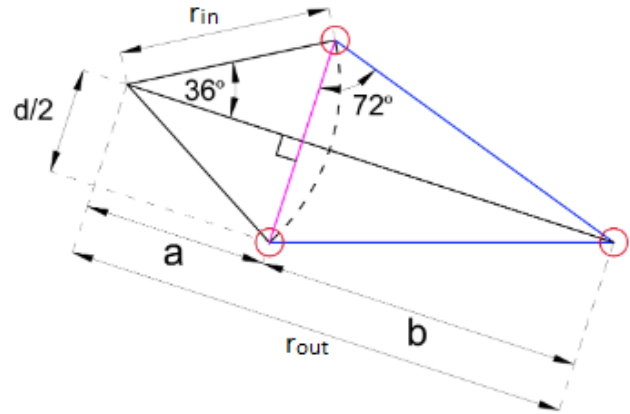


Fig. 4. Star Formation Points

In Figure 4, the value  $a$  indicates the height of the triangle formed inside the inner circle.  $b$  is the height of the triangle formed by the corner of the star.  $d$  is the distance between the points in the inner circle. These values are calculated according to equations 3-8.

$$\alpha = \frac{360}{5 * 2} = 36 \tag{3}$$

$$\beta = \frac{180 - 36}{2} = 72 \tag{4}$$

$$a = \frac{d}{2} \cot(36) \tag{5}$$

$$b = \frac{d}{2} \cot(72) \tag{6}$$

$$r_{out} = a + b \tag{7}$$

$$r_{in} = \frac{d}{2\sin(36)} \tag{8}$$

After calculating the values of  $r_{in}$  and  $r_{out}$ , the ratio between the radius is obtained according to equation 9.

$$ratio = \frac{r_{out}}{r_{in}} = 2.618 \tag{9}$$

After the ratio between the radiuses is calculated, two circle formations are created. The radius of the circle to be formed on the outer side is calculated as 2.618 times bigger, and the starting angle is calculated as  $36^\circ$  bigger. After the circle points are calculated, the star formation points are calculated by combining them.

2) *V Formation*: In order to calculate the points required for the V formation, first, a circle is created that encompasses the V shape to be formed. Then the intervals of these three points are filled with points according to the desired distance parameter between the agents. V formation points are shown in Figure 5.

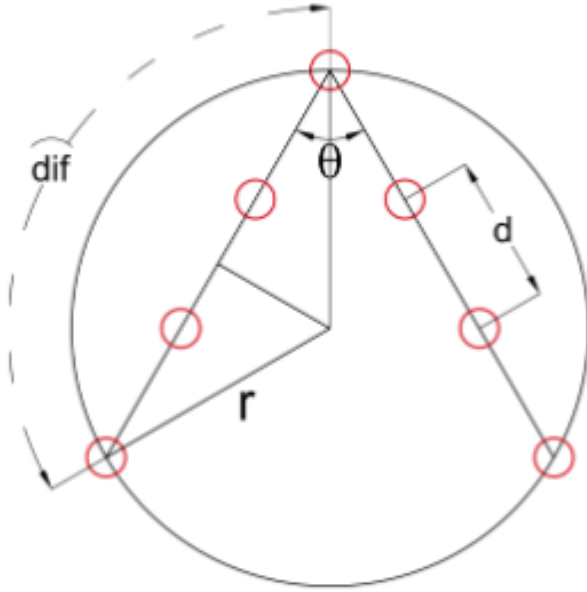


Fig. 5. V Formation Points

$\theta$  denotes the angle of the triangle forming the V-shape,  $d$  is the distance between the formation points denotes  $r$  is the radius of the circle enclosing the triangle formed as a result of the formation.

$$r = d \frac{\text{intervalCount}}{2\cos(\frac{\theta}{2})} \quad (10)$$

To calculate the points in a circle, it is enough to calculate three angles.

$$\text{diff} = 180 - \theta \quad (11)$$

And the angles are:

$$(90 \ 90+\text{diff} \ 90-\text{diff})$$

The positions are formed as follows after the angle and difference values are found.

$$\begin{pmatrix} r\cos(90)+x_0 & r\sin(90)+y_0 & z_0 \\ r\cos(90+\text{diff})+x_0 & r\sin(90+\text{diff})+y_0 & z_0 \\ r\cos(90-\text{diff})+x_0 & r\sin(90-\text{diff})+y_0 & z_0 \end{pmatrix}$$

### C. Assigning Formation Points

After the formation points are calculated, they need to be assigned to the robots in the swarm. Hungarian Algorithm, a simultaneous assignment algorithm, was used for this assignment process. In order to minimize the path taken by the

robots in the swarm and reduce the probability of collision, the formation points should be assigned to the most suitable robot. For this reason, the cost for each robot to go to each formation point is calculated, and these values are collected in a matrix. The values in this cost matrix are calculated according to equations 12-13.

$$\text{dist} = \sqrt{(x_1 - x_0)^2 + (y_1 - y_0)^2} \quad (12)$$

$$\text{Cost}_{ij} = \text{dist}^2 \quad (13)$$

After the cost matrix is calculated, this matrix is given as input to the Hungarian algorithm. The algorithm returns the required row and column indexes for the scenario with the least cost. By using these indexes, formation points are assigned to the relevant swarm robots.

### D. Avoiding Obstacles

An artificial potential field strategy is used for robots to avoid obstacles while reaching the desired goal. According to the basic principle of this strategy, the workspace is under the influence of artificial potential forces. The positions of the obstacles and the positions of the agents have a repulsive effect, pushing the robot away from the obstacles, while the destination point, where the robot needs to go, has an attractive effect, pulling the robot towards the destination. Robot control is achieved by finding the sum of the repulsive and attractive forces (net force) at any point [29].

The total force is calculated as in Equation 14-16 [29].

$$F_{net} = F_{rep}(X_R) + F_{att}(X_R) \quad (14)$$

where  $F_{net}$  is net force,  $F_{rep}$  is repulsive force,  $F_{att}$  is attractive and  $X_R$  is robots' position. The repulsive force can be defined as,

$$F_{rep}(X_R) = \begin{cases} d \leq d_m, & K_r \times \left( \frac{1}{(d(X_R - X_o) - d_0)^2} - \frac{1}{(d_m - d_0)^2} \right), \\ d > d_m, & 0 \end{cases} \quad (15)$$

where  $d(X_R, X_o)$  is distance between robot and obstacle,  $d_m$  is distance threshold,  $d_0$  is minimum safety distance from obstacles,  $K_r$  is repulsive potential field constant. The attractive force can be defined as,

$$F_{att}(X_R) = K_a \times d(X_R - X_G) \quad (16)$$

where  $d(X_R - X_G)$  is distance from robot to goal point,  $K_a$  is attractive potential field constant. Figure 6 shows the robots moving towards the obstacle. Figure 7 shows the robots avoiding the obstacle.



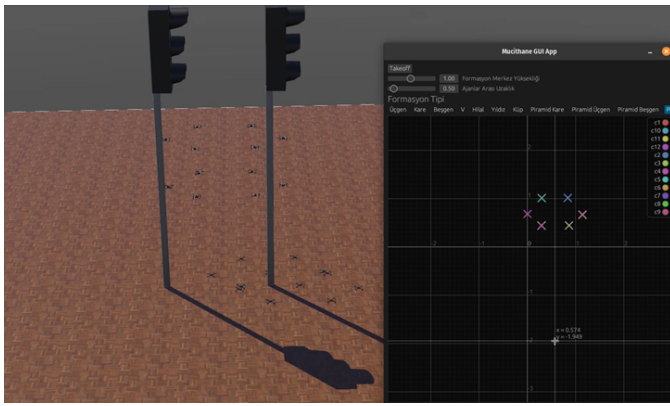


Fig. 6. Robots with Obstacles

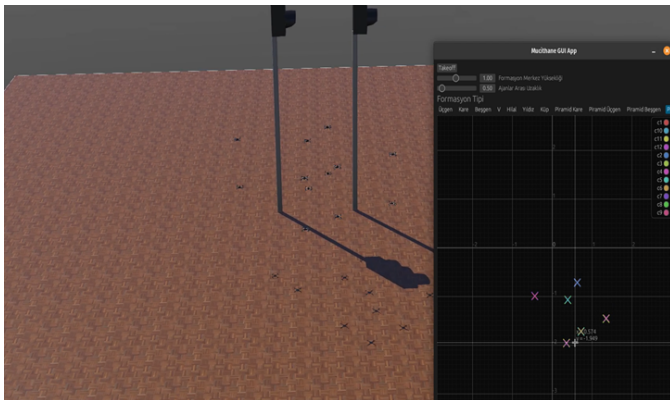


Fig. 7. Robots Avoiding Obstacles

#### IV. SIMULATION AND TEST

Webots simulation environment was used to test the swarm system. Crazyflie quadcopters were selected for swarm robots. The crazyflie firmware was used in the simulation to ensure that the crazyflie robots produce the most realistic output. Webots allows the development of separate plugins to use any robot with ROS2. The developed plugins work as a separate process by running each robot. The firmware for Crazyflie was developed using C programming language. However, the developers have provided the relevant links to use these codes in Python programming language. These links were used in the Webots plugin to generate the motor control outputs required for simulation.

Various test environments have been created in order to test the robots in the swarm according to different algorithms during the simulation process. Since the formation types are dependent on the number of robots in the swarm, environments with different numbers of robots were created. For testing the obstacle avoidance of the swarm, some static obstacles with known positions were added to some environments.

##### A. 2D Simulation Results

Three different formations were tested for the two-dimensional simulation test.

1) *Square Formation*: The simulation result is in Figure 8.

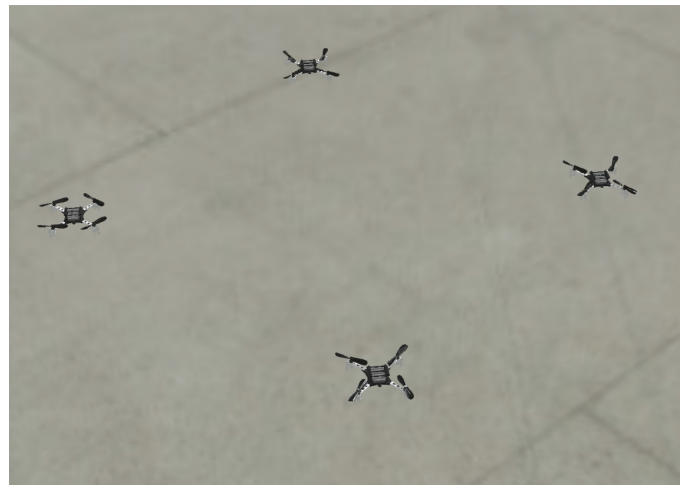


Fig. 8. Square Formation

2) *Star Formation*: The simulation result is in Figure 9.



Fig. 9. Star Formation

3) *V Formation*: The simulation result is in Figure 10.

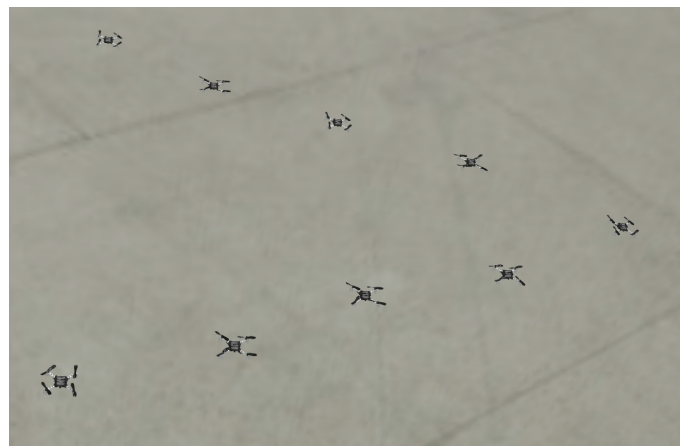


Fig. 10. V Formation

### B. 3D Simulation Results

Three different formations were tested for the three-dimensional simulation test.

1) *Cube Formation*: The simulation result is in Figure 11.



Fig. 11. Cube Formation

2) *Pyramid Formation*: The simulation result is in Figure 12.



Fig. 12. Pyramid Formation

3) *Hexagonal Prism Formation*: The simulation result is in Figure 13.



Fig. 13. Hexagonal Prism Formation

simulation environment was used to test the methods. Crazyflie robots were selected to create a swarm robot. There are static obstacles in the simulation environment where the experiments are performed. The artificial potential field method was used for obstacle avoidance. Two-dimensional formation shapes such as square, star and v formation were tested. Cube and pyramid formations were created as three-dimensional formations. When the simulation results were analyzed, it was observed that success was achieved. As an extension of this study, it is considered to manage robots not from a single center but with a distributed control method.

### REFERENCES

- [1] L. Bayındır, "A review of swarm robotics tasks," *Neurocomputing*, vol. 172, pp. 292–321, 2016.
- [2] E. Şahin, "Swarm robotics: From sources of inspiration to domains of application," in *International workshop on swarm robotics*. Springer, 2004, pp. 10–20.
- [3] Z. Jiao, Y. Zhang, J. Xin, L. Mu, Y. Yi, H. Liu, and D. Liu, "A deep learning based forest fire detection approach using uav and yolov3," in *2019 1st International conference on industrial artificial intelligence (IAI)*. IEEE, 2019, pp. 1–5.
- [4] N. Jayapandian, "Cloud enabled smart firefighting drone using internet of things," in *2019 International Conference on Smart Systems and Inventive Technology (ICSSIT)*. IEEE, 2019, pp. 1079–1083.
- [5] M. Silvagni, A. Tonoli, E. Zenerino, and M. Chiaberge, "Multipurpose uav for search and rescue operations in mountain avalanche events," *Geomatics, Natural Hazards and Risk*, vol. 8, no. 1, pp. 18–33, 2017.
- [6] F. Tükenmez, "Harita mühendisliğinde iha ile karayolu projelendirme," *Türkiye Fotogrametri Dergisi*, vol. 3, no. 2, pp. 53–61, 2021.
- [7] H. T. Do, H. T. Hua, M. T. Nguyen, C. V. Nguyen, H. T. Nguyen, H. T. Nguyen, and N. T. Nguyen, "Formation control algorithms for multiple-uavs: a comprehensive survey," *EAI Endorsed Transactions on Industrial Networks and Intelligent Systems*, vol. 8, no. 27, pp. e3–e3, 2021.
- [8] G. Skorobogatov, C. Barrado, and E. Salami, "Multiple uav systems: A survey," *Unmanned Systems*, vol. 8, no. 02, pp. 149–169, 2020.
- [9] A. Saeed, A. Abdelkader, M. Khan, A. Neishaboori, K. A. Harras, and A. Mohamed, "On realistic target coverage by autonomous drones," *ACM Transactions on Sensor Networks (TOSN)*, vol. 15, no. 3, pp. 1–33, 2019.

### V. CONCLUSION

In this study, formation control of UAV swarm was realized. Formations are both two-dimensional and three-dimensional. All control commands of the robots come from a single location. In this respect, it has a centralized control method. The study was implemented in the ROS environment. Webots

- [10] V. Sharma, H.-C. Chen, and R. Kumar, "Driver behaviour detection and vehicle rating using multi-uav coordinated vehicular networks," *Journal of Computer and System Sciences*, vol. 86, pp. 3–32, 2017.
- [11] J. Scherer, S. Yahyanejad, S. Hayat, E. Yanmaz, T. Andre, A. Khan, V. Vukadinovic, C. Bettstetter, H. Hellwagner, and B. Rinner, "An autonomous multi-uav system for search and rescue," in *Proceedings of the first workshop on micro aerial vehicle networks, systems, and applications for civilian use*, 2015, pp. 33–38.
- [12] J. Dong, E. Nelson, V. Indelman, N. Michael, and F. Dellaert, "Distributed real-time cooperative localization and mapping using an uncertainty-aware expectation maximization approach," in *2015 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2015, pp. 5807–5814.
- [13] B. D. Anderson, B. Fidan, C. Yu, and D. Walle, "Uav formation control: Theory and application," in *Recent advances in learning and control*. Springer, 2008, pp. 15–33.
- [14] A. S. Brandao and M. Sarcinelli-Filho, "On the guidance of multiple UAV using a centralized formation control scheme and delaunay triangulation," *Journal of Intelligent Robotic Systems*, vol. 84, no. 1-4, pp. 397–413, Nov. 2015. [Online]. Available: <https://doi.org/10.1007/s10846-015-0300-5>
- [15] D. Kim, H. Wang, G. Ye, and S. Shin, "Decentralized control of autonomous swarm systems using artificial potential functions: analytical design guidelines," in *2004 43rd IEEE Conference on Decision and Control (CDC) (IEEE Cat. No.04CH37601)*. IEEE, 2004. [Online]. Available: <https://doi.org/10.1109/cdc.2004.1428623>
- [16] Z. Li, W. Xue, D. Li, D. Lin, and J. Huang, "Distributed differential game for formation control of multi-uav with obstacle avoidance," in *2022 China Automation Congress (CAC)*. IEEE, 2022, pp. 4877–4882.
- [17] H. S. Yavuz, H. Goktas, H. Cevikalp, and H. Saribas, "Optimal task allocation for multiple uavs," in *2020 28th Signal Processing and Communications Applications Conference (SIU)*. IEEE, 2020, pp. 1–4.
- [18] T. Wang, Q. Dang, and P. Pan, "A multi-robot system based on a hybrid communication approach," *Studies in Media and Communication*, vol. 1, no. 1, pp. 91–100, 2013.
- [19] M. A. Kamel, X. Yu, and Y. Zhang, "Formation control and coordination of multiple unmanned ground vehicles in normal and faulty situations: A review," *Annual reviews in control*, vol. 49, pp. 128–144, 2020.
- [20] Q. Ouyang, Z. Wu, Y. Cong, and Z. Wang, "Formation control of unmanned aerial vehicle swarms: A comprehensive review," *Asian Journal of Control*, vol. 25, no. 1, pp. 570–593, 2023.
- [21] N. H. Li and H. H. Liu, "Formation uav flight control using virtual structure and motion synchronization," in *2008 American Control Conference*. IEEE, 2008, pp. 1782–1787.
- [22] X.-p. Xu, X.-t. Yan, W.-y. Yang, K. An, W. Huang, and Y. Wang, "Algorithms and applications of intelligent swarm cooperative control: A comprehensive survey," *Progress in Aerospace Sciences*, vol. 135, p. 100869, 2022.
- [23] D. Xu, X. Zhang, Z. Zhu, C. Chen, P. Yang *et al.*, "Behavior-based formation control of swarm robots," *mathematical Problems in Engineering*, vol. 2014, 2014.
- [24] L. Joseph, *Robot operating system (ros) for absolute beginners*. Springer, 2018.
- [25] L. Joseph and J. Cacace, *Mastering ROS for Robotics Programming: Design, build, and simulate complex robots using the Robot Operating System*. Packt Publishing Ltd, 2018.
- [26] M. Quigley, K. Conley, B. Gerkey, J. Faust, T. Foote, J. Leibs, R. Wheeler, A. Y. Ng *et al.*, "Ros: an open-source robot operating system," in *ICRA workshop on open source software*, vol. 3, no. 3.2. Kobe, Japan, 2009, p. 5.
- [27] W. Giernacki, M. Skwierczyński, W. Witwicki, P. Wroński, and P. Kozierski, "Crazyflie 2.0 quadrotor as a platform for research and education in robotics and control engineering," in *2017 22nd International Conference on Methods and Models in Automation and Robotics (MMAR)*. IEEE, 2017, pp. 37–42.
- [28] J. A. Preiss, W. Honig, G. S. Sukhatme, and N. Ayanian, "Crazyswarm: A large nano-quadcopter swarm," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2017, pp. 3299–3304.
- [29] M. J. Hong and M. Arshad, "A balance-artificial potential field method for autonomous surface vessel navigation in unstructured riverine environment," *Procedia Computer Science*, vol. 76, pp. 198–202, 2015.



**Abdulmelik Bekmez** is currently studying in his bachelor's degree in the Department of Computer Engineering, Engineering Faculty, at the Fatih Sultan Mehmet Vakıf University. His current research areas include robotics and formation control.



**Kadir Aram** is research assistant in Computer Engineering department at Fatih Sultan Mehmet Vakıf University. He completed his PH.D. in Computer Science and Engineering at Istanbul Sabahattin Zaim University in 2023. His research interest are mobile robotics and natural language processing.



**Burhanettin Can** is Professor in Computer Engineering department at Fatih Sultan Mehmet Vakıf University. He completed his PH.D. Nuclear Technology at Istanbul Technical University in 1984. His research interest are robotics and fuzzy logic.



# Bayesian Optimization-based CNN Framework for Automated Detection of Brain Tumors

Mahir Kaya


**Abstract**— Brain tumors, capable of yielding fatal outcomes, can now be identified through MRI images. However, their heterogeneous nature introduces challenges and time-consuming aspects to manual detection. This study aims to design the optimal architecture, leveraging Convolutional Neural Networks (CNNs), for the automated categorization of brain tumor types within medical images. CNN architectures frequently face challenges of overfitting during the training phase, mainly attributed to the dual complexities of limited labeled datasets and complex models within the medical domain. The depth and width hyperparameters in these architectures perform a vital role, in determining the extent of learning parameters engaged in the learning process. These parameters, encompassing filter weights, fundamentally shape the performance of the model. In this context, it is quite difficult to manually determine the optimum depth and width hyperparameters due to many combinations. With Bayesian optimization and Gaussian process, we identified models with optimum architecture from hyperparameter combinations. We performed the training process with two different datasets. With the test data of dataset 1, we reached 98.01% accuracy and 98% F1 score values. With the test data of dataset 2, which has more data, 99.62% accuracy and F1 score values were obtained. The models we have derived will prove valuable to clinicians for the purpose of brain tumor detection.

**Index Terms**— Deep learning, CNN, Bayesian optimization, Brain tumor.

## I. INTRODUCTION

**B**RAIN TUMORS are fatal, and life span of patients can be quite short [1]. Magnetic resonance imaging (MRI) is the most well-known and successful tool for detecting and classifying brain tumors due to its aptitude for distinguishing between structure and tissue depending on contrast levels [2]. The detection and classification of brain tumor is often determined manually by the clinician; the duration of this process is quite long and sometimes can lead to erroneous results [3]. Early detection of such tumors can improve the effectiveness of the therapeutic process and increases the likelihood of long-term survival [4]. The inherent heterogeneity within brain tumor cells poses a hurdle in precisely classifying the tumor type, consequently hindering treatment planning.

**MAHİR KAYA**, is with Department of Computer Engineering, Faculty of Engineering and Architecture, Tokat Gaziosmanpaşa University, Tokat, Türkiye. (e-mail: mahir.kaya@gop.edu.tr)

 <https://orcid.org/0000-0001-9182-271X>

Manuscript received Jan 16, 2023; accepted October 4, 2023.

DOI: [10.17694/bajece.1346818](https://doi.org/10.17694/bajece.1346818)

The principal objective of this study is to correctly identify brain tumor types, thereby facilitating informed decisions regarding suitable treatment approaches.

Artificial intelligence applications are used in many areas such as cloud computing [5, 6] and computer-assisted disease diagnosis [7]. Convolutional Neural Networks (CNNs), which are a part of deep learning, demonstrate exceptional performance in end-to-end learning. During the training phase, they autonomously generate feature maps from input images and subsequently execute classification in the network's final stages. The network's parameters undergo updates based on error rates during the concluding phase [8]. One of the important contrasts between machine learning and deep learning models hinges upon the choice of features employed for classification. While traditional machine learning relies on manually crafted features, deep learning directly uncovers features from data, eliminating the need for external intervention [9, 10]. As a result, CNN models require an ample supply of images to achieve successful training outcomes. However, within the domain of healthcare, there's generally a shortage of labeled data. This shortage, combined with the intricate nature of CNN models, often leads to the occurrence of memorization issues during the training phase [11]. Identifying the optimal architecture (comprising depth and width) along with a suitable hyperparameter combination is a crucial challenge when aiming for successful learning within CNN models constrained by a limited dataset.

CNN architectures are typically composed of sequential convolutional, max-pooling, and fully connected layers. In the Convolution layer, feature maps are generated by applying various filters to the raw image. This process is iterated across subsequent convolutional layers. In later layers, a higher number of filters is commonly employed to yield more intricate feature maps. The max-pooling layer reduces the feature map dimensions. These two layers collaboratively conduct the essential task of feature extraction from raw images within CNN architectures [12]. During the training phase, the filter weights—located where the learning process occurs—are updated using the backpropagation algorithm at the outcome layer, based on the error rate [8]. These filter weights are initially set randomly. The fully connected layer is often designated as the classification layer. Depending on the class count, the last layer incorporates either sigmoid or softmax functions. Of paramount importance is the determination of the optimal count of convolutional layers, along with the quantity



and size of filters within each convolutional layer, to foster successful learning in CNN architectures.

Increasing the model's depth and width typically leads to two prevalent challenges. One of these issues pertains to the vanishing gradient problem, while the other revolves around the risk of overfitting the training dataset. The vanishing gradient problem can be mitigated through the integration of residual connections. Nonetheless, even when employing generalization techniques such as batch normalization, data augmentation, and dropout, identifying the point at which excessive learning initiates remains crucial. This can be discerned by monitoring loss/accuracy graphs during the training phase. In shallower models with reduced model depth and width, learning is hindered, resulting in underfitting and lower performance. A major concern to tackle is the process of identifying the most suitable depth and width hyperparameters within CNN models. Given the often extensive training period these models require, attempting all conceivable combinations to obtain results can be a time-intensive process. Addressing this, Bayesian optimization in conjunction with Gaussian processes offers an effective means to efficiently pinpoint the optimal hyperparameter combinations. Within this study, we identified the optimal CNN architecture to achieve accurate brain tumor type detection using two distinct datasets. Employing Bayesian optimization, we systematically arrived at optimal hyperparameter values by iteratively combining them with model accuracy data. The resulting models hold promise in aiding clinicians with brain tumor type detection.

The contributions of this study can be listed as follows:

- A novel and effective CNN model has been developed to detect brain tumor types.
- This study presents an analysis of the efficacy of different CNN architectures, encompassing a range of widths and depths, in the context of brain tumor type detection.
- The optimal hyperparameter combinations were identified using Bayesian optimization in conjunction with Gaussian processes.
- In the context of two separate datasets, it's notable that the models' performance sees improvements in instances characterized by a larger data volume.

The structure of this work is as follows: In the Related Works section, we provide a summary of previous research on brain tumor detection and classification. Section 3 delves into the specifics of the dataset features, the Bayesian optimization method, and the overall structure of the CNN model. Section 4 presents the outcomes of our proposed models and compares them with findings from other research. In the conclusion, we underscore the significance of this study, its broader contributions, and potential paths for future research.

## II. RELATED WORKS

Brain MR images are frequently used in research in a variety of fields, including tumor segmentation and tumor type

classification. The research on brain tumor classification can be divided into three categories: tumor detection, tumor type classification, and tumor detection and classification.

The first of these is the studies carried out to detect the presence of tumor in brain MR images. The datasets contain two classes, tumor and normal. Toğaçar et al. [13] proposed a model for identifying brain tumors and the study used a dataset of 253 MRI images. The proposed model employs the attention module as well as the hypercolumn technique. The attention module is used to detect important areas of the image, and the hypercolumn technique provides more effective feature selection by allowing data from each layer to be used in the final layer, according to the study. The proposed model reached a 96.05 percent accuracy in brain tumor detection. Using the data augmentation technique, the authors [14] augmented the unbalanced dataset (155 tumor, 98 normal) to include an equal number of data for the tumor (155) and normal (155) classes. In the study, features obtained by hypercolumn technique using pre-trained AlexNet and VGG16 architectures were reduced by the recursive feature elimination (RFE) method and classified by SVM. The models that were classified using 200 and 300 features yielded the best accuracy values in the study, 96.77%. Balamurugan and Gnanamanoharan [15] proposed a hybrid CNN model for the detection of brain tumors. The dataset used in the study comprises 271 tumor images and 98 non-tumor images. There are 173 images set aside for training, 50 for validation, and 30 for testing. A Laplacian Gaussian filter (LOG) was used for data preprocessing in the study, and a Fuzzy C Means with Gaussian mixture model (FCM-GMM) algorithm was used for brain tumor segmentation. 13 features were determined using the VGG-16 architecture as a feature extractor, and classification was performed using the proposed enhanced LuNET algorithm. The proposed model's performance metrics are as follows: accuracy (99.7%), sensitivity (98.2%), specificity (98.6%), precision (99.4), F-Score (98.2), and recall (99.8%).

The second category includes studies on brain tumor classification. The researcher employs various datasets containing three or four types of brain tumors. Deepak and Ameer [16] undertook a study with the objective of classifying three different brain tumor varieties. In the study, GoogleNet architecture was used with transfer learning method for the purpose of feature extraction from brain MRIs. In addition to softmax as a classifier, SVM and KNN algorithms have also been tried. In the study, the best accuracy rate of 98% was obtained with the KNN algorithm and 80% of the dataset was used as the training set. The model was trained using 70%, 50% and 25% of the dataset, and classifications were performed with SVM. It has been reported that shrinking the training dataset does not significantly affect performance. To classify brain tumors, Başaran [17] proposed a hybrid model. Gray level co-occurrence matrix (GLCM) and Local Binary Pattern (LBP) algorithms, as well as four CNN models: AlexNet, VGG16, EfficientNetB0, and ResNet50, were used to obtain the features. Artificial Bee Colony (ABC), Particle Swarm Optimization (PSO), and Genetic algorithms (GA) were implemented to reduce these properties, and SVM was used to classify them.

The highest accuracy obtained in the study was 98.22% with 5-fold cross validation and an 86%/14% training/test ratio. Ayadi et al. [18] presented a model designed for brain tumor classification and this model was tested using three datasets. The proposed CNN model comprises of 10 convolutional layers, a dense layer, and a classifier based on softmax. The study's best performance was obtained with the Adagrad optimization algorithm and 0.003 learning rate, training for 20 epochs, and batch size 16. The model was trained and tested using the Figshare dataset in the study. The tests were then repeated with two different datasets (Radiopaedia and REMBRANDT) to provide additional validation. The dataset accuracy values are as follows: dataset1 (94.74%), dataset2 (93.71%), and dataset3 (five sub-datasets: 100%, 97.22%, 97.02%, 88.86%, and 95.72%). Ait-Amou et al. [19] use Bayesian optimization to determine the proposed CNN model's hyperparameters. The study involved an examination of the classification of three brain tumor types, namely glioma, meningioma, and pituitary. Their base model includes 5 convolutional blocks, a dense block, and a classification block. The input size is 64\*64 pixels and softmax activation function was used for classification. Bayesian optimization was employed to identify the optimal dropout rate, number of dense nodes, activation function, batch size, and optimization algorithm that yielded the best results in the study. The best performance, according to the results, was 98.70%. Alhassan and Zainon [20] explored the effect of the activation function on brain tumor classification. Feature extraction from brain images was accomplished using the Histogram of Oriented Gradients (HOG) technique, and a classification model employing the Hard Swish-based ReLU activation function was introduced for brain tumor classification. Assessment of the suggested model's effectiveness was conducted using two different architectures (CNN and RNN), a selection of three activation functions (sigmoid, tanh, and Hard Swish-based ReLU), and four varied training and testing ratios (20:80, 40:60, 60:40, 80:20). The top accuracy score was obtained by the CNN model that utilized the Hard Swish-based ReLU activation function, in conjunction with an 80:20 ratio. The approach presented by Aurna et al. [21] involves a two-stage process for feature selection in brain tumor classification. A total of four datasets were used in the study, including three individual datasets and one obtained by combining them. The best feature extractors were identified among the set of five pre-trained models and a new model (Scratched CNN). In the first stage, the three best models (EfficientNet-B0, ResNet-50, and Scratched CNN) were combined in pairs, and the model pairs with the highest accuracy value were determined. In the second stage, features were acquired using these pairs of models. During the feature reduction phase, Principal Component Analysis was used. For the classification process, the performance of Softmax, SVM, RF, KNN, and AdaBoost algorithms was compared, and Softmax produced the best results. The proposed model for the combined dataset produced the best accuracy performance of 98.96%. Accuracy values for other data sets are 99.67 for dataset1, 98.16 for dataset2, and 99.76 for dataset3, respectively. Mehnatkesh et al. [10]

conducted a hyperparameter optimization study to classify brain tumors. First, the images' empty space was cropped, and data augmentation was implemented. The performance of seven different state-of-the-art models was then analyzed on the dataset, and the ResNet model with the best result was chosen. The optimization study yielded 99.02% classification success with the ResNet model and the Improved Ant Colony Optimization algorithm. Kazemi et al. [22] used the Figshare and TCIA datasets in their research to determine the classification of brain tumors in MRI scans. In the proposed model, AlexNet and VGG16 architectures are trained concurrently. SVM, KNN, and Decision Tree methods were used to choose the most essential characteristics derived from the two architectures. The Softmax classifier was used to predict the tumor class. They investigated the suggested model's performance in binary and multiclass classification. The best accuracy result with Figshare dataset is 99.14% in binary class and 98.78% in multi-class. Gomez-Guzman et al. [23] compared the performance of seven different CNN models for brain tumor classification. A 17-layer CNN model with four convolutional layers was proposed. In addition, six pre-trained architectures: ResNet50, MobileNetV2, Xception, InceptionV3, InceptionRes-NetV3, and EfficientNetB0 were used in the study. A dataset of 7023 MRI images with four classes: no-tumor, glioma, meningioma, and pituitary was used in the study. The best accuracy was 97.12% with the InceptionV3 model. Türkoğlu[24] proposed a four-stage hybrid system for brain tumor diagnosis. The images are enhanced by preprocessing first. Transfer learning method was then applied to DenseNet and AlexNet architectures and 4096 and 1000 deep features were obtained respectively. The most significant features were determined in the third stage through feature reduction using the MrMr algorithm on the combined features. The SVM algorithm was employed in the final stage to ascertain the class to which the tumors belong. The Bayesian Optimization Algorithm was implemented to optimize the SVM classifier's hyperparameters. The author used the figshare dataset, which included 3064 brain MRI images and three classes of brain tumors. The proposed model was tested by selecting eight different numbers (from 500 to 5000) of combined deep features and the best accuracy performance was 98.04% using 2500 features. In addition, the author conducted an extensive experimental study on feature extraction from CNN models, feature selection (MrMr), and the optimization of SVM hyperparameters using Bayesian optimization for classification. The features obtained from the CNN models are reduced by the MrMr method and then used as input for the machine learning model. The difference in our work is that with the Gaussian process-based Bayesian optimization algorithm, many hyperparameters in the CNN architecture such as the number of convolution layers, number and size of filters, learning rate, dropout rate, and optimizer were optimized. We optimized directly on the CNN architecture without the need for features extracted from CNN models and a two-stage training process.

In the final category, there are studies that both detect and classify brain tumors. Mondal and Shrivastava [25] conducted

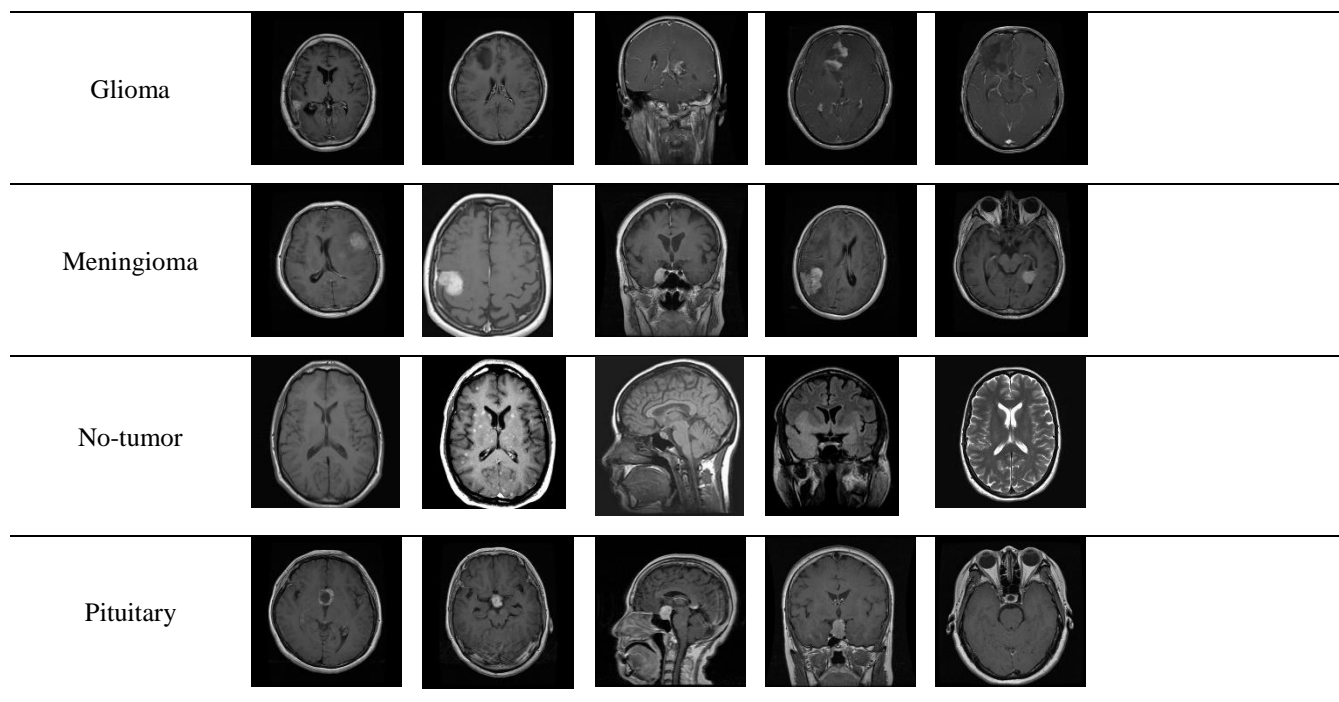


Fig. 1. Examples of three types of tumors and normal brain MR images

a study involving two separate datasets to identify and categorize brain tumors. They introduced a novel activation function named Parametric Flatten-p Mish (PFpM) alongside the BMRI-Net model. The model accomplished an accuracy of 99.00% in the detection of brain tumors and 99.57% in the classification of tumor type. Saurav et al. [26] devised a CNN model that employs channel-attention blocks to concentrate on pertinent areas within the image when classifying tumors. The selection of the pertinent feature maps is carried out via channel-attention blocks. The suggested model's performance was evaluated using four different datasets. The BT-small-2c and BT-large-2c datasets were utilized for tumor detection, and the other two were used for tumor categorization (BT-large-3c and BT-large-4c). According to the datasets, the following accuracy values were obtained: BTsmall-2c (96.08%), BT-large-2c (99.83%), BT-large-3c (97.23%), and BT-large-4c (95.71%). Turk et al. [27] proposed a system that detects and classifies brain tumors. The study was performed in three stages. First, brain tumor detection was performed with 2 classes (tumor and normal), then brain tumor classification was performed with 4 classes (glioma, meningioma, pituitary and normal) and finally Class Activation Maps were created. In the study, 3441 MR images for the first stage and 3362 MR images for the second stage from two different datasets were used. The transfer learning method was performed with the ensemble DL approach using ResNet50, VGG19, InceptionV3 and MobileNet architectures. The highest accuracy rate was 100% for brain tumor detection (with InceptionV3, MobileNet and ResNet50 architectures) and 96.45% for tumor classification (with ResNet50 architecture). Alanazi et al. [28] designed three scratch CNN models with 19, 22 and 25 layers to detect brain tumor and compared their performance. They obtained the best accuracy of 92.67% with the 22-layer CNN model. Then, they trained a model that detects the type of brain tumor using the

22-layer CNN model with a fine-tuning approach using the transfer learning. The test accuracy of this model is 95.75%. Kang et al. [29] proposed a feature ensemble-based model for brain tumor classification and investigated the performance of nine different ML classifiers. They used 13 different CNN architectures to extract features to be used in classification. The three models that provided the best features were determined using ML classifiers. DenseNet-169, ShuffleNet V2, and MnasNet provided the best features for four-class (glioma, meningioma, pituitary and normal) classification. The features from these three models were combined and fed into ML classifiers to identify brain tumor classes. The best accuracy for brain tumor diagnosis was 93.72% with SVM (RBF) for four-class classification and 98.83% for two-class classification.

Numerous studies in the literature have explored brain tumor detection. Transfer learning methods have been employed, typically yielding successful results in scenarios with limited datasets. However, current state-of-the-art CNN models tend to be intricate, as they are primarily tailored to vast datasets like ImageNet. In the context of medical images with limited labeled data [30], these models often grapple with issues such as overfitting or memorization during training. Despite the application of techniques like batch normalization, data augmentation, L2 regularization, and dropout to mitigate these challenges, they often fall short of being entirely effective. While the approach of utilizing CNN models for feature extraction followed by machine learning classification has become prevalent, it necessitates a two-stage training process

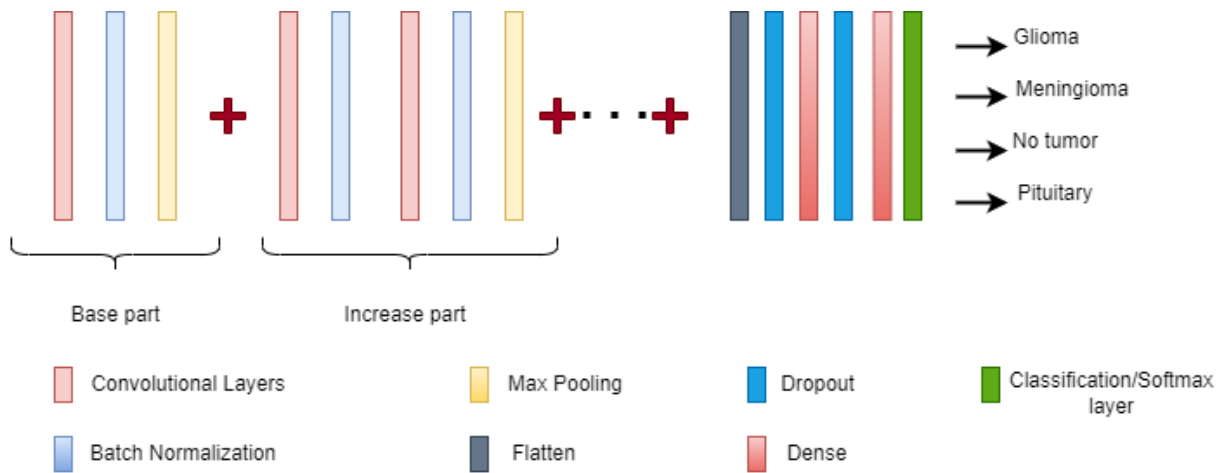


Fig. 2. Bayesian optimization-based proposed model structure

and can be demanding to implement more complexity. The key to success often lies in a well-designed CNN architecture, particularly in optimizing hyperparameters such as depth and width to suit the specific dataset. In this study, we addressed this challenge by determining the optimal depth and width hyperparameters for CNN architectures using Bayesian optimization.

### III. MATERIALS AND METHOD

#### A. Dataset

Two datasets were used in this study. Dataset-1 [31], a publicly accessible brain tumor dataset, has a total of 3264 brain MRIs. The dataset-1 has four classes: glioma, meningioma, pituitary and healthy (no tumor). This dataset contains 926 gliomas, 937 meningiomas, 901 pituitary, and 500 healthy images. Dataset-2 [32] represents a publicly available dataset containing brain tumors and has a total of 7023 brain MRIs. This dataset was created by merging three separate datasets (Figshare, SARTAJ, and Br35H). The dataset consists of four distinct classes. These are MR images of healthy, meningioma, pituitary, and glioma patients' brains. There are 2000 images of healthy people, 1621 gliomas, 1645 meningiomas, and 1757 pituitary tumors. The dataset divisions for training, validation, and testing are presented in Table 1. Fig. 1 shows examples of three types of tumors and normal brain MRI images from dataset 2.

TABLE I  
TRAIN, VALIDATION AND TEST PART OF DATASETS

	Train	Validation	Test
<b>Dataset 1<sup>1</sup></b>	2351	261	652
<b>Dataset 2<sup>2</sup></b>	5141	571	1311

<sup>1</sup>Sartaj dataset <https://www.kaggle.com/sartajbhuvaji/brain-tumor-classification-mri>

<sup>2</sup>Massoud dataset <https://www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset?select=Training>

#### B. Bayesian optimization

The optimization approach involves iteratively generating neural networks with defined hyperparameters, executing the

training process, and determining the optimum hyperparameter set among these constructed networks. The Sequential model-based optimization (SMBO) approach, that is Bayesian optimization, was utilized in this study to estimate hyperparameters and network design [33, 34].

Bayesian optimization is a technique based on probabilistic modeling, used to find the peaks or lowest points of objective functions that are costly to assess. It can be used when the goal function has no closed-form expression but observations of this function can be acquired at sampling values [35, 36]. An acquisition function is used within the context of Bayesian optimization to effectively select the next sampling location. It automates the balance between exploration and exploitation. Exploration takes place when there's uncertainty in the objective function, whereas exploitation centers on utilizing  $x$  values where the objective function is anticipated to be at its peak [35]. The primary goal of this optimization method is to reduce the count of evaluations for the objective function, making it advantageous. The Bayes theorem is employed to compute the posterior probability of an event, taking into account both the prior probability and the likelihood probability of the event. In cases where the objective function is uncertain, the Bayesian model provides an elegant approach to defining attributes of the objective function with the aid of informative priors, such as approximate locations of the maximum or its smoothness [33, 36].

In this research, we utilized the Gaussian Process (GP) method and gathered the initial  $D_i(x_i, y_i)$  data. Through the input  $X$ , we conducted training on the dataset to acquire the function output  $y$ . At any given point  $x$ , the value of  $f_x$  is treated as a stochastic variable. The random variables  $f_{x_i}$  and  $f_{x_j}$ , corresponding to distinct  $x_i$  and  $x_j$  points, exhibit correlation. To represent these random variables, we employed a Gaussian

distribution ( $f(x) \sim N(\mu(x), \sigma^2)$ ). Given that we conducted optimization for several hyperparameters, we employed the Gaussian Process (GP) as presented in Equation 1. The fundamental aspect of a GP is its function distribution, and this is completely characterized by the mean and covariance functions. The kernel function in Equation 2 was applied to a



set of hyperparameters with  $m$  dimensions. It quantifies the similarity between two distinct predictions. We employed the maximum likelihood estimate (MLE) approach, as described in Equation 3 [37], for hyperparameter selection. In this method, we evaluate the likelihood of the observations  $f(x_{1:n})$  with respect to the prior distribution,  $P(f(x_{1:n})|\varphi)$ , which follows a multivariate normal density. Here,  $\varphi$  represents the hyperparameter vector. Afterward, we estimate  $\varphi$  using the maximum a posteriori (MAP) estimation, which corresponds to the value of  $\varphi$  that maximizes the posterior distribution [37].

$$f_x \sim GP(m(x), k(x, x')) \quad (1)$$

$$k(x_i, x_j) = \exp\left(-\frac{1}{2}\|x_i - x_j\|^2\right) \quad (2)$$

$$\begin{aligned} \varphi^* &= \operatorname{argmax}_{\varphi} P(\varphi|f(x_{1:n})) \\ &= \operatorname{argmax}_{\varphi} P(f(x_{1:n})|\varphi)P(\varphi) \end{aligned} \quad (3)$$

$$\begin{aligned} EI(x) &= \begin{cases} (\mu(x) - f(x^*))\Phi(Z) + \sigma(x)\phi(Z) & \text{if } \sigma(x) > 0 \\ 0 & \text{if } \sigma(x) = 0 \end{cases} \end{aligned} \quad (4)$$

$$Z = \frac{\mu(x) - f(x^*)}{\sigma(x)}$$

Within Bayesian optimization, a critical step revolves around the criteria used to determine the upcoming collection of hyperparameters derived from the surrogate function. The Expected Improvement, provided in Equation 4, stands out as the most commonly adopted criterion [35]. By considering the expected value of the improvement function concerning the Gaussian process's predictive distribution, we achieve a harmonious equilibrium between exploration and exploitation. During exploration, our emphasis is on pinpointing locations characterized by substantial surrogate variance. Conversely, during exploitation, we direct our attention to points with elevated surrogate means [37].

### C. Proposed Method

In this study, we initially identified a fundamental block comprising convolutional, batch normalization, and max pooling layers. Following this foundational block, we introduced a max pooling layer after every 2 convolutional and batch normalization layers and elucidated the growth block structure for each model proposal. Fig. 2 shows the general structure of the proposed method.

Following activation functions in the convolutional layers, the batch normalization layer aids in achieving quicker convergence towards optimal values for the models, while also preventing overfitting during the training phase. In deep learning models, as the number of convolutional layers increases—leading to greater depth—it is anticipated that more intricate feature maps will be acquired. In light of this, we devised a block structure that incorporates max pooling after every two successive convolutional layers. By amplifying the count of these blocks, we effectively enhance the depth within the models. Increasing the depth of a model often gives rise to two primary challenges. The first involves the vanishing gradient, while the second pertains to the potential overfitting

of the training dataset. The vanishing gradient dilemma can be alleviated via the use of residual connections. However, even with generalization techniques like batch normalization in place, recognizing the juncture at which undue learning takes hold remains pivotal. This can be deduced by closely observing the loss/accuracy graphs during the training phase. In this study, we employ Bayesian optimization to identify the most suitable model depth-width and optimal hyperparameters from a range of possibilities.

## IV. EXPERIMENT AND RESULTS

CNN architectures automate the extraction and classification of features directly from input images, bypassing the need for manual feature extraction as required by traditional machine learning algorithms. Within CNN architectures, the training process involves the movement of each image through the network, and learning takes place as filter weights are updated using the backpropagation algorithm based on error rates at the output layer. Alongside this, various hyperparameters are defined for each training stage.

The determination of optimal hyperparameters, which maximize classification performance, stands as a significant concern. Hyperparameter value ranges are presented in Table 2. These ranges were established for the brain tumors dataset following numerous trial-and-error iterations.

TABLE II  
HYPERPARAMETERS AND VALUES

Hyperparameters	Values
Convolutional layer size	5, 7, 9, 11
Kernel size	3x3, 5x5
Filters size	min value :=16, max value :=256, step :=16
Dropout-rate	0.0, 0.2, 0.3, 0.4, 0.5, 0.6
Optimizer	Adam, SGD with Nesterov
Learning rate	0.001, 0.0001

CNN architectures commonly achieve favorable outcomes through the implementation of deep networks, allowing for detailed feature extraction. Nevertheless, in cases where the available labeled data is scarce, the training phase frequently encounters challenges related to memorization, thereby detrimentally impacting overall performance. Consequently, despite the models exhibiting high training success, their performance on unseen test datasets remains notably suboptimal.

As can be seen in Table 3, models with four different depths and filter numbers are reported on dataset 1. Out of the various configurations, the 9-conv-layer CNN architecture coupled with the hyperparameter settings in Model 3, determined through Bayesian optimization, yielded the most optimal results for dataset 1. When working with a constrained dataset size, elevating the model's depth can result in reduced performance on the test dataset, primarily due to the risk of memorization during the training phase. In Dataset 1, Model 3 exhibited the

most impressive performance, achieving a test accuracy of 98.01%.

TABLE III  
MODEL STRUCTURES BASED ON BAYESIAN OPTIMIZATION FOR DATASET 1

Layers	CNN Model 1	CNN Model 2	CNN Model 3	CNN Model 4
Conv-1	3x3, 16	3x3, 48	3x3, 112	3x3, 112
Conv-2	5x5, 16	3x3, 48	5x5, 112	3x3, 112
Conv-3	3x3, 176	5x5, 128	3x3, 112	3x3, 112
Conv-4	5x5, 256	3x3, 80	5x5, 112	5x5, 112
Conv-5	5x5, 208	3x3, 128	3x3, 240	3x3, 240
Conv-6	-	5x5, 256	3x3, 240	5x5, 240
Conv-7	-	5x5, 48	5x5, 240	3x3, 16
Conv-8	-	-	5x5, 144	5x5, 16
Conv-9	-	-	5x5, 48	5x5, 112
Conv-10	-	-	-	5x5, 48
Conv-11	-	-	-	3x3, 16
Dropout-rate-1	0,6	0,6	0,4	0
Dense-1	256	256	208	64
Dropout-rate-2	0	0	0	0
Dense-2	256	256	64	256
Optimizer	SGD with Nesterov	SGD with Nesterov	SGD with Nesterov	SGD with Nesterov
Learning-rate	0.001	0.0001	0.0001	0.001
Accuracy Score (%)	95.40	96.47	98.01	96.78

axa,b : a stands for kernel size and b stands for filter size in convolutional layers

The optimal model configuration for dataset 2 is illustrated in Fig. 3. In this dataset, the 9-conv-layer convolutional architecture demonstrated exceptional performance, achieving a test accuracy of 99.62%. Beyond Bayesian optimization, the notable success of this architecture can be attributed to the substantial increase in the volume of data. In CNN architectures, it's generally expected that the number of filters will be higher in the later layers to facilitate more intricate feature extraction. An intriguing observation we made using Bayesian optimization in both datasets is the decline in filter counts in the last layers. This reduction can be attributed to Bayesian optimization addressing instances of overfitting during the training phase, often associated with excessive filter numbers.

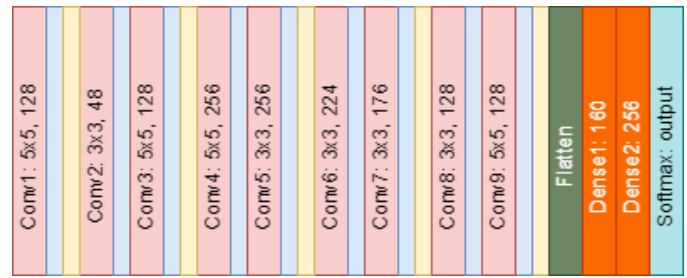
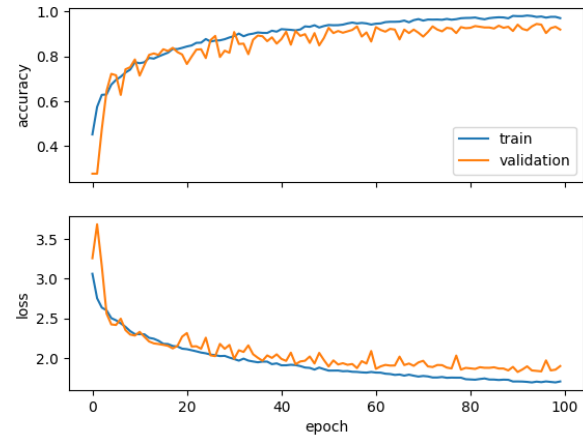
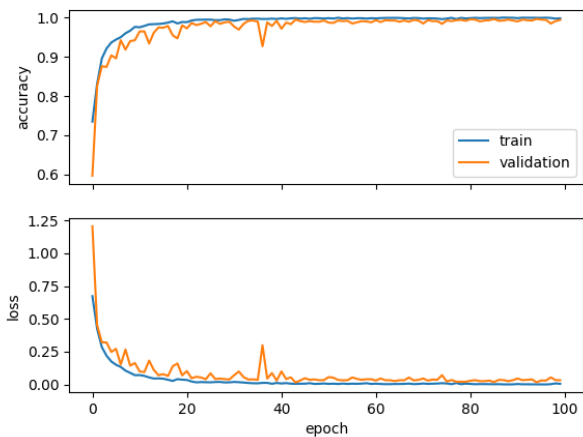


Fig. 3. Optimum model structure for dataset 2



(a)



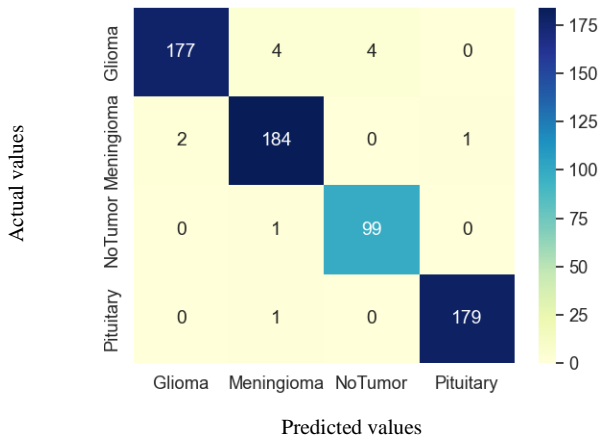
(b)

Fig. 4. Training and Validation accuracy/loss values for (a) dataset 1 (b) dataset 2

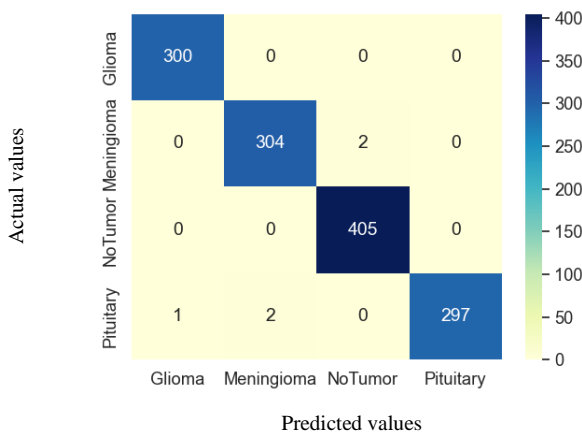
Fig. 4 depicts the training and validation loss/accuracy plots for both dataset 1 and dataset 2. From these graphs, it's evident that the models achieving optimal outcomes do not exhibit signs of memorization during the training phase. In cases of memorization, the training accuracy consistently improves while the validation accuracy tends to decline after reaching a certain epoch value. Similarly, during memorization, the training loss demonstrates a continuous decline in the loss graph, while the validation loss tends to rise after a specific epoch value. Across both graphs, a consistent trend of increase/decrease is observed in the training and loss graphs,

with instances of overlap at various points. Upon examining Fig. 4b, it becomes evident that due to the increased volume of data in dataset 2, the training and validation curves exhibit a noticeable overlap during the training phase.

Additionally, within the Meningioma column, 2 Pituitary images were erroneously labeled as Meningioma.



(a)



(b)

Fig. 5. Confusion matrix of (a) dataset 1 and (b) dataset 2

Fig. 5 presents the confusion matrix results generated by the best models on dataset 1 and dataset 2. Upon inspection of Figure 5a, it becomes evident that out of the total 185 glioma disease images, 177 were accurately predicted. Consequently, the True Positive count stands at 177. Furthermore, there were instances where 4 images, originally categorized as gliomas, were wrongly identified as meningiomas, while another 4 were classified as No tumor. These misclassifications amount to a total of 8 images, which constitute the False Negative instances. Moreover, examining the column corresponding to glioma, we find that 2 meningioma images were erroneously labeled as gliomas. These 2 instances contribute to the False Positive values within the column. Upon reviewing Fig. 5b, it's evident that all glioma images were accurately classified. Furthermore, upon closer inspection of the Meningioma category, it was revealed that 2 images that were supposed to be classified as meningioma were instead wrongly predicted as No tumor.

TABLE IV  
PERFORMANCE METRICS OF PROPOSED MODEL FOR DATASET 1

Classes	Precision	Recall	F1-Score	ICSI	Support
<b>Glioma</b>	0.9888	0.9568	0.9725	0.9456	185
<b>Meningioma</b>	0.9684	0.9840	0.9761	0.9524	187
<b>No-tumor</b>	0.9612	0.99	0.9754	0.9512	100
<b>Pituitary</b>	0.9944	0.9944	0.9944	0.9888	180
<b>Accuracy</b>	0.9801				652
<b>Macro avg</b>	0.9782	0.9813	0.9796		652
<b>Weighted avg</b>	0.9803	0.9801	0.98		652

TABLE V  
PERFORMANCE METRICS OF PROPOSED MODEL FOR DATASET 2

Classes	Precision	Recall	F1-Score	ICSI	Support
<b>Glioma</b>	0.9967	1.00	0.9983	0.9967	300
<b>Meningioma</b>	0.9935	0.9935	0.9935	0.987	306
<b>No-tumor</b>	0.9951	1.00	0.9975	0.9951	405
<b>Pituitary</b>	1.00	0.99	0.995	0.99	300
<b>Accuracy</b>	0.9962				1311
<b>Macro avg</b>	0.9963	0.9959	0.9961		1311
<b>Weighted avg</b>	0.9962	0.9962	0.9962		1311

TABLE VI  
PERFORMANCE METRICS OF PROPOSED MODELS FOR DATASET 1 AND DATASET 2

	MCC	Kappa	CSI
Proposed Model (Dataset 1)	0.9731	0.9730	0.9595
Proposed Model (Dataset 2)	0.9949	0.9949	0.9922

Performance metric results, derived from the confusion matrix data, are presented in Table 4, Table 5 and Table 6. Within medical image-based diagnostic systems, the computation of precision and recall values holds paramount importance. Precision denotes the proportion of True Positives to all Positives (TP/(TP+FP)), whereas recall measures the model's accurate recognition of True Positives (TP/(TP+FN)). The F1 score, representing the harmonic average of precision and recall, offers a comprehensive evaluation. It's worth noting that precision and recall values can exhibit disparities, particularly in scenarios of class imbalances within the dataset.

The performance metrics for each class in dataset 1 and dataset 2 are given in Table 4 and Table 5. The precision, recall, and F1-score values in Table 4 and Table 5 were calculated as described in the reference [38]. In addition, it was also verified with the scikit-learn library. Machine learning makes use of the Matthews Correlation Coefficient (MCC) as a statistical gauge to assess the accuracy of binary and multiclass classifications. The MCC assesses the model's overall performance by taking into account not just the accuracy of predictions but also the possibility of random agreement, which is particularly relevant in scenarios involving multiple classes. A higher MCC score signifies superior model performance across all classes, whereas a lower score indicates a weaker alignment between predictions and actual labels. When it comes to multiclass classification, the Kappa score enhances the evaluation of classification performance by taking chance agreement into consideration, making it especially advantageous in scenarios where classification models need assessment and there are multiple classes or imbalances in the dataset. In the context of classification assessment, the Individual Classification Success Index (ICSI) serves as a class-specific symmetric metric. Averaging the Individual Classification Success Index (ICSI) scores for all individual classes yields the Classification Success Index (CSI), offering a holistic assessment of the classification performance. A CSI value close to 1 indicates that the classification performance is very good [39]. MCC, Kappa, and Classification Success Index values for Dataset 1 and Dataset 2 are given in Table 6. In our study, the performance metrics showcased successful outcomes.

TABLE VII  
COMPARISON OF THE RECOMMENDED MODEL WITH EXISTING STUDIES

Study	Year	Accuracy (%)	F1-Score (%)
Mehnatkesh et al. [10] <sup>3</sup>	2023	98.69	98.46
Deepak&Ameer [16] <sup>3</sup>	2019	97.17	97.2
Turkoglu [24] <sup>3</sup>	2021	98.04	97.95*
Alanazi et al. [28] <sup>1</sup>	2022	95.75	95.72*
Saurav et al. [26] <sup>1</sup>	2022	95.71	95.98
Kang et al.[29] <sup>1</sup>	2021	93.72	-
Ayadi et al. [18] <sup>2</sup>	2021	98.49	98.3*
Aurna et al. [21] <sup>2</sup>	2022	98.96	99
Gomez-Guzman et al. [23] <sup>2</sup>	2023	97.12	97.28*
Proposed Model	Dataset-1 <sup>1</sup>	98.01	98
	Dataset-2 <sup>2</sup>	99.62	99.62

<sup>1</sup>Sartaj dataset <https://www.kaggle.com/sartajbhuvaji/brain-tumor-classification-mri>

<sup>2</sup>Masoud dataset <https://www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset?select=Training>

<sup>3</sup>Figshare dataset [https://figshare.com/articles/dataset/brain\\_tumor\\_dataset/1512427](https://figshare.com/articles/dataset/brain_tumor_dataset/1512427)

\* F1-score is calculated from the confusion matrix or precision-recall values.

In Table 7, you can find a comparison between our model and prior studies. Some studies [10, 16 and 24] used the Figshare dataset in their study, which involved a three-class classification task (Glioma, Meningioma, and Pituitary). In contrast, other studies used datasets that incorporated normal

brain MRI images (four-class classification). When we compare our proposed model with other studies, it becomes evident that our model performs better. Transfer learning and ensemble learning methods have been widely used in existing studies. In addition, CNN models are used for feature extraction, and the resulting feature dataset is trained on a machine-learning algorithm for classification. Aurna et al. [21] used the two-stage method of feature extraction. They obtained an accuracy of 98.96% on Dataset 2. Apart from these, there are studies that propose modified CNN models. Ayadi et al. [18] proposed a customized CNN model with 10 convolutional layers. They tried to find the best optimizer and learning rate manually. Gomez-Guzman et al. [23] proposed a 17-layer CNN model with four convolution layers. However, they found the best result using the InceptionV3 model with transfer learning. Their accuracy on Dataset 2 is 97.12%. In Table 7, our proposed model also gives better results when compared with the studies in Dataset 1. In our proposed CNN model, we directly optimize the depth, width, and other hyperparameters in the CNN architecture using Gaussian process-based Bayesian optimization.

## V. CONCLUSION

Brain tumors, a deadly type of cancer impacting both genders, have historically been diagnosed through risky biopsy procedures. Yet, the safer alternative of magnetic resonance imaging (MRI) has become more common in recent times. The main objective of this study is to differentiate brain tumor types from the MRI image, thereby guiding suitable treatment approaches. CNNs excel in autonomously extracting and classifying features from medical images, yielding favorable outcomes. Nevertheless, to ensure proficient learning within CNN architectures, a generous dataset is imperative. The focal issue revolves around pinpointing the most suitable depth and width hyperparameters that can facilitate optimal learning with the available limited data. Given the extended duration of the training phase and the considerable time investment required to explore all possible combinations, we adopted the Bayesian Optimization technique. This approach streamlined the process of identifying the optimal hyperparameter combinations. This study's focus was on determining the ideal hyperparameter configurations across two separate datasets. In Dataset 1, we reached an accuracy of 98.01% and an F1 score of 98%. In Dataset 2, our endeavors led to an impressive accuracy of 99.62%, also reflected in the F1 score. The model we have developed presents a valuable tool for clinicians, aiding in the precise identification of brain tumor types. As we look ahead, upcoming studies will explore diverse sets of hyperparameters and alternate datasets.

## REFERENCES

- [1] Siegel, R. L., Miller, K. D., & Jemal, A. (2019). Cancer statistics, 2019. *CA: a cancer journal for clinicians*, 69(1), 7-34.
- [2] Bhatlele, K. R., & Bhadauria, S. S. (2020). Brain structural disorders detection and classification approaches: a review. *Artificial Intelligence Review*, 53(5), 3349-3401.



- [3] Nazir, M., Shakil, S., & Khurshid, K. (2021). Role of deep learning in brain tumor detection and classification (2015 to 2020): A review. *Computerized Medical Imaging and Graphics*, 91, 101940.
- [4] Sharif, M. I., Li, J. P., Naz, J., & Rashid, I. (2020). A comprehensive review on multi-organs tumor detection based on machine learning. *Pattern Recognition Letters*, 131, 30-37.
- [5] Kaya, M. and Çetin-Kaya, Y. (2021). Seamless computation offloading for mobile applications using an online learning algorithm. *Computing*, vol. 103, no.5, pp. 771-799.
- [6] Miao, Y., Wu, G., Li, M., Ghoneim, A., Al-Rakhami, M., & Hossain, M. S. (2020). Intelligent task prediction and computation offloading based on mobile-edge cloud computing. *Future Generation Computer Systems*, 102, 925-931.
- [7] Rashed, A. E. E., Elmorsy, A. M., & Atwa, A. E. M. (2023). Comparative evaluation of automated machine learning techniques for breast cancer diagnosis. *Biomedical Signal Processing and Control*, 86, 105016.
- [8] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature*, 521(7553), 436-444.
- [9] Alzubaidi, L., Zhang, J., Humaidi, A.J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaria, J., Fadhel, M.A., Al-Amidie, M. and Farhan, L., (2021). Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. *Journal of big Data*, 8(1), pp.1-74.
- [10] Mehnatkesh, H., Jalali, S. M. J., Khosravi, A., & Nahavandi, S. (2023). An intelligent driven deep residual learning framework for brain tumor classification using MRI images. *Expert Systems with Applications*, 213, 119087.
- [11] Liu, Z., Tong, L., Chen, L., Jiang, Z., Zhou, F., Zhang, Q., ... & Zhou, H. (2023). Deep learning based brain tumor segmentation: a survey. *Complex & intelligent systems*, 9(1), 1001-1026.
- [12] Krizhevsky, A., Sutskever I. and Hinton G. E. (2012). ImageNet classification with deep convolutional neural networks," *Proc - Neural Information Processing System Conference*, pp. 1-9.
- [13] Toğaçar, M., Ergen, B., & Cömert, Z. (2020). BrainMRNet: Brain tumor detection using magnetic resonance images with a novel convolutional neural network model. *Medical hypotheses*, 134, 109531.
- [14] Toğaçar, M., Cömert, Z., & Ergen, B. (2020). Classification of brain MRI using hyper column technique with convolutional neural network and feature selection method. *Expert Systems with Applications*, 149, 113274.
- [15] Balamurugan, T., & Gnanamanoharan, E. (2023). Brain tumor segmentation and classification using hybrid deep CNN with LuNetClassifier. *Neural Computing and Applications*, 35(6), 4739-4753.
- [16] Deepak, S., & Ameer, P. M. (2019). Brain tumor classification using deep CNN features via transfer learning. *Computers in biology and medicine*, 111, 103345.
- [17] Başaran, E. (2022). A new brain tumor diagnostic model: Selection of textural feature extraction algorithms and convolution neural network features with optimization algorithms. *Computers in Biology and Medicine*, 148, 105857.
- [18] Ayadi, W., Elhamzi, W., Charfi, I., & Atri, M. (2021). Deep CNN for brain tumor classification. *Neural processing letters*, 53, 671-700.
- [19] Ait Amou, M., Xia, K., Kamhi, S., & Mouhafid, M. (2022). A Novel MRI Diagnosis Method for Brain Tumor Classification Based on CNN and Bayesian Optimization. In *Healthcare* (Vol. 10, No. 3, p. 494). MDPI.
- [20] Alhassan, A. M., & Zainon, W. M. N. W. (2021). Brain tumor classification in magnetic resonance image using hard swish-based RELU activation function-convolutional neural network. *Neural Computing and Applications*, 33, 9075-9087.
- [21] Aurna, N. F., Yousuf, M. A., Taher, K. A., Azad, A. K. M., & Moni, M. A. (2022). A classification of MRI brain tumor based on two stage feature level ensemble of deep CNN models. *Computers in biology and medicine*, 146, 105539.
- [22] Kazemi, A., Shiri, M. E., & Sheikhhamedi, A. (2022). Classifying tumor brain images using parallel deep learning algorithms. *Computers in Biology and Medicine*, 148, 105775.
- [23] Gómez-Guzmán, M.A., Jiménez-Beristain, L., García-Guerrero, E.E., López-Bonilla, O.R., Tamayo-Pérez, U.J., Esqueda-Elizondo, J.J., Palomino-Vizcaino, K. & Inzunza-González, E. (2023). Classifying brain tumors on magnetic resonance imaging by using convolutional neural networks. *Electronics*, 12, 955.
- [24] Türkoğlu, M. (2021). Brain Tumor Detection using a combination of Bayesian optimization based SVM classifier and fine-tuned based deep features. *Avrupa Bilim ve Teknoloji Dergisi*, (27), 251-258.
- [25] Mondal, A., & Shrivastava, V. K. (2022). A novel Parametric Flatten-p Mish activation function based deep CNN model for brain tumor classification. *Computers in Biology and Medicine*, 150, 106183.
- [26] Saurav, S., Sharma, A., Saini, R., & Singh, S. (2023). An attention-guided convolutional neural network for automated classification of brain tumor from MRI. *Neural Computing and Applications*, 35(3), 2541-2560.
- [27] Turk, O., Ozhan, D., Acar, E., Akinci, T. C., & Yilmaz, M. (2022). Automatic detection of brain tumors with the aid of ensemble deep learning architectures and class activation map indicators by employing magnetic resonance images. *Zeitschrift für Medizinische Physik*. <https://doi.org/10.1016/j.zemedi.2022.11.010>
- [28] Alanazi, M.F.; Ali, M.U.; Hussain, S.J.; Zafar, A.; Mohatram, M.; Irfan, M.; Albarrak, A.M. (2022). Brain tumor/mass classification framework using magnetic-resonance-imaging-based isolated and developed transfer deep-learning model. *Sensors*, 22, 372
- [29] Kang, J., Ullah, Z., & Gwak, J. (2021). Mri-based brain tumor classification using ensemble of deep features and machine learning classifiers. *Sensors*, 21(6), 2222.
- [30] Nergiz, M. (2023). Classification of Precancerous Colorectal Lesions via ConvNeXt on Histopathological Images. *Balkan Journal of Electrical and Computer Engineering*, 11(2), 129-137.
- [31] Sartaj Bhubvaji, Brain tumor classification (MRI). <https://www.kaggle.com/sartajbhubvaji/brain-tumor-classification-mri>, 2020. (Accessed 1 Jan 2023).
- [32] Masoud Nickparvar, Brain Tumor MRI Dataset. <https://www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset?select=Training> (Accessed 5 Jan 2023).
- [33] Fernandes, V., Junior, G. B., de Paiva, A. C., Silva, A. C., & Gattass, M. (2021). Bayesian convolutional neural network estimation for pediatric pneumonia detection and diagnosis. *Computer Methods and Programs in Biomedicine*, 208, 106259.
- [34] Pelikan, M., Goldberg, D. E., & Cantú-Paz, E. (1999). BOA: The Bayesian optimization algorithm. In *Proceedings of the genetic and evolutionary computation conference GECCO-99* (Vol. 1, No. 1999).
- [35] Brochu, E., Cora, V. M., & De Freitas, N. (2010). A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning. *arXiv preprint arXiv:1012.2599*.
- [36] Bergstra, J., Bardet, R., Bengio, Y., & Kégl, B. (2011). Algorithms for hyper-parameter optimization. *Advances in neural information processing systems*, 24.
- [37] Frazier, P. I. (2018). A tutorial on Bayesian optimization. *arXiv preprint arXiv:1807.02811*.
- [38] Kaya, M., Ulutürk, S., Çetin Kaya, Y., Altıntaş, O., & Turan, B. (2023). Optimization of Several Deep CNN Models for Waste Classification. *Ahmet ZENGİN, Sakarya University, Türkiye*, [azengin@sakarya.edu.tr](mailto:azengin@sakarya.edu.tr), 6(2), 91.
- [39] Koukoulas, S., & Blackburn, G. A. (2001). Introducing new indices for accuracy evaluation of classified images representing semi-natural woodland environments. *Photogrammetric Engineering and Remote Sensing*, 67(4), 499-510.

## BIOGRAPHIES



**MAHİR KAYA**, He graduated from the Industrial Engineering Department of İstanbul Technical University in 2000. He received his M.S and Ph.D. degrees in 2010 and 2016, respectively from the Department of Information Systems at Middle East Technical University. His research field is machine learning, deep learning, mobile cloud computing,

and optimization. Currently, he works as an Assistant Professor in the Department of Computer Engineering at Tokat Gaziosmanpaşa University.

# 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: October 2023  
Volume: 11  
Issue: 4

## CONTENTS

Research Article	İsa Ataş, Cüneyt Özdemir, Musa Ataş, Yahya Doğan; Forensic Dental Age Estimation Using Modified Deep Learning Neural Network, ..... 298-305
Research Article	Mustafa Can Bingol, Ömür Aydoğmuş; Development of a Human-Robot Interaction System for Industrial Applications, ..... 306-315
Research Article	Abdülkadir Albayrak; Vision Transformer Based Photo Capturing System, ..... 316-321
Research Article	Kadir Yılmaz, Taner Dindar, Murat Ayaz, Serkan Aktaş, Serkan Sezen; Finite Element Analysis of SMC Core Magnetic Gear for Vehicle Powertrain Systems, .....322-328
Research Article	Yasin Bektaş; African Vultures Optimization Algorithm-Based Selective Harmonic Elimination for Multi-level Inverter, ..... 329-339
Research Article	Cüneyt Özdemir; Designing Effective Models for COVID-19 Diagnosis through Transfer Learning and Interlayer Visualization, ..... 340-345
Research Article	Olayemi Olaniyi, Shefiu Ganiyu; Intelligent Video Surveillance System Using Faster Regional Convolutional Neural Networks, ..... 346-351
Research Article	Cihan Bilge Kayasandık; Automated Detection of Collagen Bundles in Second Harmonic Generation Microscopy Images, ..... 352-363
Research Article	Ahmet Saim Yılmaz, Haydar Cukurtepe, Emin Kugu; Geo-Location Spoofing on E-Scooters; Threat Analysis and Prevention Framework, ..... 364-372
Research Article	Övül Eski, Kemal Şahin, Sevilay Çetin; A Modular Multilevel Converter-Based Pulsed Electric Field Generator Design for Electroporation Applications, ..... 373-379
Research Article	Çiğdem Bakır; Different Induction Motor Faults by New Proposed Random Forest Method, ..... 380-386
Research Article	Abdülmelik Bekmez, Kadir Aram; Three Dimensional Formation Control of Unmanned Aerial Vehicles in Obstacle Environments, ..... 387-394
Research Article	Mahir Kaya; Bayesian Optimization-based CNN Framework for Automated Detection of Brain Tumors, ..... 395-402

# 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](mailto:bajece@hotmail.com)

