Automated Gesture Recognition Using Zebra Optimization Algorithm with Deep Learning Model for Visually Challenged People

Mohammed Basheri*[1]

[1]Information Technology Department, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah 21589, Saudi Arabia
Emails: mbasheri@kau.edu.sa

Abstract

Gesture recognition for visually challenged people plays a vital role in improving their convenience and interaction with digital gadgets and environments. It includes improvement of systems that permit them to relate with digital devices by using hand actions or gestures. To improve user-friendliness, these systems select in-built and effortlessly learnable gestures, often integrating wearable devices prepared with sensors for precise detection. Incorporating auditory or haptic feedback devices offers real-time cues about achievement of familiar gestures. Machine learning (ML) and deep learning (DL) methods are useful tools for accurate gesture detection, with customization choices to accommodate individual preferences. In this view, this article concentrates on design and development of Automated Gesture Recognition using Zebra Optimization Algorithm with Deep Learning (AGR-ZOADL) model for Visually Challenged People. The AGR-ZOADL technique aims to recognize the gestures to aid visually challenged people. In the AGR-ZOADL technique, the primary level of data pre-processing is involved by median filtering (MF). Besides, the AGR-ZOADL technique applies NASNet model to learn complex features from the preprocessed data. To enhance performance of NASNet technique, ZOA based hyperparameter procedure performed. For gesture recognition process, stacked long short term memory (SLSTM) model is applied. The performance validation of AGR-ZOADL technique carried out using a benchmark dataset. The experimental values stated that AGR-ZOADL methodology extents significant performance over other present approaches

Keywords: Gesture Recognition; Visually Challenged People; Deep Learning; Zebra Optimization Algorithm; Artificial Intelligence

1. Introduction

The communication between humans and computers is extensively improved but the realm is noticing constant expansion, with the innovative techniques acquired and approaches determined. Hand gesture recognition (HGR) is a major progressive field where computer vision (CV) and artificial intelligence (AI) are supported for enhancing communication with deaf persons, however, to help gesture-based signalling systems [1]. Sub-sectors of HGR comprise sign language identification, and special signal language detection has been employed in human activity identification, posture and pose detection, sports, physical exercise monitoring, and controlling smart home or assistive living utilization with HGR [2]. Recently, computer technologists have employed diverse computation techniques and approaches to support resolve our difficulties even though they assist our lives [3]. The exploitation of hand gestures in different software applications has been provided to increase interaction between human and computer [4]. The development of gesture recognition techniques performs an significant part in expansion of computer and human interaction, and HGR applications in diverse fields are highly developing as popular [5]. The application of hand gestures are implemented in cognitive development evaluation, augmented reality and virtual reality, assisted living, games, and so on. The current growth of HGR in several domains is taken into consideration of industry as well as human-robot interaction in manufacturing as well as controlling autonomous cars [6].

Doi: https://doi.org/10.54216/FPA.160102
Received: July 29, 2023 Revised: November 17, 2023 Accepted: April 12, 2024

Major perspectives where HGR plays a crucial part in domain of assistive technologies for persons with visual disability but proficient user communication scheme can be of great significance [7]. Any applications and devices in this domain might be prominently advantageous from natural, agile and spontaneous interaction methods, which utilize hand gestures [8]. Instances of such devices namely OrCam MyEye state text and recognize objects in image, or eyewear object identification devices that support persons with visual disabilities in a marketplace environment. It has also mobile uses like SuperVision for Cardboard that transform a smartphone and Google Cardboard devices into lower-cost electronic glasses [9]. But, these techniques have been confined to a highly precise activity, demanding the user to switch or press the application to execute other tasks. Thus, a hand gesture-based interaction performs a main function to enhance these technologies [10]. There are many kinds of research under HGR as the field has been broadly increasing, and several applications comprising both deep learning (DL) and machine learning (ML) techniques targeting to detection of a gesture that must be accomplished by a human hand.

This article mainly concentrates on design and development of Automated Gesture Recognition using Zebra Optimization Algorithm with Deep Learning (AGR-ZOADL) model for Visually Challenged People. The AGR-ZOADL technique aims to recognize the gestures to aid visually challenged people. In the AGR-ZOADL technique, the primary level of data pre-processing is involved by median filtering (MF). Besides, the AGR-ZOADL technique applies NASNet model to learn complex features from the preprocessed data. To enhance performance of NASNet technique, ZOA based hyperparameter procedure performed. For gesture recognition process, stacked long short term memory (SLSTM) model applied. The performance validation of AGR-ZOADL technique carried out by employing a benchmark dataset.

2. Related works

Subudhi et al. [11] introduced an automatic HGR system by employing DL framework. The presented structure categorizes input hand gestures, each one signified by feature vectors into pre-defined amount of gesture classes. DL technique employed in order to identify hand gestures. The hand gestures images kept in a database and then pre-processed to acquire into appropriate method for processing. Then, histogram-oriented gradient (HOG) feature vector removed from pre-processed gesture image. Adithya and Rajesh [12] propose an effective technique for HGR, which is a main part of sign language vocabulary, depending on effectual deep convolutional neural networks (CNNs) framework.

Mujahid et al. [13] developed a lightweight technique that depends on YOLOv3 and DarkNet-53 CNNs for gesture detection without further pre-processing, enhancement of images and image filtering. The projected technique attained high accurateness even in difficult atmospheres, and it effectively identified gestures in lower-resolution image method. Alashhab et al. [14] developed a shared method for mobile devices managed by hand gestures which permit consumer to control device and utilize numerous help tools by creating easy static as well as dynamic hand gestures. The system mainly depends on a multi-head NN, which originally discovers and classifies hand gesture, and then, executes a 2nd phase which implements consistent action.

Adeel et al. [15] developed a gesture-based confidence assessment (GCA) method for gesture detection for recognizing mental conditions from their hand actions at time of interview. Additionally, this method is beneficial for visually decreased persons, whether their behavior or seem for an interview. GCA model makes audio alerts regarding confidence level of objective consumers. In past, there was no effort to identify a person’s psychological state by employing hand gestures. This technique completely relies on CNN with LSTM in order to capture temporal data. Can et al. [16] goal is to project a DL CNN technique which can able to categorize hand gestures efficiently from diagnosis of colored natural images and near infrared. This research paper develops a novel DL technique that depends on CNN for recognizing hand gestures, enhancing detection rate, training as well as test time. The presented technique contains data augmentation to increase training.

Ryumin et al. [17] proposed a dual deep neural networks (DNNs) based model framework such as AVSR and gesture detection. The chief innovation concerning audio-visual talking classification lies in finetuning plans for visual as well as acoustic features and developed end-to-end technique, which reflects 3 modality fusion models namely feature level, model level and prediction level. Parvathy et al. [18], vision based HGR method developed by using ML. The proposed method trained as well as tested by employing Sebastian Marcel's static hand posture database that obtainable in online. Modified Speed Robust Feature extraction (SRFE) and Discrete wavelet transform (DWT) techniques were utilized in order to remove revolution and then scale invariant key descriptor. Lastly, Bag of Word (BoW) model applied to improve secure input dimension vector which need for SVM.

Doi: https://doi.org/10.54216/FPA.160102
Received: July 29, 2023 Revised: November 17, 2023 Accepted: April 12, 2024

3. The proposed method

In this study, we have designed the AGR-ZOADL model for Visually Challenged People. The AGR-ZOADL technique aims to recognize the gestures to aid visually challenged people. It contains four major processes involved as MF MF-based preprocessing, NASNet based feature extraction, ZOA based hyperparameter, and SLSTM based classification. Fig. 1 illustrates entire process of AGR-ZOADL model.

A. Image Preprocessing

At primary level, data pre-processing is involved by MF. MF is commonly employed image processing method in order to decrease noise and improve image quality [19]. Working on a pixel-by-pixel base, this model includes swapping every pixel's strength value with median value of its neighboring pixels. When compared to traditional linear filters, median filtering is very effective in conserving edges and fine particulars, making it suitable for situations where salt-and-pepper noise exists. By choosing median value, filter efficiently reduces impact of outlier intensity values, donating to smoothing of image while upholding its general structural integrity.

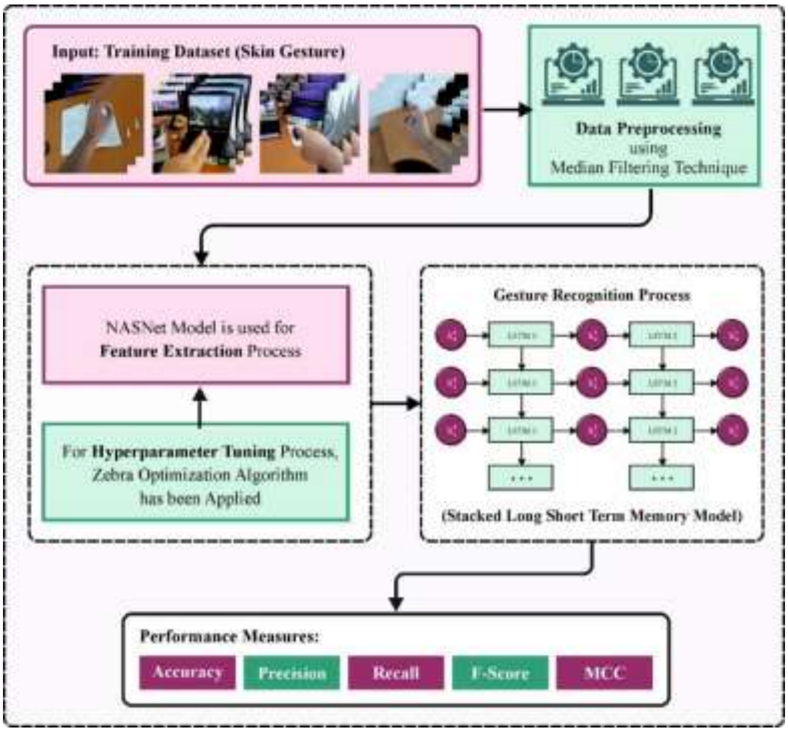

Figure 1: Overall process of AGR-ZOADL model

B. NASNet based feature extraction

Here, AGR-ZOADL technique applies NASNet model to learn complex features from the preprocessed data. NASNet belongs to the family of DNNs developed by Google, with the ability to select and search best-fit CNN architecture by using the concept of Reinforcement Learning (RL) [20]. The objective is to find best mixture of parameters namely strides, layer numbers, kernel size, etc, in certain search range.

In this method, the conception of NAS is motivated by NASNet model. NAS exploits a special RNN (controller)-based sample network contains different architectures. NAS algorithm works by considering that general framework of network is traditional and focuses mostly on accurateness to impose network performance. Each sample neural network is equipped with capability to adapt accurateness on a predetermined validation set that later utilized to update the controller for determining the fittest network architecture.

Cells in network are defined by RL-based technique though the general architecture of NASNet is predefined. The usual cell is a convolution operation block that often provides a feature map without reducing dimension or size. The reduction block is also convolution operation layer provides a reduced-sized feature map.

The working process of controller is given below.

Step1: Select a state amongst dissimilar hidden states $(h_i, h_{i-1}, etc.)$ generated by prior block.

Doi: https://doi.org/10.54216/FPA.160102
Received: July 29, 2023 Revised: November 17, 2023 Accepted: April 12, 2024

Step 2: Follow step 1, and select multiple hidden states.

Step 3: Select and use a suitable operation to state defined in step1.

Step 4: Follow step3, and implement a similar operation for state defined in step.

Step 5: Select a suitable architecture to combine outcomes of step3 and step4.

The set of possible operations comprises $3\times3/5\times5$ max-pooling, $33\times3/5\times5/7\times7$ depthwise convolution, $1\times7/7\times1/1\times1/3\times3$ convolutions, etc. Concretely, NAS- controller follows a LSTM based technique consists of 100 hidden blocks in all layers. It approximately makes 25B softmax predictions in LSTM for 2 Conv blocks (B is considered as 5). The popular versions of NASNet are NASNet-A, NASNet-B, and NASNet- C. In this study, NASNet-A and NASNet-C have been implemented.

C. Hyperparameter using ZOA

To improve performance of NASNet model, ZOA based hyperparameter process is performed. Zebras are type of horse inherent to Africa's east as well as south areas [21]. This type of animal is famous for its lined design of white and black hair that shields its full body. These lines frequently set in a upright design on zebra's body and neck, and they aid dual reasons such as defending zebras from existence seen by possible hunters and preventing sharp flies. The subsequent is a list of circumstances and their features. The main key source for foraging and protective strategies near attackers. At time of hunting for food, zebra lead makes it potential for enduring packs to track in its paths so they acquire nearer to food sources. As an outcome, pack in crowd travel over grassland by subsequent this discoverer zebra's principal.

ZOA technique is a populace based model and zebras create up a portion of that populace. When seen from scientific viewpoint, every zebra signifies probable response to problem and situated denotes search space for issue. A preliminary place of zebras in hunt space defined by a arbitrary procedure. Eq. (1) covers parameters that determine ZOA population matrix.

$$P = \begin{bmatrix} P_1 \\ \vdots \\ P_i \\ \vdots \\ P_N \end{bmatrix}_{N\times m} = \begin{bmatrix} p_{1,1} & \cdots & p_{1,j} & \cdots & p_{1,m} \\ \vdots & \cdots & \vdots & . & \vdots \\ p_{i,1} & \cdots & p_{i,j} & \cdots & p_{i,m} \\ \vdots & . & \vdots & . & \vdots \\ p_{N,1} & \cdots & p_{N,j} & \cdots & p_{N,m} \end{bmatrix}_{N\times m} \tag{1}$$

Whereas P signifies zebra population, P_i means ith zebra candidate, $p_{i,j}$ means j^{th} problem variable proposed by i^{th} zebra candidates, N represents amount of search agents, and m designates quantity of variables be adjusted. Every zebra views a possible response to optimizer problem. As an outcome, we evaluate fitness function by equating recommended solutions from every zebra. The values of fitness function defined by Eq. (2).

$$F = \begin{bmatrix} F_1 \\ \vdots \\ F_i \\ \vdots \\ F_N \end{bmatrix}_{N\times 1} = \begin{bmatrix} F(P_1) \\ \vdots \\ F(P_i) \\ \vdots \\ F(P_N) \end{bmatrix}_{N\times 1} \tag{2}$$

Whereas F signifies a column vector that contains candidates for fitness function (FF), and F_i denotes value that defined for i^{th} zebra. By evaluating fitness function candidates, one capable of observing excellence of candidate solutions properly that equal to issue and define which solution is optimal. When commerce with connecting minimization problems, zebra where FF has lowermost value considered as optimal candidate solution. These actions are mentioned below:

Foraging action.

Anti-predators' defensive methods.

In the first phase, individuals in population acquire upgrades by employing models of zebra action during foraging. Zebras acquire bulk of their food from sedges and grasses, but while nutrients are little, they only consume other plant parts such as fruits, buds, bark as well as leaves. This animal can pledge 60 to 80 percentage of their time consumption depending on excellence as well as flora. The grasslands zebra makes situations for animals that need smaller and more grasslands by overwhelming a shelter of higher and fewer healthful grasses. Zebras' location upgrades across foraging time by employing Eqs. (3) and (4) as follows.

Doi: https://doi.org/10.54216/FPA.160102
Received: July 29, 2023 Revised: November 17, 2023 Accepted: April 12, 2024

$$p_{i,j}^{new,S1} = p_{i,j} + r \cdot (ZL_j - I \cdot p_{i,j}) \tag{3}$$

$$P_i = \begin{cases} P_i^{new,S1}, \; F_i^{new,S1} < F_i \\ \\ P_i, \quad else \end{cases} \tag{4}$$

Whereas $P_i^{new,S1}$ signifies upgrade of ith zebra affording to 1st stage, $p_{i,j}^{new,S1}$ means its jth dimension value, $F_i^{new,S1}$ denotes its FF, ZL signifies zebra leader, ZL_j refers to its j^{th} dimension, r portrays random value from 0 to 1, $I = round\ (1 + rand)$, where rand denotes random value from 0 to 1. As an outcome, value of I can be any 1 or 2, and if value of I is equivalent to 2, then excessive contract changes in populace flexibility.

Next, ZOA populace owns locations in hunt space are upgraded by utilizing simulations of zebras' protection strategies against attackers. Lions are considered as main hunters of zebras. Zebras run the danger of flattering targets for crocodiles. Therefore, finest technique for zebra to protect itself from a lion attack is to escape by employing a zigzag running method and a full rapidity unexpected turns. Once challenged with attacks by tiny-sized enemies such as canines and hyaenas who confuse and frighten target by gathering, zebras answer with a greater level of violence.

When zebras attacked by lions, this method monitors them to escape from region they located to prevent additional danger. This permits them to save from presence consumed by lions. Eq. (5) signifies this method from a measured viewpoint.

$$p_{i,j}^{new,S2} = p_{i,j} + R \cdot (2r - 1) \cdot \left(1 - \frac{r}{T}\right) \cdot p_{i,j}, \; P_s \leq 0.5 \tag{5}$$

Whereas $p_{i,j}^{new,S2}$ means jth dimension value affording to 2nd phase, t signifies current iteration, T portrays maximal amount of iterations, R represents a constant value of 0.01, and P_s symbolizes prospect of choosing this approach that is arbitrarily selected to take a value among 0 and unity.

When one zebra in a crowd attacked by a starving animal, others will go nearer and make struggle to construct a difficulty of defence in order to frighten as well as fabricate attacker. This zebra method signified arithmetically by Eq. (6) below. When crowd positions upgraded, a zebra's upgraded location decides if it profits an enhanced outcome for fitness function. Eq. (7) denotes this upgrading principle.

$$p_{i,j}^{new,S2} = p_{i,j} + r \cdot (AZ_j - I \cdot p_{i,j}), P_s > 0.5 \tag{6}$$

$$P_i = \begin{cases} P_i^{new,S2}, \; F_i^{new,S2} < F_i \\ \\ P_i, \quad else \end{cases} \tag{7}$$

Whereas $P_i^{new_r S2}$ signifies upgrade of ith zebra linked to 2nd phase, $F_i^{new,S2}$ portrays its FF value, AZ denotes condition of criticized zebra, and AZ_j represents its jth dimension value. Fig. 2 demonstrates steps involved in ZOA.

Doi: https://doi.org/10.54216/FPA.160102
Received: July 29, 2023 Revised: November 17, 2023 Accepted: April 12, 2024

Figure 2: Steps involved in ZOA

The ZOA method originates a FF to obtain better classifier accuracy. It regulates a positive integer to signify superior act of candidate solutions. Here, minimization of detection error rate measured as FF provided in Eq. (8).

$$fitness(x_i) = ClassifierErrorRate(x_i)$$

$$= \frac{No.of\ misclassified\ samples}{Total\ No.of\ samples} * 100 \tag{8}$$

D. Gesture detection using SLSTM model

Eventually, the SLSTM methodology can be implemented. LSTM technique is a specific recurrent neural networks (RNNs) that can solve issue of long-term reliance in time sequence as well as gradient withdrawal and disappearing of RNN in long series training procedure [22]. The storage part of LSTM's grip link among long and short-term time sequence. This method capable of upgrading, preserve as well as remove data in storage unit via 3 gates namely forget, input, and output. The gate plan holds data for a long time as well as manage information movement. When evaluated to usual RNNs, LSTM executes superior in longer orders. The LSTM technique executed by employing MATLAB $2022b$ version deep learning toolbox. The detailed steps of the LSTM technique mentioned below:

Step1. The forget gate reads information of h_{t-1} and x_t, as well as selects whether to preserve data of earlier time via sigmoid function (σ) :

$$f_t = \sigma\left(a_f \cdot [h_{t-1}, x_t] + b_f\right) \tag{9}$$

Whereas f_t refers forget gate function, a_f signifies weight, x_t is input at time t, h_{t-1} is earlier intended output, b_f denotes bias of forget gate.

Step2. The input gate adopts data kept in cell state. Primary, define upgrade value i_t by sigmoid function and produce novel candidate value $\tilde{C}_t$ by tanh function:

$$i_t = \sigma(a_i \cdot [h_{t-1}, x_t] + b_i) \tag{10}$$

$$\tilde{C}_t = \tanh(a_c \cdot [h_{t-1}, x_t] + b_c) \tag{11}$$

Whereas a_i and a_c denotes weights, b_i and b_c refers biases.

Step3. Upgrade old cell state via forget and input gates for producing an upgraded value:

$$C_t = f_t x C_{t-1} + i_t x \tilde{C}_t \tag{12}$$

Doi: https://doi.org/10.54216/FPA.160102
Received: July 29, 2023 Revised: November 17, 2023 Accepted: April 12, 2024

Step4. The output gate consequences depend on cell state. Primary, define output value 0_t of cell state by sigmoid function. An upgrade value C_t refers normalization by $\tan h$ function and increased with Q, and o_t attain result value of (h) at time t:

$$o_t = \sigma(a_o \cdot [h_{t-1}, x_t] + b_o) \tag{13}$$

$$h_t = o_t \times \mathrm{tann}h(C_t) \tag{14}$$

Whereas a_o and b_o are said to be weight and bias. SLSTM architecture is a DL technique that influences numerous layers of LSTM elements to capture complex temporal addictions and hierarchical features in sequential data. Every LSTM layer contains memory cells and gates that permit network to selectively store and recover data over extended sequences, making it actual for errands concerning time-series data or natural language processing. Stacking many LSTM layers improves model's capability to learn difficult patterns by hierarchically removing and demonstrating features at dissimilar stages of concept.

4. Experimental Validation

In this part, gesture recognition outcomes of AGR-ZOADL model tested using gesture dataset [23] comprising 4000 samples with four classes as defined in Table 1.

Table 1 Details on dataset

Classes	No. of Samples
Point	1000
Drag	1000
Loupe	1000
Pinch	1000
Total Samples	4000

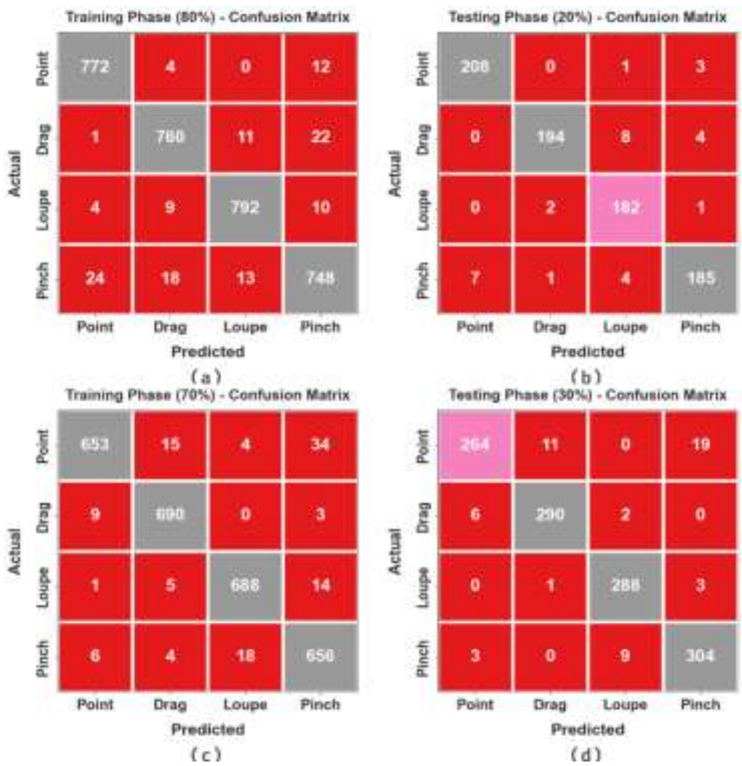

Figure 3: Confusion matrices of (a-b) 80:20 TRP/TSP and (c-d) 70:30 of TRP/TSP

Doi: https://doi.org/10.54216/FPA.160102
Received: July 29, 2023 Revised: November 17, 2023 Accepted: April 12, 2024

Fig. 3 represents confusion matrices accomplished by AGR-ZOADL system on 80:20 and 70:30 of training phase (TRP)/testing phase (TSP). The acquired outcomes show proficient recognition with 4 classes.

The gesture detection results of AGR-ZOADL technique with 80:20 of TRP/TSP are reported in Table 2 and Fig. 4. The outcome demonstrates that AGR-ZOADL method recognizes different kinds of gestures proficiently. According to 80% of TRP, the AGR-ZOADL algorithm provides an average $accu_y$ of 98%, $prec_n$ of 95.99%, $reca_l$ of 96%, F_{score} of 95.99%, and MCC of 94.66%. At same time, on 20% of TSP, the AGR-ZOADL technique provides an average $accu_y$ of 98.06%, $prec_n$ of 96.10%, $reca_l$ of 96.14%, F_{score} of 96.09%, and MCC of 94.83% correspondingly.

Table 2: Gesture detection outcomes of the AGR-ZOADL method on 80:20 TRP/TSP

Classes	$Accu_y$	$Prec_n$	$Reca_l$	F_{score}	MCC
TRP (80%)					
Point	98.59	96.38	97.97	97.17	96.24
Drag	97.97	96.08	95.72	95.90	94.55
Loupe	98.53	97.06	97.18	97.12	96.13
Pinch	96.91	94.44	93.15	93.79	91.74
Average	98.00	95.99	96.00	95.99	94.66
TSP (20%)					
Point	98.62	96.74	98.11	97.42	96.49
Drag	98.12	98.48	94.17	96.28	95.07
Loupe	98.00	93.33	98.38	95.79	94.54
Pinch	97.50	95.85	93.91	94.87	93.23
Average	98.06	96.10	96.14	96.09	94.83

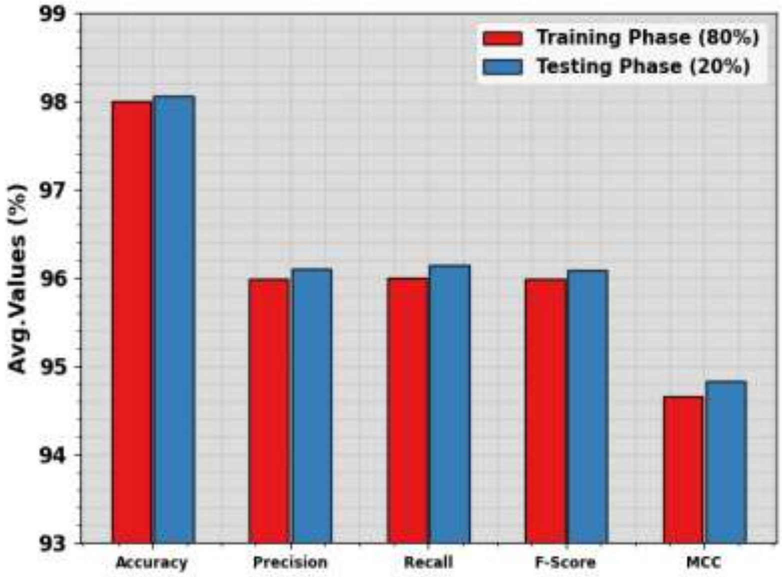

Figure 4: Average outcomes of the AGR-ZOADL model on 80:20 TRP/TSP

The gesture detection analysis of AGR-ZOADL technique with 70:30 of TRP/TSP can be examined in Table 3 and Fig. 5. The results displays that AGR-ZOADL system excellently recognizes diverse types of gestures. Based on 70% of TRP, AGR-ZOADL algorithm gives an average $accu_y$ of 97.98%, $prec_n$ of 95.98%, $reca_l$ of 95.97%,

Doi: https://doi.org/10.54216/FPA.160102
Received: July 29, 2023 Revised: November 17, 2023 Accepted: April 12, 2024

F_{score} of 95.95%, and MCC of 94.63%. Concurrently, with 30% of TSP, the AGR-ZOADL algorithm provides average $accu_y$ of 97.75%, $prec_n$ of 95.58%, $reca_l$ of 95.49%, F_{score} of 95.49%, and MCC of 94.03% correspondingly.

Table 3: Gesture detection analysis of the AGR-ZOADL approach on 70:30 TRP/TSP

Classes	$Accu_y$	$Prec_n$	$Reca_l$	F_{score}	MCC
TRP (70%)					
Point	97.54	97.61	92.49	94.98	93.41
Drag	98.71	96.64	98.29	97.46	96.60
Loupe	98.50	96.90	97.18	97.04	96.03
Pinch	97.18	92.79	95.91	94.32	92.47
Average	97.98	95.98	95.97	95.95	94.63
TSP (30%)					
Point	96.75	96.70	89.80	93.12	91.10
Drag	98.33	96.03	97.32	96.67	95.56
Loupe	98.75	96.32	98.63	97.46	96.64
Pinch	97.17	93.25	96.20	94.70	92.79
Average	97.75	95.58	95.49	95.49	94.03

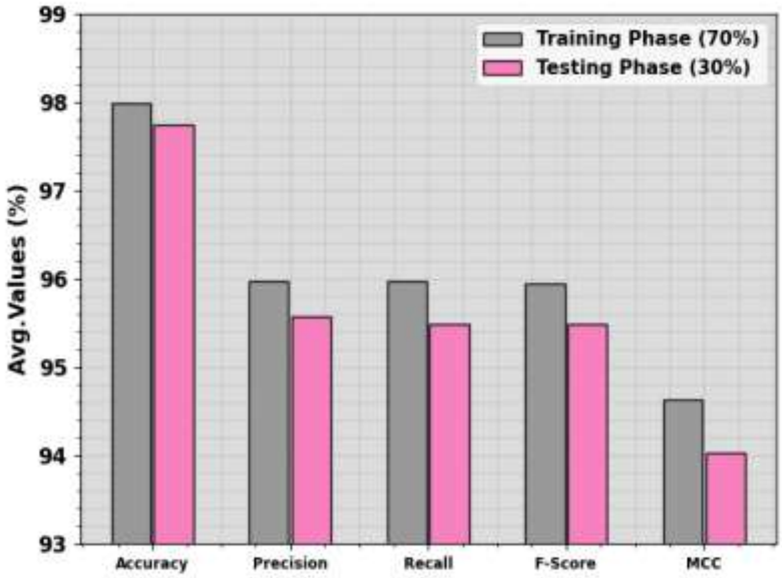

Figure 5: Average outcomes of the AGR-ZOADL technique with 70:30 TRP/TSP

Doi: https://doi.org/10.54216/FPA.160102
Received: July 29, 2023 Revised: November 17, 2023 Accepted: April 12, 2024

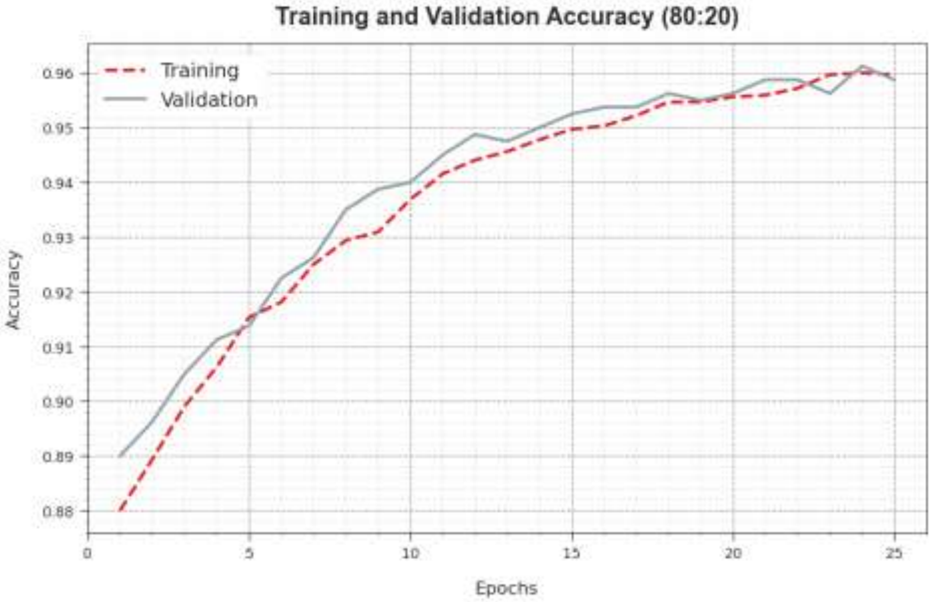

Figure 6: $Accu_y$ curve of the AGR-ZOADL method with 80:20 TRP/TSP

The $accu_y$ curves for training (TR) and validation (VL) exhibited in Fig. 6 for the AGR-ZOADL algorithm with 80:20 TRP/TSP gives respected insights into its effectiveness with several epoch count. Specifically, it can depend on augmentation in TR and TS $accu_y$ with improving epoch count, displays the model's capability to recognize and learn patterns in TR and TS data. The increasing trends in TS $accu_y$ exhibits the model's adaptability for dataset of TR and its capabilities for creating exact predictions on unnoticed data, emphasizing supreme generalization proficiencies.

Fig. 7 indicates a comprehensive review of TR and TS loss values to AGR-ZOADL methodology with 80:20 TRP/TSP in plentiful epoch count. The TR loss is consistently minimized as the model develops its weights for decreasing classifier error rates with these datasets. These loss curves remarkably signify the model's alignment with database of TR, underscoring proficiencies in capturing patterns in these databases. The continuous parameter changes in the AGR-ZOADL technique, are targeted at lessening discrepancies among actual and predicted TR labels.

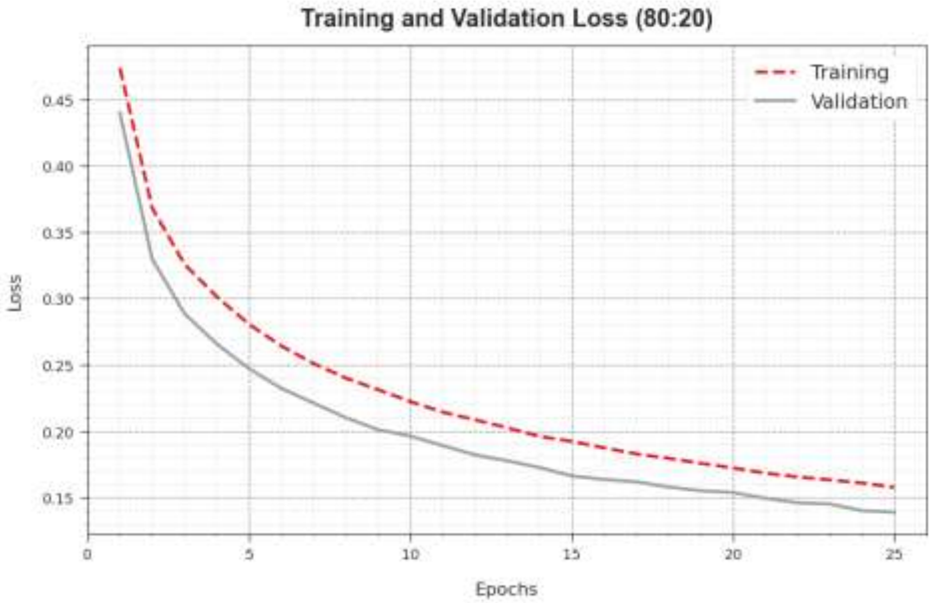

Figure 7: Loss curve of the AGR-ZOADL algorithm on 80:20 TRP/TSP

Doi: https://doi.org/10.54216/FPA.160102
Received: July 29, 2023 Revised: November 17, 2023 Accepted: April 12, 2024

Regarding the precision-recall (PR) curve displayed in Fig. 8, the findings unequivocally affirm that the AGR-ZOADL model on 80:20 TRP/TSP consistently attains increased PR values through each class. These outcomes underscore the model's active capacity for discriminating amongst diverse classes, emphasizing its effectiveness in recognizing class labels accurately.

Moreover, in Fig. 9, we exhibit ROC curves produced by the AGR-ZOADL approach under 80:20 TRP/TSP, representing its proficiency in differentiating between classes. These curves offer valuable insights into how the tradeoffs between FPR and TPR differs at diverse classification epochs and thresholds. The accomplished outcomes underscore the model's accurate classification efficiency in different class labels, highlighting its effectiveness in addressing numerous classification challenges.

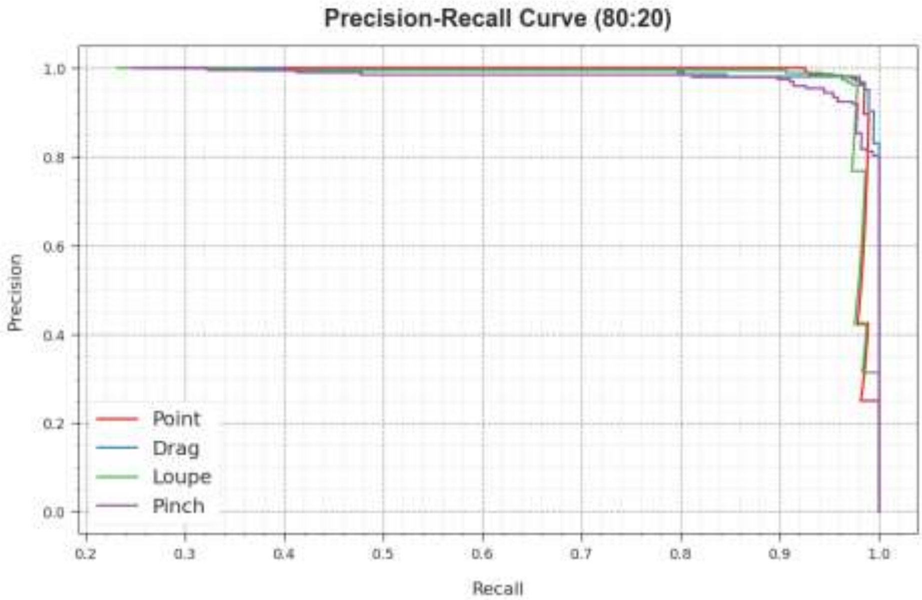

Figure 8: PR analysis of the AGR-ZOADL technique under 80:20 TRP/TSP

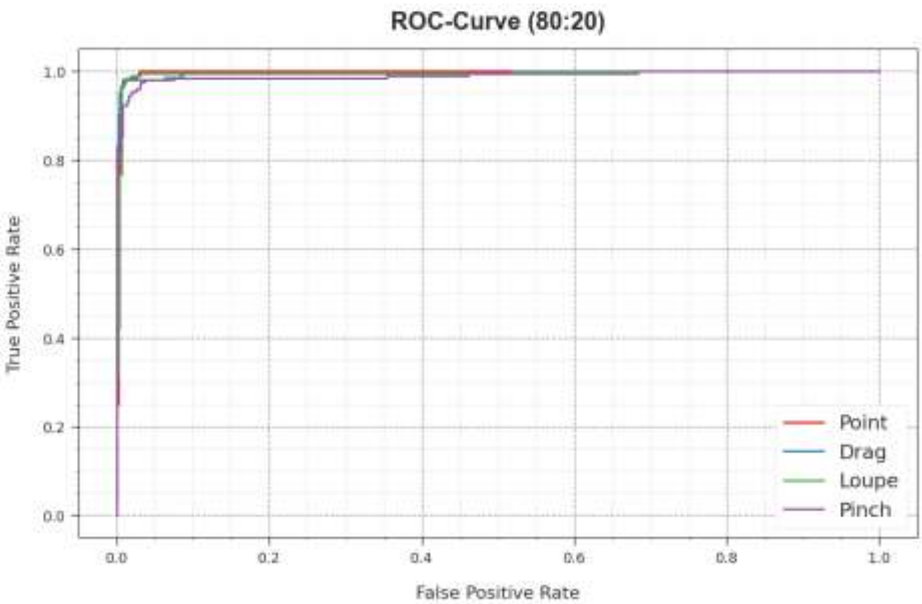

Figure 9: ROC analysis of the AGR-ZOADL method on 80:20 TRP/TSP

Table 4: Comparison outcome of the AGR-ZOADL approach with other algorithms

Methods	$Accu_y$	$Prec_n$	$Reca_l$	F_{score}

Doi: https://doi.org/10.54216/FPA.160102
Received: July 29, 2023 Revised: November 17, 2023 Accepted: April 12, 2024

ResNet-50	97.25	95.58	94.60	95.06
Inception-v3	96.27	95.03	94.67	94.83
MobileNet-v1	97.45	95.29	95.12	95.21
EfficientNet-B0	97.66	95.41	95.33	95.32
EfficientNet-B1	96.09	95.06	94.69	94.91
AGR-ZOADL	98.06	96.10	96.14	96.09

Table 4 and Fig. 10 represent a brief comparison analysis of the AGR-ZOADL method in terms of different measures [14]. The results stated that Inception-v3 and EfficientNet-B1 models obtained decreased performance. Followed by, ResNet-50, MobileNet-v1, and EfficientNet-B0 models provide closer performance. However, the AGR-ZOADL technique gains better results with increased $accu_y$ of 98.06%, $prec_n$ of 96.10%, $reca_l$ of 96.14%, and F_{score} of 96.09%.

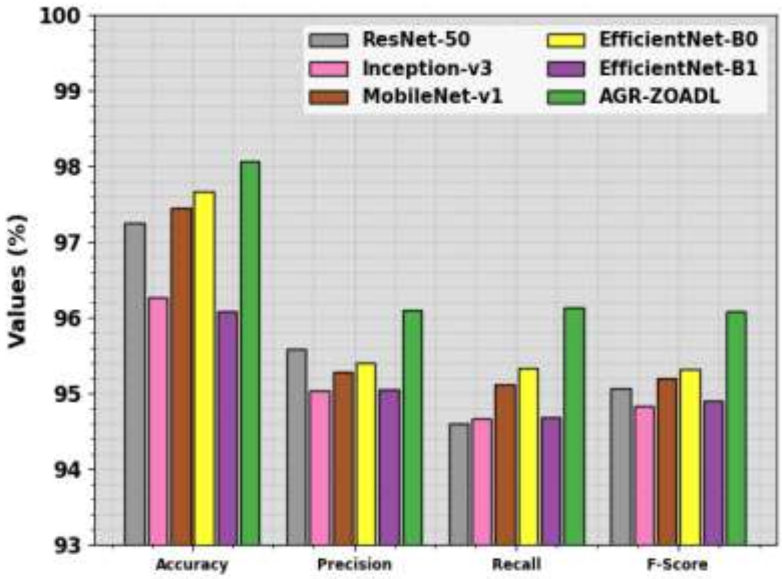

Figure 10: Comparative outcome of AGR-ZOADL model with other systems

Table 5 and Fig. 11 display a detailed comparison outcome of AGR-ZOADL methodology with respect of diverse measures. The acquired outcome shows EfficientNet-B0, Inception-v3 and ResNet-50 systems get increased performance. Next, EfficientNet-B1 and MobileNet-v1 methods give nearby performance. However, AGR-ZOADL technique attains higher results with minimized CT values of 4.18s respectively. Therefore, AGR-ZOADL method can be employed for an effective gesture detection method.

Table 5: CT analysis of AGR-ZOADL approach with other algorithms

Methods	Computational Time (sec)
ResNet-50	11.74
Inception-v3	11.87
MobileNet-v1	07.56
EfficientNet-B0	17.97
EfficientNet-B1	10.45
AGR-ZOADL	04.18

Doi: https://doi.org/10.54216/FPA.160102
Received: July 29, 2023 Revised: November 17, 2023 Accepted: April 12, 2024

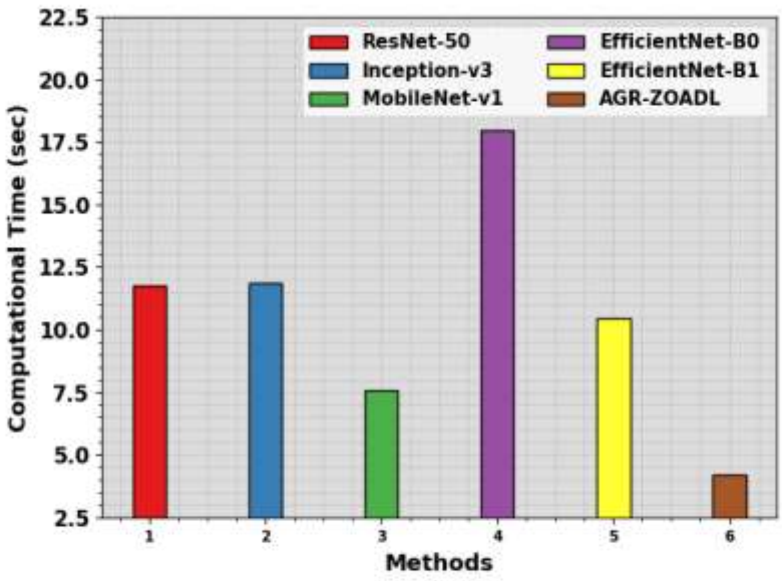

Figure 11: CT analysis of the AGR-ZOADL model with other algorithms

5. Conclusion

In this article, we have developed the AGR-ZOADL algorithm for Visually Challenged People. The AGR-ZOADL technique aims to recognize gestures to aid visually challenged people. It contains four major processes involved as MF MF-based preprocessing, NASNet based feature extraction, ZOA based hyperparameter, and SLSTM based classification. Besides, the AGR-ZOADL technique applies NASNet model to learn complex features from the pre-processed data. The ZOA based hyperparameter process is performed to optimize the accuracy of NASNet model. Finally, the SLSTM model is applied. The performance validation of the AGR-ZOADL method is carried out using a benchmark dataset. The experimental values stated that AGR-ZOADL technique reaches significant performance over other existing approaches.

Funding: “This research received no external funding”

Conflicts of Interest: “The authors declare no conflict of interest.”

References

[1] Dhivyasri, S., KB, K.H., Akash, M., Sona, M., Divyapriya, S. and Krishnaveni, V., 2021, May. An efficient approach for interpretation of Indian sign language using machine learning. In 2021 3rd International Conference on Signal Processing and Communication (ICPSC) (pp. 130-133). IEEE.

[2] Maro, J.M., Ieng, S.H. and Benosman, R., 2020. Event-based gesture recognition with dynamic background suppression using smartphone computational capabilities. Frontiers in neuroscience, 14, p.275.

[3] Yuksel, B.F., Fazli, P., Mathur, U., Bisht, V., Kim, S.J., Lee, J.J., Jin, S.J., Siu, Y.T., Miele, J.A. and Yoon, I., 2020, July. Human-in-the-loop machine learning to increase video accessibility for visually impaired and blind users. In Proceedings of the 2020 ACM Designing Interactive Systems Conference (pp. 47-60).

[4] Siddiqui, U.A., Ullah, F., Iqbal, A., Khan, A., Ullah, R., Paracha, S., Shahzad, H. and Kwak, K.S., 2021. Wearable-sensors-based platform for gesture recognition of autism spectrum disorder children using machine learning algorithms. Sensors, 21(10), p.3319.

[5] Gupta, S., Chakraborti, S., Yogitha, R. and Mathivanan, G., 2022, May. Object Detection with Audio Comments Using YOLO V3. In 2022 International Conference on Applied Artificial Intelligence and Computing (ICAAIC) (pp. 903-909). IEEE.

[6] Fatmi, R., Rashad, S. and Integlia, R., 2019, January. Comparing ANN, SVM, and HMM based machine learning methods for American sign language recognition using wearable motion sensors. In 2019 IEEE 9th annual computing and communication workshop and conference (CCWC) (pp. 0290-0297). IEEE.

Doi: https://doi.org/10.54216/FPA.160102
Received: July 29, 2023 Revised: November 17, 2023 Accepted: April 12, 2024

[7] Aithal, C.N., Ishwarya, P., Sneha, S., Yashvardhan, C.N., Kumar, D. and Suresh, K.V., 2023, January. Hand Gesture Recognition in Complex Background. In Cognition and Recognition: 8th International Conference, ICCR 2021, Mandya, India, December 30–31, 2021, Revised Selected Papers (pp. 243-257). Cham: Springer Nature Switzerland.
[8] Afif, M., Ayachi, R., Pissaloux, E., Said, Y. and Atri, M., 2020. Indoor objects detection and recognition for an ICT mobility assistance of visually impaired people. Multimedia Tools and Applications, 79, pp.31645-31662.
[9] Chen, B., Chen, C., Hu, J., Sayeed, Z., Qi, J., Darwiche, H.F., Little, B.E., Lou, S., Darwish, M., Foote, C. and Palacio-Lascano, C., 2022. Computer Vision and Machine Learning-Based Gait Pattern Recognition for Flat Fall Prediction. Sensors, 22(20), p.7960.
[10] Jadhav, A., Padwad, H., Chandak, M.B. and Raut, R., 2022. Use of Assistive Techniques for the Visually Impaired People. Intelligent Systems for Rehabilitation Engineering, pp.115-127.
[11] Subudhi, B.N., Veerakumar, T., Harathas, S.R., Prabhudesai, R., Kuppili, V. and Jakhetiya, V., 2023. Deep Learning in Autoencoder Framework and Shape Prior for Hand Gesture Recognition. In Smart Computer Vision (pp. 223-242). Cham: Springer International Publishing.
[12] Adithya, V. and Rajesh, R., 2020. A deep convolutional neural network approach for static hand gesture recognition. Procedia Computer Science, 171, pp.2353-2361.
[13] Mujahid, A., Awan, M.J., Yasin, A., Mohammed, M.A., Damaševičius, R., Maskeliūnas, R. and Abdulkareem, K.H., 2021. Real-time hand gesture recognition based on deep learning YOLOv3 model. Applied Sciences, 11(9), p.4164.
[14] Alashhab, S., Gallego, A.J. and Lozano, M.Á., 2022. Efficient gesture recognition for the assistance of visually impaired people using multi-head neural networks. Engineering Applications of Artificial Intelligence, 114, p.105188.
[15] Adeel, M.I., Asad, M.A., Zeeshan, M.R., Amna, M., Aslam, M. and Martinez-Enriquez, A.M., 2022, March. Gesture Based Confidence Assessment System for Visually Impaired People Using Deep Learning. In Future of Information and Communication Conference (pp. 135-147). Cham: Springer International Publishing.
[16] Can, C., Kaya, Y. and Kılıç, F., 2021. A deep convolutional neural network model for hand gesture recognition in 2D near-infrared images. Biomedical Physics & Engineering Express, 7(5), p.055005.
[17] Ryumin, D., Ivanko, D. and Ryumina, E., 2023. Audio-Visual Speech and Gesture Recognition by Sensors of Mobile Devices. Sensors, 23(4), p.2284.
[18] Parvathy, P., Subramaniam, K., Prasanna Venkatesan, G.K.D., Karthikaikumar, P., Varghese, J. and Jayasankar, T., 2021. Development of hand gesture recognition system using machine learning. Journal of Ambient Intelligence and Humanized Computing, 12, pp.6793-6800.
[19] Singh, K., Kansal, A. and Singh, G., 2019. An improved median filtering anti-forensics with better image quality and forensic undetectability. Multidimensional Systems and Signal Processing, 30(4), pp.1951-1974.
[20] Hazarika, R.A., Kandar, D. and Maji, A.K., 2022. An experimental analysis of different deep learning based models for Alzheimer's disease classification using brain magnetic resonance images. Journal of King Saud University-Computer and Information Sciences, 34(10), pp.8576-8598.
[21] Trojovská, E., Dehghani, M. and Trojovský, P., 2022. Zebra optimization algorithm: A new bio-inspired optimization algorithm for solving optimization algorithm. IEEE Access, 10, pp.49445-49473.
[22] Shi, Y.F., Yang, C., Wang, J., Zheng, Y., Meng, F.Y. and Chernogor, L.F., 2023. A Hybrid Deep Learning-Based Forecasting Model for the Peak Height of Ionospheric F2 Layer. Space Weather, 21(10), p.e2023SW003581.
[23] https://www.dlsi.ua.es/~jgallego/datasets/gestures/

Doi: https://doi.org/10.54216/FPA.160102
Received: July 29, 2023 Revised: November 17, 2023 Accepted: April 12, 2024