Facial paralysis image analysis for stroke detection using deep ensemble transfer learning and optimization

Kiruthiga Subramaniyan[1], Chinnasamy Anbuananth[1], Dhilip Kumar Venkatesan[2]

[1]Department of Computer Science and Engineering, Faculty of Engineering and Technology, Annamalai University, Chidambaram, India

[2]Department of Computer Science and Engineering, School of Computing, Vel Tech. Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai, India

Article Info

ABSTRACT

Facial paralysis (FP) weakens facial muscles, leading to asymmetric facial actions and complicating stroke diagnosis. Machine learning (ML) and deep learning (DL) systems have been explored for diagnosing FP, but the effectiveness of these methods is hindered by the limited size and diversity of available datasets. This study proposes a novel deep ensemble transfer learning method for accurate stroke diagnosis using facial paralysis imaging (DETLM-ASDFPI). The method leverages pre-trained models to reduce computation costs on edge devices. The framework includes data acquisition, preparation, and pre-processing, with image rescaling to standardize input dimensions. Feature extraction is performed using a deep capsule network (DCapsNet) to capture complex features. For stroke detection, an ensemble transfer learning model integrates three classifiers: gated recurrent unit (GRU), deep convolutional neural network (DCNN), and stacked sparse auto-encoder (SSAE). The hippopotamus optimization algorithm (HOA) is applied to fine-tune model parameters. The method was validated using two benchmark datasets, Massachusetts eye and ear infirmary (MEEI) and YouTube facial palsy (YFP), achieving an accuracy of 97.06%, outperforming recent approaches. This research demonstrates the effectiveness of the DETLM-ASDFPI method in accurately diagnosing strokes from FP images while addressing challenges related to dataset limitations and computational efficiency.

This is an open access article under the CC BY-SA license.

Corresponding Author:

Kiruthiga Subramaniyan
Department of Computer Science and Engineering, Faculty of Engineering and Technology
Annamalai University
Chidambaram, Tamilnadu, India
Email: kiruthigaresearchscholar@gmail.com

1. INTRODUCTION

A stroke, the second paramount reason for demise globally and a primary reason is a disability, can be a neurological deficiency that outcomes mostly from severe focal damage of the central nervous system (CNS) owing to vascular reasons such as intracerebral haemorrhage, subarachnoid haemorrhage, and cerebral infarction [1]. It's a medical crisis and hence needs instant treatment. To accurately detect suspected stroke patients, Facial paralysis (FP) has been extensively applied for emergency medical treatment. FP damages facial actions due to nerve pathology [2]. It causes damage to the function of intentional facial muscles strengthened through the facial nerve, resulting in facial anomalies [3]. Peripheral FP corresponds to nerve disorder inside the pontine of the brain stem. It affects the muscles of the face in the upper, middle, and lower areas of the face on

one side. Central FP (because of stroke) becomes nerve failure within the motor cortical regions; the lower part of one side of the face can be influenced [4]. Most of the FP patients become infected with peripheral FP. Then, it has an emotional impact on most facial muscles of the face on one side; it's problematic for the patient to carry out the regular actions of the mouth, eyebrows, and eyes [5].

Detecting the facial palsy process is significant in measuring the difficulty of the muscle failure and facial nerve, recording physical developments after treatment, and observing the patient [6]. Computerized automated FP detection is significant in developing systematized devices for treatment, monitoring, and medical valuation, and medical prices have decreased over the addition of automated methods [7]. Moreover, computerized methods are predicted to offer convenient devices, shortly, for patient monitoring in the house. Mainly in computer vision (CV), the study of facial movements has inspired several investigations on automatic facial nerve function valuation from the biological visual capture of the face [8]. Recently, with the assistance of AI, investigators have suggested gradually more precise detection and assessment models for CNS conditions. In the current investigation, deep learning (DL) and machine learning (ML) models are presented for FP detection and prediction of the severity [9]. Nevertheless, ML models have restrictions as they depend on facial landmark methods and physical face paralysis area extraction models to get spatial data. Meanwhile, DL techniques normally utilize CNNs to remove in-depth features from facial features [10]. They identify slight modifications and detect patterns from facial images that assist in improved recognition of FP.

This study presents a new deep ensemble transfer learning method for accurate stroke diagnosis using facial paralysis imaging (DETLM-ASDFPI) method. The DETLM-ASDFPI method aims to classify strokes that exist in FP images proficiently. This study focuses on an advanced diagnostic framework integrating data acquisition, preparation, and pre-processing of FP images. The pre-processing phase includes rescaling the images to standardize input dimensions. Also, the deep capsule network (DCapsNet) method is used for feature extraction to learn complex features from the pre-processed data. For the stroke detection process, an ensemble transfer learning (TL) model utilizes three classifiers such as gated recurrent unit (GRU), deep convolutional neural network (DCNN), and stacked sparse auto-encoder (SSAE). Eventually, the hippopotamus optimization algorithm (HOA) is utilized to optimize parameter tuning for the three ensemble techniques. To improve the prediction results of the DETLM-ASDFPI methodology, a sequence of experiments is implemented on two benchmark datasets such as Massachusetts eye and ear infirmary (MEEI) and YouTube facial palsy (YFP). The major contribution of the DETLM-ASDFPI methodology is listed as follows:

–   The DETLM-ASDFPI model integrates a robust pre-processing step to rescale medical images, confirming uniform input sizes for the subsequent feature extraction and model training stages. This normalization assists in maintaining consistency across diverse image formats and improves the performance of the feature extraction process. Standardizing image dimensions enables the method to learn patterns more effectually, improving stroke detection accuracy.

–   The DETLM-ASDFPI method utilizes DCapsNet to capture spatial hierarchies and part-whole relationships in medical images. This allows for more precise and robust feature extraction, which is significant for detecting stroke-related patterns. By employing DCapsNet, the approach enhances the accuracy and reliability of stroke detection from complex medical imaging data.

–   The DETLM-ASDFPI model incorporates three classifiers, GRU, DCNN, and SSAE, into an ensemble TL model to enhance stroke detection. This approach utilizes the merits of every classifier, integrating sequential, spatial, and abstract feature learning for more accurate outcomes. The model benefits from pre-trained knowledge utilizing TL, improving generalization and performance on medical datasets.

–   The DETLM-ASDFPI methodologies utilize HOA for optimal hyperparameter tuning, confirming improved model performance in stroke detection. HOA assists in avoiding premature convergence by effectively exploring the solution space and fine-tuning model parameters. This enhances the technique's accuracy and robustness, resulting in more reliable stroke detection results.

–   The DETLM-ASDFPI approach incorporates ensemble TL with three distinct classifiers, GRU, DCNN, and SSAE, to utilize their complementary merits in stroke detection. This integration enhances the technique's capability to capture sequential, spatial, and abstract features from diverse data sources. Moreover, using the HOA for parameter tuning improves the model's efficiency by avoiding premature convergence and optimizing performance. This integration of models results in a more robust and adaptive model for precise stroke detection across varied datasets.

2.   RELATED WORKS

Ou *et al*. [11] introduced an innovative multimodal DL technique based on the face arm speech test (FAST) for evaluating suspicious stroke patients showing symptoms like speech disorders, limb weakness, and facial paralysis in severe sceneries. Based on the FAST, the author gathered a dataset containing audio and video recordings of intensive care unit patients enacting specified speech tests, limb movements, and facial expressions. The author equated the developed DL method to process multimodal datasets with six

preceding methods, which attained great movement classifier execution, containing the multiscale vision transformer (MViT), inflated 3D (I3D), SlowFast, extended 3D (X3D), and temporal pyramid network (TPN). Umirzakova et al. [12] introduce a DL-based technique for diagnosing FP illnesses like Bell's palsy and stroke. This method uses multi-task networks, incorporating facial asymmetry, face analysis, and type improvement to identify the costs accompanying conventional diagnosing techniques like computed tomography (CT) and magnetic resonance imaging (MRI) scanned images. Spatial differences were tackled by a depth-map estimating module, which influences a case-specified kernel technique. Chowdhury et al. [13] present an automatic system to identify the stroke from pre-processed data utilizing CNN and other DL methods. The presented approach is primarily to determine the stroked person's face from the expressions or normal face. For classification, the author gave pre-processed stroke images for training, served them into different deep structures, and ultimately, based on the categorized expressions, the author categorized stroke and normal patients.

Cai et al. [14] present a new multimodal DL method, DeepStroke, to attain computer-aided stroke incidence evaluation by diagnosing models of trivial facial muscle discoordination and speech incapability for patients with stroke impression in a critical situation. The presented DeepStroke endures 1-minute facial audio and video data freely accessible in stroke triages for local FP recognition and global speech illness examination. TL has been implemented to decrease face-attribute biases and enhance generalizability. Elhanashi et al. [15] propose a new technique to confront these crucial requirements by presenting a real-time stroke recognition method that depends on DL using federated learning (FL) to improve precision and confidentiality maintenance. The main aim of this case is to advance an effective and precise method of discriminating between non-stroke and stroke patients at present, enabling medical experts to make intellectual decisions. Phienphanich et al. [16] propose the usage of a facial image dataset encompassing smiling and neutral faces to identify facial faults that can be a usual symptom of stroke. The facial image dataset contains face images of stroke patients and normal subjects. This added dataset encompasses a set of smiling and neutral facial images created from freely available datasets that are augmented to develop two further smiling images at eight age groups. Gomes et al. [17] presented to concentrate on analyzing this asymmetry utilizing a DL technique without trusting manual computations, proposing the facial point graphs (FPG) method, a new technique that surpasses in studying geometrical data and efficiently managing differences after the possibility of handcrafted computations. FPG enables the technique to efficiently identify the facial disability caused by a stroke by utilizing video data.

3. THE PROPOSED MODEL

This work presents a new DETLM-ASDFPI technique. The major intention of the DETLM-ASDFPI technique is to identify strokes in FP images proficiently. The DETLM-ASDFPI technique has data preparation and pre-processing, a feature extractor, ensemble classification processes, and hyperparameter tuning to accomplish that. Figure 1 portrays the complete process of the DETLM-ASDFPI model. Stroke diagnosis is complicated by FP, which affects face muscles and produces asymmetry. Insufficient datasets restrict diagnostics using ML and DL. A new deep ensemble transfer learning approach for stroke diagnosis using facial paralysis imaging is proposed in this paper. Pre-trained models lower edge device computational expenses. The framework includes data gathering, picture rescaling, and deep capsule network feature extraction. A stroke detection ensemble transfer learning model uses GRU, CNN, and stacked sparse auto-encoder classifiers. Hippopotamus optimization optimizes parameters. On MEEI and YFP benchmark datasets, DETLM-ASDFPI outperforms current techniques with 97.06% accuracy. This method addresses dataset and computational issues while improving stroke diagnosis using FP pictures.

3.1. Data preparation and pre-processing

Data pre-processing is vital in the combined dataset to increase quality by executing image pre-processing and augmenting tasks [18]. These cropped images in the organized dataset were rescaled to normal dimensions of 224×224 for input to the TL method. Additionally, these are resized within the interval (0–1) by separating each pixel value by 225. Rescaling has been applied to normalize the image data in the normal variety of (0–1). Certain classes get more images than others, while some images within the datasets are noisy and characterize insufficient data about FP. Initially, noise images were extracted from the dataset. Added images have been made with data augmentation models to attain the offset between each class. Conventional image-augmenting models like rotation, zooming, flipping, and shearing are applied for data augmentation. The goal is to achieve similar image counts in every FP class to improve the training data diversities and enhance the generalizability of the algorithm. This augmentation phase is critical in making a higher-quality dataset to train a precise FP classification method. Either image augmentation or pre-processing operations are carried out with the Keras library ImageDataGenerator class.
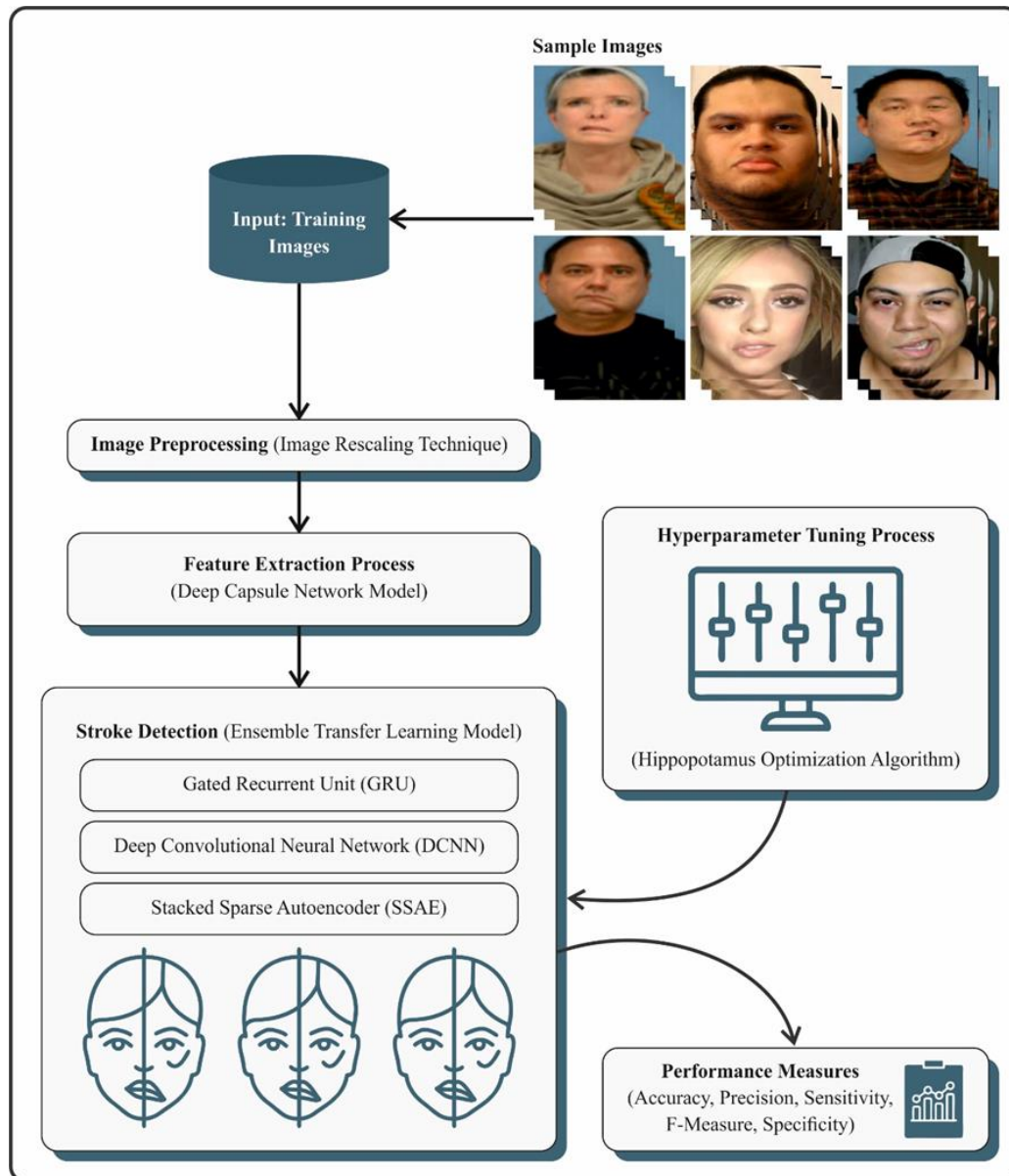
Figure 1. Overall process of DETLM-ASDFPI methodology

3.2. DCapsNet-based feature extraction

Besides, the DCapsNet model is employed for the feature extraction process to learn complex features from the pre-processed data [19]. This model gives a significant advantage in feature extraction tasks because it can capture spatial hierarchies and preserve part-whole relationships in data. Unlike conventional CNNs, which face difficulty handling viewpoint discrepancies and pose changes, DCapsNet utilizes capsules groups of neurons that encode features' presence and pose information. This allows the network to recognize objects or patterns more robustly, even in occlusions or transformations. Additionally, DCapsNet's dynamic routing mechanism between capsules enables effectual data propagation, resulting in more accurate and compact representations. By choosing DCapsNet over conventional methods, the methodology can improve generalization, mitigate the requirement for extensive data augmentation, and improve the interpretability of learned features. This makes it particularly effectual for complex tasks such as image classification, where spatial relationships play a significant role. Figure 2 illustrates the overall structure of the DCapsNet model.

The DCapsNet is utilized to process the merged variational image. According to the consistent samples, they can lastly gain an expert network. The DCapsNet for any variation recognition is achieved in the subsequent phases: i) Choose dual $n \times n$ samples, relate them directly into a $n \times 2n$ dimension, and utilize these for the network input; ii) Set the input within the Conv layer, with several convolutional kernels,

to remove various easy feature information. These main caps layer additionally chooses the removed feature data and merges the feature data into vectors. This digit caps layer standardizes this vector and categorizes it into a group of vectors; iii) Reshaping this vector within the 1D vectors and modifying it to include various image blocks of a selected dimension. Formerly, they are inputted within the system in the same way as earlier; and iv) Calculate the L2 norm from the vectors to classify and get the last classifier outcomes.
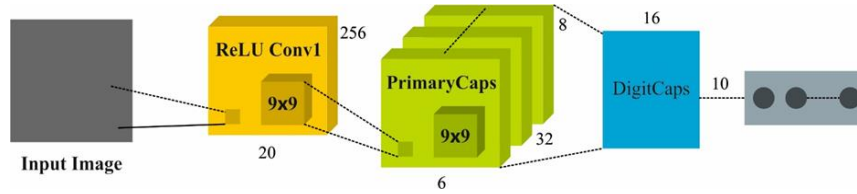
Figure 2. Overall structure of DCapsNet methodology

The neurons in every layer were separated into clusters, such as capsules. The conventional neuron output is reshaped and improved like a vector. The path consistencies model maintains the entity's position information and additional information. The training of the network has been calculated by employing different functions conventionally by utilizing the task (1).

$$L_c = T_c \max(0, m^+ - ||v_c||)^2 + \lambda(1 - T_c)\max(0, ||v_c|| - m^-)^2 \tag{1}$$

On the other hand, c signifies a kind of classification, and T_c denotes a parameter. If c occurs, T_c is 1, or 0. $m^+ = 0.9, m^- = 0.1$ represents missing cases in the current classification conditions. L_c is named the marginal loss. In CapsNet, the reconstruction loss is integrated with the margin loss to improve classification accuracy and learned representations' quality. The loss function enforces margin-based regularization, promoting intra-class compactness and inter-class separation, with the hyperparameter λ balancing these terms for enhanced model performance in tasks like metric learning.

3.3. Ensemble classification process

An ensemble TL model involves three classifiers for the stroke detection process: GRU, DCNN, and SSAE. An ensemble TL model integrating GRU, DCNN, and SSAE presents a robust approach to stroke detection by utilizing the merits of every model type. GRU is highly efficient for capturing sequential dependencies in time-series data, such as EEG signals or patient history, allowing it to detect temporal patterns crucial for stroke detection. DCNN outperforms in extracting spatial features from images, such as MRI or CT scans, making it ideal for detecting stroke-related anomalies in medical imaging. On the contrary, SSAE is proficient in learning sparse, high-level representations from unstructured data, which assists in detecting subtle features and patterns that may not be easily discernible. The ensemble approach improves the model's performance by incorporating the complementary merits of these diverse models, enhancing accuracy, robustness, and generalization related to single-model methods. This integration confirms that diverse facets of the stroke detection process, such as temporal, spatial, and abstract feature extraction, are optimally addressed.

3.3.1. GRU classifier

GRUs are an recurrent neural network (RNN) structure applied to process sequential data [20]. GRUs were initially presented as an easier substitution for the more composite long short-term memory (LSTM) system. GRUs were intended to tackle the problem of vanishing gradients that may occur in normal RNNs. If the loss function gradient becomes inadequate, upgrade the network parameters efficiently. The GRU structure contains gate mechanisms controlling the system's information flow. The gate, usually an update gate and a reset gate, controls what amount of the preceding condition has been maintained and what amount of original information is incorporated within the present condition. The reset gate controls which components of the prior condition must be forgotten. In contrast, the update gate regulates the amount of the original input that must be included in the present condition. The critical feature of GRUs is their capability to discard or maintain data from the preceding time phase, which makes them efficient for demonstrating longer-term dependencies in sequential data. GRUs were exposed to surpass other RNN techniques through various tasks such as speech recognition, machine translation, and language modelling. They were utilized in different applications in signal processing and CV, including anomaly detection, image captioning, and music

generation. GRUs are versatile and powerful devices for demonstrating sequential data, and their efficiency and simplicity make them a standard selection in the DL area. Figure 3 shows the infrastructure of the GRU model.
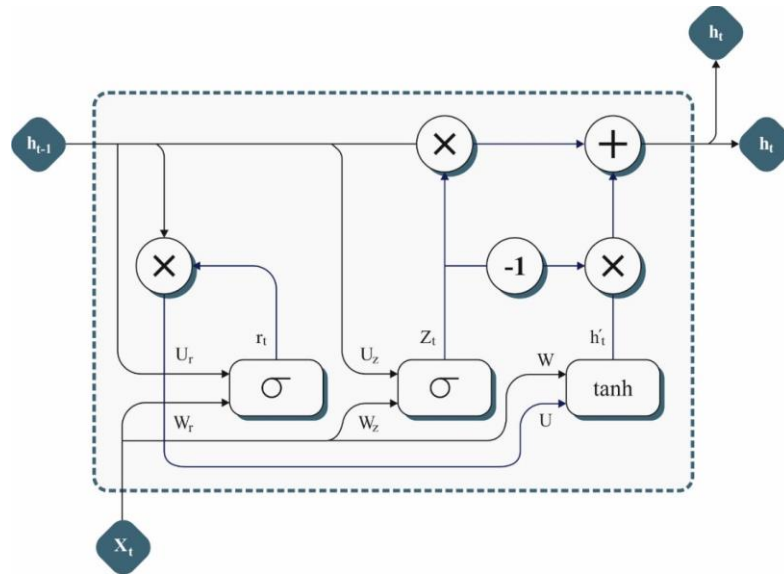

Figure 3. GRU architecture

3.3.2. DCNN Classifier

The DCNN network's final layer is SoftMax, using similar network node counts for the number of training data classifications [21]. Afterwards, the training was finished, and the final layer output of the network was chosen to remove the image features. Typically, the greater the network architecture, the multiple network layers, and the additional parameters to be studied, the larger the overfitting probability and the slowest training speed. However, the training might not function when the network contains minimal layers. In general, this work applies a network framework with three convolution layers. Figure 4 depicts the structure of the DCNN model.
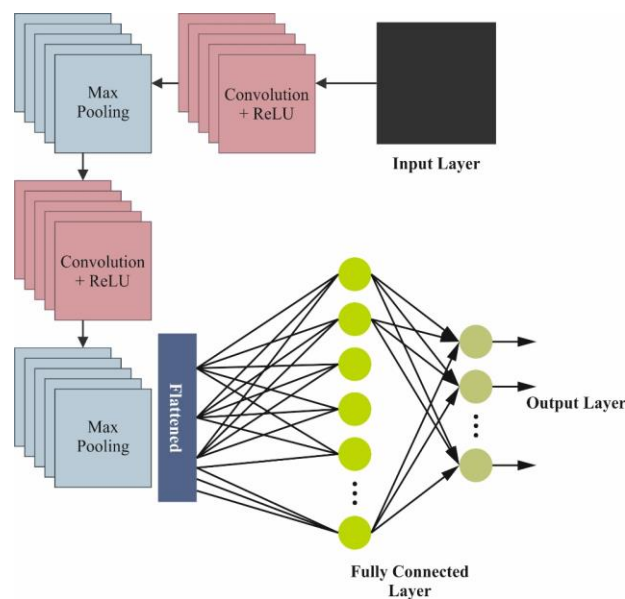

Figure 4. Structure of DCNN model

The DCNN with three convolutional layers has been applied. Besides the convolutional layer, dual fully connected (FC) layers exist, all of which can be accompanied by the pooling layer (sub-sampling layer). The input image is an image of a solitary target of 80×100 dimensions. The primary DCNN layer represents the convolutional layer related to the input and is represented by convolutional layer. The input of a solitary targeted image can be convoluted with twenty-four filters to present twenty-four feature mapping with Gaussian filters. This network utilizes dual types of convolutional kernels of various dimensions. The dimensions of the greater rectangle box within the figure are 5×5. Additionally, the dimensions of the lesser rectangle box correspond to the dimensions of the convolutional kernels of 3×3, and the dimensions of the convolutional kernel of these layers are 5×5. The convolutional phase dimensions are one. For the convolutional function to operate by the pixels by the image edge, they filled in two-pixel widths (pads) on every image side.

Then, convolutional layer contains twenty-four convolutional samples, and every convolution kernel needs 5×6 parameters. One of the parameters is 26×28, so the total parameter count for convolutional layer is 26×28. The SoftMax classifier has been utilized within the output layer. Unlike logistic regression, which can handle non-linear binary classification issues, SoftMax can handle multiple classification difficulties. A single instance can relate to a class; classes are equally exclusive. The node counts are similar for the number of training data classes.

The activation functions of every FC layer and convolutional layer are rectified linear unit (ReLU). The function of ReLU denotes a non-linear activation function, and the function of ReLU is significantly less mathematically than another function. Utilizing the function of ReLU will cause numerous network points to output zero, so it contains a particular self-consciousness resulting in overfitting.

3.3.3. SSAE classifier

An SAE is a standard 3-layer ANN, which can utilize unsupervised learning to acquire a compressed input data code [22]. The compressed code recognizes the dimension decrease of the raw input. Significantly, the AE is a great feature extraction for the DL network. Figure 5 demonstrates the architecture of the SSAE model.
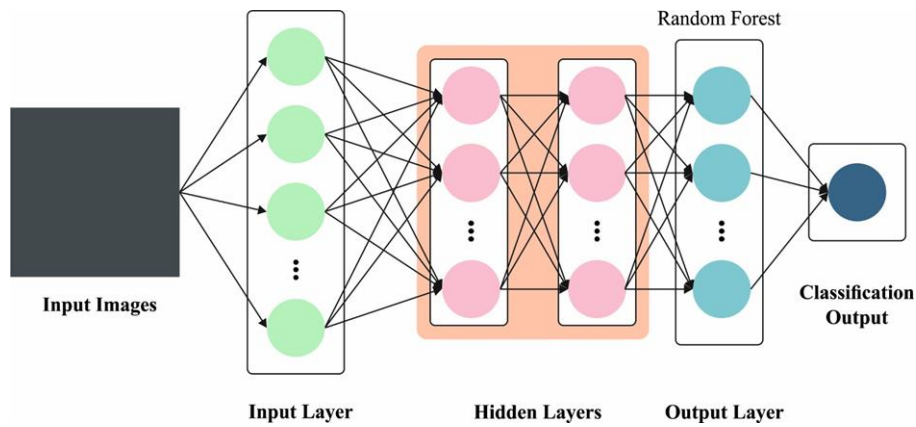

Figure 5. Architecture of the SSAE approach

In general, an SAE is built from encoding ϕ and decoding φ.

$$\phi: Z = \sigma(WX + b) \tag{2}$$

$$\varphi: X' = \sigma'(WZ + b') \tag{3}$$

Here, W, b, σ and W', b', σ' denote the weight, bias vector, and activation function of the encoder and decoder process. $X \in R^N$ means an input. $Z \in R^M$ refers to compressed code. $X' \in R^N$ denotes the reconstruction vector. The SAE can convert an input X into the latent variable Z and rebuild X' over the decoding. Therefore, the objective of AE is to focus $on\ W$ and W', which permits an output of decoding to improve the new input. For an assumed pair of data X, the reconstruction error of mathematical formulation as mentioned in (4).

$$L(x, x') = \|x - x'\|^2 = \|x - \sigma'[W'\sigma(wx + b) + b']\|^2 \tag{4}$$

The stacked self-encoder system contains numerous AE layers, which employ an output of previous AE layers as an input for the consequent one. I denote the number of samples for a complete set of training samples, $x^{(i)}$. The loss functions for SAE are in (5).

$$J(w, b) = \frac{1}{m} \sum_{i=1}^{m} [\, T\eta(j)$$ (5)

Here, m denotes a sum of samples, $and\ Y$ signifies the equivalent label of X. The regularization of weight attenuation is the 2nd term employed for preventing over-fitting of the method, and λ refers to weight attenuation co-efficient and is utilized to regulate the relative weight of 1st and 2nd terms. s_l^c and s_l^r indicate the total amount of neurons in $the\ lth$ layer, and W_{ji}^l refers to the connection weight among the neurons in dual layers.

The feature's sparsity is attained by including words in (6), and the complete feature extractor procedure is enhanced. Next, the loss function of standard stacked SAE (SSAEs) has been (7).

$$L_{sparse}(W, b) = J(W, b) + \beta \sum_{j=1}^{s_2} K\,L(\rho\square\hat{\rho}_j)$$ (6)

$$KL(\rho\hat{\rho}_j) = \rho\log\frac{\rho}{\hat{\rho}_j} + (1 - \rho)\log\frac{1-\rho}{1-\hat{\rho}_j}$$ (7)

Here, the symbols $\rho\ and\ \beta$ correspondingly represent the constant of sparseness and divergence. s_2 denotes the amount of hidden layer neurons. $\hat{\rho}$ refers to the value of mean activation.

3.4. HOA-based parameter optimizer

Eventually, the HOA is employed for the optimal parameter tuning of the three ensemble techniques [23]. The HOA is a novel metaheuristic inspired by the behaviour of hippopotamuses in their natural habitat. It is particularly effective in parameter optimization due to its unique balance between exploration and exploitation. The algorithm's ability to mimic the dynamic movement patterns of hippos-alternating between grazing in open fields and diving underwater-makes it highly suited for navigating complex, high-dimensional search spaces with local and global optima. Compared to traditional techniques like genetic algorithms or particle swarm optimization (PSO), HOA shows better robustness in avoiding premature convergence and can more effectively handle multimodal, noisy, and uncertain objective functions. Its simplicity and relatively fewer control parameters make implementing and tuning for diverse optimization tasks easier. Therefore, HOA offers an appealing alternative when tackling intricate and high-dimensional parameter optimization problems. Figure 6 shows the steps involved in the HOA methodology.
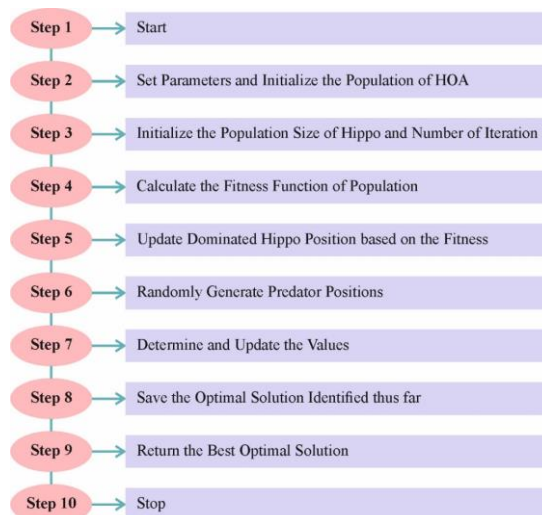

Figure 6. Steps involved in the HOA approach

The HOA model is between the most often applied and advanced bio-inspired metaheuristic optimization algorithms for solving composite optimization issues. This method has been stimulated by their herds' defence mechanisms and social behaviour. The HOA model contains 3 stages,

i) Phase 1: hippopotamuses updated location in the pond or river

This stage targets discovering the search space corresponding to how hippos move near their surroundings, the water. Hence, the location and movement of individual males $(x_{ij\,Mhippo})$, females, and the dominant hippo (D^{hippo}) control the exploration process. In (14), the D^{hippo} demonstrates the ideal solution to travelling another entity to its area. These distances are, in sequence, a function of not just the dominant hippo but also arbitrary vectors $(x_i i)$. (x_i), and (y_1) and an integer (I_1) signifying intrinsic variabilities witnessed in exploration.

$$x_i^{M\,hippo} : x_{ij}^{M\,hippo} = x_{ij} + y_1 \cdot \left(D^{hippo} - I1 x_{ij} \right) \tag{8}$$

When a female or male hippopotamus's location leads to a larger value of the objective function than the hippo's present dominance, the dominant location is substituted with that individual location. These mechanisms ensure that the exploration process constantly searches for improved solutions.

ii) Phase 2: the defence act of hippopotamus against predators (exploration)

The defence method is acquired if there is something to defend against; for example, the flock identifies crocodiles from the Nile. Instant defence from the hunter has been assumed, together with loud vocalization. In (9) represents the fast try near the danger. The arbitrary predator's movement $(\mathcal{P} \backslash : redator_j)$ can be represented by a randomly generated vector $(\vec{r}_8)$ range between (0-1) inside the lower (ll_j) and upper (ub_j) bounds of the decision variables at (j^{th}). Moreover, in (10) represented the modification in the distance between hippos and predators after the self-defence mechanism.

$$Predator \backslash :: \backslash : \mathcal{P} \backslash : redator_j \backslash := ll_j + \vec{r}_8 \cdot \left(ub_j - ll_j \right), j = 1,2,\dots,m \tag{9}$$

$$\vec{\mathcal{D}} = \left| \mathcal{P} \backslash : redator_j \backslash : - x_{i \backslash varvec j} \right| \tag{10}$$

iii) Phase 3: hippopotamus escaped from the predator

The hippos might escape by overcoming predator assaults or conditions in which they can't intensify an adequate defence to safer regions. Accurately, these stochastic refugia are demonstrated to pretend the randomness of escaping paths. Meanwhile, an original position offers an improved objective function value, representing a superior solution. These could describe how a hippo escapes, representing an effective escape.

The HO model is an advanced optimization algorithm inspired by the life habits and behaviors of hippos, such as their defensive mechanisms and group dynamics. It efficiently addresses complex optimization problems by simulating these biological strategies, improving both the robustness and adaptability of the search process. Through iterative stages, the model refines its solutions, making it particularly effective in tackling challenging optimization tasks. Its ability to balance exploration and exploitation enhances performance across various domains. The next is the flowchart and pseudocode illustrating the HOA model as shown in Algorithm 1.

Algorithm 1. Pseudocode of HOA
Start
Describe the optimization problem available.
Establishing the maximal iteration counts represented as "it" and determining the number of hippopotamuses mentioned as "N".
Generating the initial hippopotamus population and evaluating the main functions of this primary group.
For $i = 1$, it
Updated the location of the dominant hippos by the standard of main function value condition
Stage1: updated the location of the hippos within the pond or river
For $i = 1 : N/2$
Compute a novel location for i^{th} hippos
Updated the location for i^{th} hippos
End for
Stage2: hippopotamus defense against predators
For $i = 1 + N/2 : N$
Make arbitrary locations for predator
Compute a novel location for i^{th} hippos
Updated the location for i^{th} hippos
End for
Stage3: hippopotamus escaped from the predator

Compute new boundaries of the variable's decision
For $i = 1: N$
Compute the novel location for i^{th} hippos
Updated the location for i^{th} hippos
End for
Saving the optimum candidate performance found so far
End for
Output the optimum solutions for the primary function
END

The HOA raises a fitness function (FF) to get superior classification performances. It identifies a positive integer to signify the greater implementation of the candidate outcomes. In this work, the reduction of the classifier error rate can be examined as the FF, as indicated in (11).

$$fitness(x_i) = ClassifierErrorRate(x_i) = \frac{no.of\ misclassified\ samples}{Total\ no.of\ samples} \times 100 \qquad (11)$$

4.   RESULT ANALYSIS AND DISCUSSION

The DETLM-ASDFPI technique's analysis is studied using two datasets: the MEEI dataset [24] and the YFP dataset [25], [26], which are combined to create a balanced dataset, as represented in Table 1. Figure 7 shows performance evaluation of the DETLM-ASDFPI Method on the test database. Figures 7(a) and 7(b) displays the confusion matrices with accurate classification and recognition of all 6 class labels on a 70:30 TRAST/TESST. Figure 7(c) shows the study of precision-recall curve, representing greater performances overall of 6 class labels. Eventually, Figure 7(d) demonstrates the study of receiver operating characteristic (ROC), signifying efficient values with better values of ROC for different classes.

Table 2 and Figure 8 signify the classifier study of the DETLM-ASDFPI methodology under 70%TRAST and 30%TESST. The findings stated that the DETLM-ASDFPI methodology appropriately identified the samples. With 70%TRAST, the DETLM-ASDFPI approach provides average $accu_y$, $prec_n$, $sens_y$, $spec_y$, and $F_{measure}$ of 97.06%, 91.20%, 91.20%, 98.24%, and 91.17%, respectively. Meanwhile, with 30%TESST, the DETLM-ASDFPI technique offers average $accu_y$, $prec_n$, $sens_y$, $spec_y$, and $F_{measure}$ of 96.93%, 90.82%, 90.83%, 98.15%, and 90.81%, correspondingly.

In Figure 9, the training $accu_y$ (TRAAC) and validation $accu_y$ (VLAAC) values of the DETLM-ASDFPI technique are displayed. The rate of $accu_y$ is estimated for 0-100 epoch counts. The figure underlined that the values of TRAAC and VLAAC display an increasing trend, which informed the capability of the DETLM-ASDFPI technique with enhanced execution over various iterations. Moreover, the TRAAC and VLAAC stay nearer over the epochs, which shows minimum overfitting and displays greater execution of the DETLM-ASDFPI model, promising constant prediction on unnoticed samples.

In Figure 10, the TRA loss (TRALS) and VLA loss (VLALS) graph of the DETLM-ASDFPI approach is exhibited. The rate of loss is estimated for 0-100 epochs. The rate of TRALS and VLALS exhibits a reducing trend, informing the capability of the DETLM-ASDFPI approach to balance a trade-off between data fitting and generalization. Moreover, the constant reduction in loss rate guarantees superior performances of the DETLM-ASDFPI technique and fine-tuning of the prediction values over time.

Table 3 presents a comparison study of the DETLM-ASDFPI method with current techniques [18], [27]. The VGG16 Net model attains 93.10% $accu_y$, while PHCNN-LSTM enhances 94.80%. TPCNN exhibits high $spec_y$ 97.09% but lower $accu_y$ 89.91%, and ResNet50 reaches 95.80% $accu_y$ but with lesser $sens_y$ 78.79%. random forest (RF) and support vector machine (SVM) classifiers give an improved accuracy of 93.65% and 94.87%, respectively, with varying $prec_n$ and $sens_y$. The ensemble classifier achieves 96.13% $accu_y$, and the DETLM-ASDFPI model outperforms all, with 97.06% $accu_y$, 91.20% $prec_n$, 91.20% $sens_y$, and 98.24% $spec_y$.

Table 1. Details of database

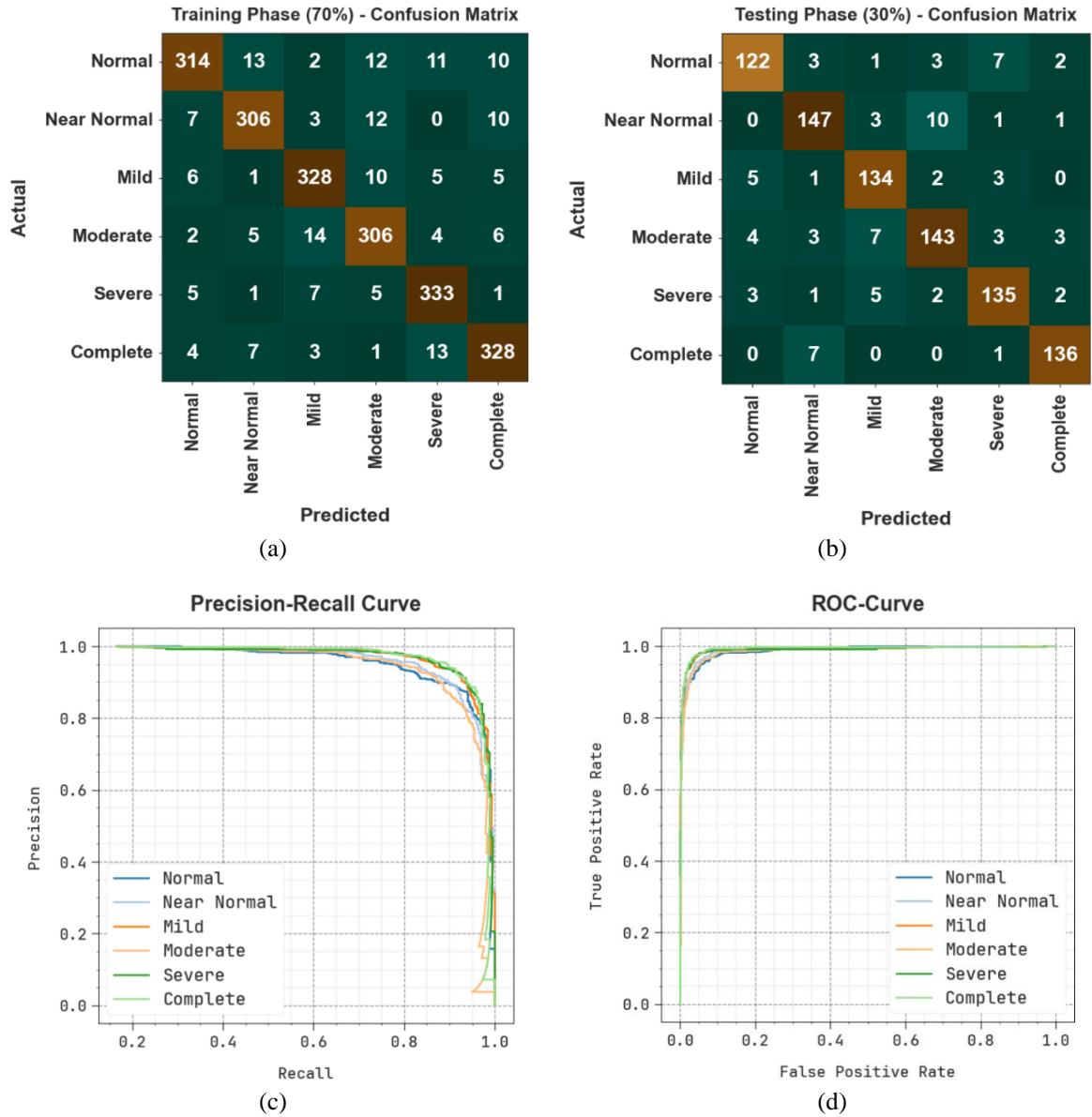| Severity classes	Number of images
Normal	500
Near normal	500
Mild	500
Moderate	500
Severe	500
Complete	500
Total images	3,000

(a)

(b)

(c)

(d)

Figure 7. Performance evaluation of the DETLM-ASDFPI method of (a) confusion matrix for training phase, (b) confusion matrix for testing phase, (c) precision-recall curve, and (d) ROC curve

Table 2. Classifier outcome of DETLM-ASDFPI technique under 70%TRAST and 30%TESST

Class	$Accu_y$	$Prec_n$	$Sens_y$	$Spec_y$	$F_{measure}$
TRAST (70%)					
Normal	96.57	92.90	86.74	98.62	89.71
Near normal	97.19	91.89	90.53	98.47	91.21
Mild	97.33	91.88	92.39	98.34	92.13
Moderate	96.62	88.44	90.80	97.73	89.60
Severe	97.52	90.98	94.60	98.11	92.76
Complete	97.14	91.11	92.13	98.17	91.62
Average	97.06	91.20	91.20	98.24	91.17
TESST (30%)					
Normal	96.89	91.04	88.41	98.43	89.71
Near normal	96.67	90.74	90.74	97.97	90.74
Mild	97.00	89.33	92.41	97.88	90.85
Moderate	95.89	89.38	87.73	97.69	88.54
Severe	96.89	90.00	91.22	98.01	90.60
Complete	98.22	94.44	94.44	98.94	94.44
Average	96.93	90.82	90.83	98.15	90.81
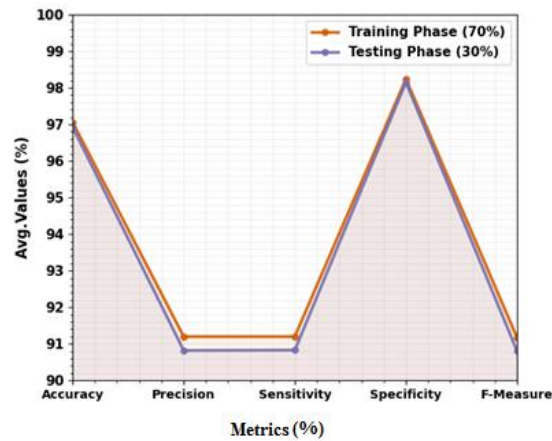
Figure 8. Average of DETLM-ASDFPI technique under 70%TRAST and 30%TESST
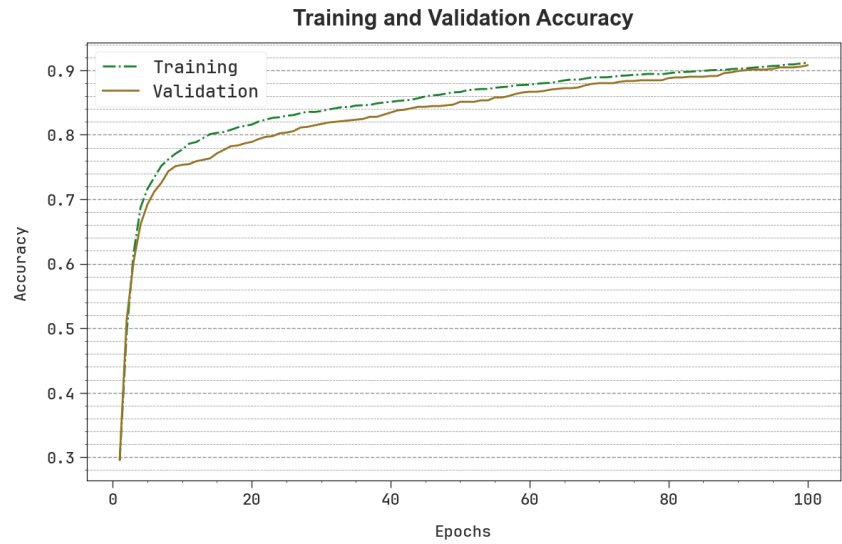

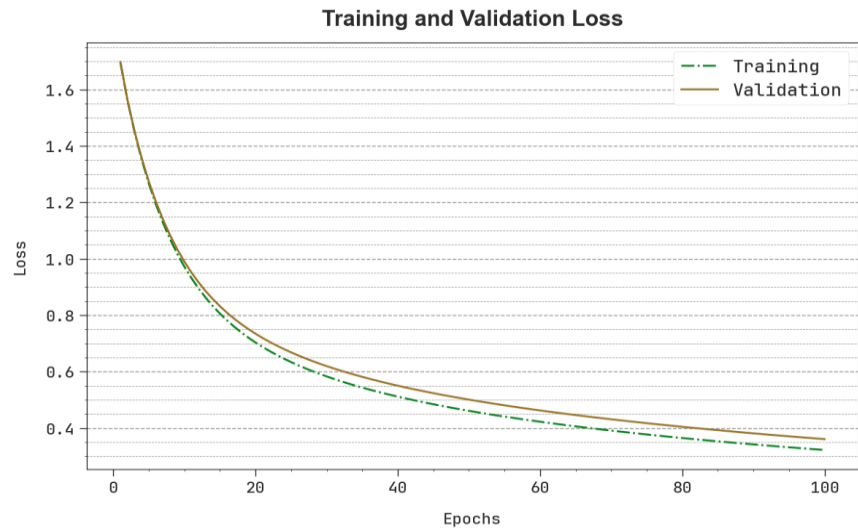Figure 9. $Accu_y$ curve of the DETLM-ASDFPI technique

Figure 10. Loss curve of DETLM-ASDFPI technique

Table 3. Comparative outcome of DETLM-ASDFPI model with other approaches [18], [27]

Classifier	$Accu_y$ (%)	$Prec_n$ (%)	$Sens_y$ (%)	$Spec_y$ (%)
VGG16 Net model	93.10	80.70	85.20	96.62
PHCNN-LSTM model	94.80	85.14	89.04	91.53
TPCNN classifier	89.91	88.65	82.89	97.09
ResNet50 method	95.80	88.14	78.79	89.99
RF	93.65	84.20	80.07	95.96
SVM classifier	94.87	79.55	88.76	94.24
Ensemble classifier	96.13	81.10	83.41	95.00
DETLM-ASDFPI	97.06	91.20	91.20	98.24

Figure 11 denotes the comparative study of the DETLM-ASDFPI model with other methods based on $accu_y$ and $prec_n$. The values stated that the DETLM-ASDFPI model outperformed has better performances. Based on $accu_y$, the DETLM-ASDFPI technique attains greater performances with an $accu_y$ rate of 97.06%. In contrast, the VGG16 Net, PHCNN-LSTM, TPCNN, ResNet50, RF, SVM, and Ensemble methods have attained lower $accu_y$ values of 93.10%, 94.80%, 89.91%, 95.80%, 93.65%, 94.87%, and 96.13%, correspondingly. Also, for $prec_n$, the DETLM-ASDFPI technique got superior execution with $prec_n$ values of 91.20%. At the same time, the VGG16 Net, PHCNN-LSTM, TPCNN, ResNet50, RF, SVM, and ensemble methods have attained minimal $prec_n$ values of 80.70%, 85.14%, 88.65%, 88.14%, 84.20%, 79.55%, and 81.10%, correspondingly.

Figure 12 signifies the comparative study of the DETLM-ASDFPI model with other techniques based on $sens_y$ and $spec_y$. The table values implied that the DETLM-ASDFPI model demonstrated superior performances. Regarding $sens_y$, the DETLM-ASDFPI technique performed better with $sens_y$ values of 91.20%. At the same time, the VGG16 Net, PHCNN-LSTM, TPCNN, ResNet50, RF, SVM, and ensemble methods have attained lower $sens_y$ values of 85.20%, 89.04%, 82.89%, 78.79%, 80.07%, 88.76%, and 83.41%, correspondingly. Moreover, based on $spec_y$, the DETLM-ASDFPI technique got greater execution with $spec_y$ values of 98.24%. In contrast, the VGG16 Net, PHCNN-LSTM, TPCNN, ResNet50, RF, SVM, and Ensemble models have obtained lesser $spec_y$ values of 96.62%, 91.53%, 97.09%, 89.99%, 95.96%, 94.24%, and 95.00%, respectively.

Figure 13 illustrates the classification results of the DETLM-ASDFPI model. The DETLM-ASDFPI model categorize the data into six distinct classes: normal, near normal, mild, moderate, severe, and complete. These classes represent diverse levels of severity or progression of the condition being analyzed. The normal class indicates healthy or unaffected instances, while near normal depicts slight deviations from the normal state. Mild and moderate reflect increasing degrees of impairment or abnormality, with Severe indicating a significant level of dysfunction or damage. The complete class portrays the most extreme state, where the condition is fully manifested, and the patient may be experiencing the most debilitating effects. The classification performance of the model is evaluated based on its capability to precisely assign instances to these categories, giving insights into the severity and progression of the condition.
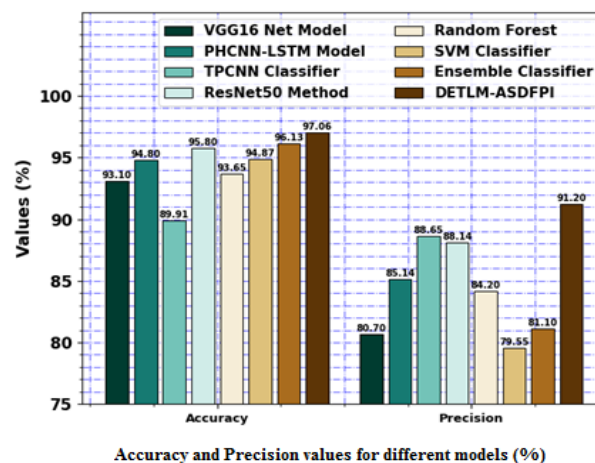
Figure 11. $Accu_y$ and $Prec_n$ analysis of the DETLM-ASDFPI model with current techniques
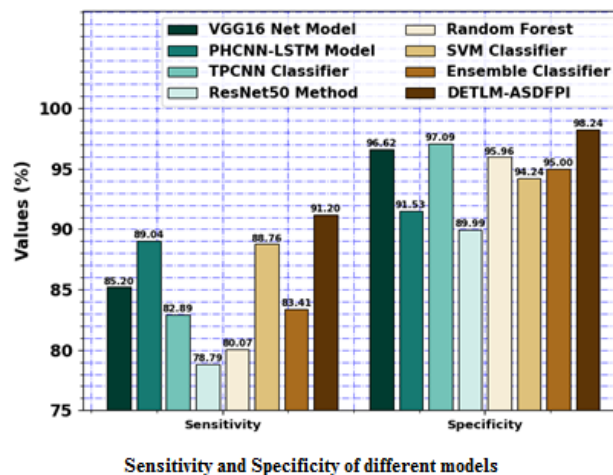
Figure 12. $Sens_y$ and $Spec_y$ analysis of DETLM-ASDFPI model with current techniques
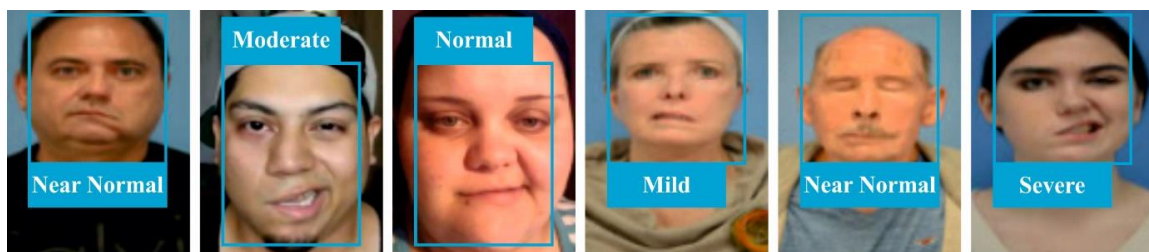

Figure 13. Classification results of the DETLM-ASDFPI model

5. CONCLUSION

In this study, a novel DETLM-ASDFPI model is presented. The primary purpose of the DETLM-ASDFPI model is to proficiently classify strokes that exist in FP images. This study focuses on an advanced diagnostic framework integrating data acquisition, preparation, and pre-processing of FP images. The pre-processing phase includes rescaling the images to standardize input dimensions. Besides, the DCapsNet model is employed for the feature extraction process to learn complex features from the pre-processed data. An ensemble TL model utilizes three classifiers, GRU, DCNN, and SSAE, for the stroke detection process. Finally, the HOA achieves the optimum hyperparameter tuning of the three ensemble techniques. To improve the prediction results of the DETLM-ASDFPI method, a sequence of experiments is implemented on two benchmark datasets such as MEEI and YFP. The experimental validation of the DETLM-ASDFPI method portrayed a superior accuracy value of 97.06% over recent approaches. The limitations of the DETLM-ASDFPI method comprise the dependence on a relatively small dataset or pre-trained dataset using ensemble transfer learning model, which may restrict the model's generalizability to larger, more diverse populations. Furthermore, the study primarily concentrates on stroke detection, and its performance may vary when applied to other medical conditions or imaging modalities. The computational complexity of the model could also pose challenges in real-time or resource-constrained environments. Moreover, the absence of interpretability in the DL-based approach may affect clinical adoption, as healthcare professionals need transparent decision-making processes. Future work should focus on expanding the dataset to comprise diverse demographics and exploring lightweight models for real-time applications with data privacy since some patients do not wish to share their raw medical records. Enhancing model interpretability through explainable AI models or ensuring privacy by adopting federated learning Techniques and testing the technique on other medical conditions will also be valuable directions for improving its utility and applicability in clinical settings.

FUNDING INFORMATION

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Kiruthiga Subramaniyan	✓	✓	✓	✓	✓	✓		✓	✓	✓			✓	
Chinnasamy Anbuananth		✓				✓	✓		✓	✓	✓	✓		
Dhilip Kumar Venkatesan	✓		✓	✓			✓			✓	✓		✓	✓

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis

I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting

Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

The authors have expressed no conflict of interest.

DATA AVAILABILITY

Data availability is not applicable to this paper as no new data were created or analyzed in this study.

REFERENCES

[1] C. Jiang *et al.*, "Automatic facial paralysis assessment via computational image analysis," *Journal of Healthcare Engineering*, vol. 2020, pp. 1–10, Feb. 2020, doi: 10.1155/2020/2398542.

[2] G. S. P.-Dominguez, R. E. S.-Yanez, and C. H. G.-Capulin, "Facial paralysis detection on images using key point analysis," *Applied Sciences*, vol. 11, no. 5, 2021, doi: 10.3390/app11052435.

[3] Z. Guo, W. Li, J. Dai, J. Xiang, and G. Dan, "Facial imaging and landmark detection technique for objective assessment of unilateral peripheral facial paralysis," *Enterprise Information Systems*, vol. 16, no. 10–11, pp. 1556–1572, Oct. 2022, doi: 10.1080/17517575.2021.1872108.

[4] H. Y. N. Oo, M. H. Lee, and J. H. Lim, "Exploring a multimodal fusion-based deep learning network for detecting facial palsy," *arXiv-Computer Science*, 2024, [Online]. Available: http://arxiv.org/abs/2405.16496.

[5] Y. Liu, Z. Xu, L. Ding, J. Jia, and X. Wu, "Automatic assessment of facial paralysis based on facial landmarks," in *2021 IEEE 2nd International Conference on Pattern Recognition and Machine Learning*, Jul. 2021, pp. 162–167, doi: 10.1109/PRML52754.2021.9520746.

[6] A. Song, Z. Wu, X. Ding, Q. Hu, and X. Di, "Neurologist standard classification of facial nerve paralysis with deep neural networks," *Future Internet*, vol. 10, no. 11, 2018, doi: 10.3390/fi10110111.

[7] M. A. Alagha, A. Ayoub, S. Morley, and X. Ju, "Objective grading facial paralysis severity using a dynamic 3D stereo photogrammetry imaging system," *Optics and Lasers in Engineering*, vol. 150, p. 106876, Mar. 2022, doi: 10.1016/j.optlaseng.2021.106876.

[8] J. J. Greene *et al.*, "The spectrum of facial palsy: the MEEI facial palsy photo and video standard set," *The Laryngoscope*, vol. 130, no. 1, pp. 32–37, Jan. 2020, doi: 10.1002/lary.27986.

[9] Y. Zhang *et al.*, "Early recognition of facial paralysis for rehabilitation of stroke patients using visual perception and AI-assisted analysis," in *2022 International Conference on Advanced Robotics and Mechatronics*, Jul. 2022, vol. 17, pp. 25–30, doi: 10.1109/ICARM54641.2022.9959402.

[10] L. Alsaedi, "New approach of estimating sarcasm based on the percentage of happiness of facial expression Using fuzzy inference system," *Journal La Multiapp*, vol. 3, no. 5, pp. 224–234, Nov. 2022, doi: 10.37899/journallamultiapp.v3i5.700.

[11] Z. Ou *et al.*, "Early identification of stroke through deep learning with multi-modal human speech and movement data," *Neural Regeneration Research*, vol. 20, no. 1, pp. 234–241, Jan. 2025, doi: 10.4103/1673-5374.393103.

[12] S. Umirzakova, S. Ahmad, S. Mardieva, S. Muksimova, and T. K. Whangbo, "Deep learning-driven diagnosis: a multi-task approach for segmenting stroke and Bell's palsy," *Pattern Recognition*, vol. 144, 2023, doi: 10.1016/j.patcog.2023.109866.

[13] N. A. Chowdhury *et al.*, "A novel approach to detect stroke from 2D images using deep learning," in *Lecture Notes in Networks and Systems*, 2024, vol. 867 LNNS, pp. 239–253, doi: 10.1007/978-981-99-8937-9_17.

[14] T. Cai *et al.*, "DeepStroke: an efficient stroke screening framework for emergency rooms with multimodal adversarial deep learning," *Medical Image Analysis*, vol. 80, 2022, doi: 10.1016/j.media.2022.102522.

[15] A. Elhanashi, P. Dini, S. Saponara, and Q. Zheng, "TeleStroke: real-time stroke detection with federated learning and YOLOv8 on edge devices," *Journal of Real-Time Image Processing*, vol. 21, no. 4, 2024, doi: 10.1007/s11554-024-01500-1.

[16] P. Phienphanich, N. Lerthirunvibul, E. Charnnarong, A. Munthuli, C. Tantibundhit, and N. C. Suwanwela, "Generalizing a small facial image dataset using facial generative adversarial networks for stroke's facial weakness screening," *IEEE Access*, vol. 11, pp. 64886–64896, 2023, doi: 10.1109/ACCESS.2023.3287389.

[17] N. B. Gomes, A. Yoshida, G. C. de Oliveira, M. Roder, and J. P. Papa, "Facial point graphs for stroke identification," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 14469 LNCS, pp. 685–699, 2024, doi: 10.1007/978-3-031-49018-7_49.

[18] W. Ali, M. Imran, M. U. Yaseen, K. Aurangzeb, N. Ashraf, and S. Aslam, "A transfer learning approach for facial paralysis severity detection," *IEEE Access*, vol. 11, pp. 127492–127508, 2023, doi: 10.1109/ACCESS.2023.3330242.

[19] W. Ma, Y. Xiong, Y. Wu, H. Yang, X. Zhang, and L. Jiao, "Change detection in remote sensing images based on image mapping and a deep capsule network," *Remote Sensing*, vol. 11, no. 6, 2019, doi: 10.3390/RS11060626.

[20] X. Li, M. Khishe, and L. Qian, "Evolving deep gated recurrent unit using improved marine predator algorithm for profit prediction based on financial accounting information system," *Complex and Intelligent Systems*, vol. 10, no. 1, pp. 595–611, 2024, doi: 10.1007/s40747-023-01183-4.

[21] F. Alrowais *et al.*, "Enhancing neurological disease diagnostics: fusion of deep transfer learning with optimization algorithm for acute brain stroke prediction using facial images," *Scientific Reports*, vol. 15, no. 1, 2025, doi: 10.1038/s41598-025-97034-y.

[22] M. Alsieni and K. H. Alyoubi, "Artificial intelligence with feature fusion empowered enhanced brain stroke detection and classification for disabled persons using biomedical images," *Scientific Reports*, vol. 15, no. 1, 2025, doi: 10.1038/s41598-025-14471-5.

[23] N. N. Razlan, R. Aminuddin, N. Sabri, S. Ibrahim, and A. A. Shari, "Visual facial paralysis detection using InceptionResNetV2," in *2024 IEEE International Conference on Automatic Control and Intelligent Systems, I2CACIS 2024 - Proceedings*, 2024, pp. 211–215, doi: 10.1109/I2CACIS61270.2024.10649871.

[24] S. Ban, H. S. Nam, and E. Park, "Detecting paralysis of stroke symptom in video: Transfer learning with gated recurrent unit using public big data of facial images," in *Proceedings - 2022 IEEE International Conference on Big Data, Big Data 2022*, 2022, pp. 6587–6589, doi: 10.1109/BigData55660.2022.10020374.

[25] A. W. Abulfaraj, A. K. Dutta, and A. R. W. Sait, "Ensemble learning-based brain stroke prediction model using magnetic resonance imaging," *Journal of Disability Research*, vol. 3, no. 5, p. 302, May 2024, doi: 10.57197/JDR-2024-0061.

[26] Z. Shagdar, A. H. Farmen, S. A. Amirshahi, and K. Raja, "A survey on computer vision-based automatic assessment of stroke and facial palsy," *IEEE Access*, vol. 13, pp. 29613–29632, 2025, doi: 10.1109/ACCESS.2025.3540658.

[27] R. Singh and N. Sharma, "Automated detection of facial palsy from facial images using convolutional neural networks," in *Proceedings of the 2024 International Conference on Artificial Intelligence and Emerging Technology, Global AI Summit 2024*, 2024, pp. 930–935, doi: 10.1109/GlobalAISummit62156.2024.10947880.

BIOGRAPHIES OF AUTHORS

Kiruthiga Subramaniyan 🆔 🔍 sc ◐ is is currently working as Assistant Professor in CSE at Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai. She obtained her Master's degree from Annamalai University and Bachelor's degree from Oxford Engineering College, Trichy. She has 11 years of teaching experience. She has published 18 papers in various international journals and Conferences. Her areas of interest include image processing, IoMT, big data and deep learning. She has attended and organized various International Symposiums, Seminars and Workshops; FDP related to her area of research. She is a Life Member in IAENG and ISTE. She can be contacted at email: kiruthigaresearchscholar@gmail.com.

Chinnasamy Anbuananth 🆔 🔍 sc ◐ is working as an Associate Professor in the Department of Computer Science and Engineering in Annamalai University, Chidambaram, India. He received Ph.D. degree in 2015 and Master of Engineering (ME) degree in 2007 from Annamalai University. He is having 21 years of experience and his research interests are wireless networks, big data analytics and image processing. He has published more than 38 articles in reputed journals and many articles presented in international conferences. He has organized 6 national level Workshop, Seminar, Hands-on-training programs funded by DST, UGC, and RUSA. He has been affiliated with professional bodies such as ISTE, CSI and IAENG as a life member. He has guided 5 research scholars and has been guiding 4 research scholars. He can be contacted at email: anbu_ananth2006@yahoo.com.

Dhilip Kumar Venkatesan 🆔 🔍 sc ◐ was awarded Ph.D. at North Eastern Hill University (A Central University of INDIA) in 2017. Presently, he is working as an associate professor in the Department of Computer Science and Engineering at Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai. He has more than 10+ year's experience of teaching as well as research. He did his B.Tech. Information Technology and M.E. Computer Science Engineering under Anna University, Chennai. He has published various international journals and international conferences in the field of vehicular communication and machine learning. He is an editorial board member and reviewer of various international journals and conferences in reputed publishers like Springer, IEEE, Elsevier, and Wiley. His area of interest is vehicular communication, soft computing techniques, internet of things, and machine learning. He is presently working on projects related to artificial intelligence for healthcare and automations. He can be contacted at email: dhilipkumar@gmail.com.