IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 10 June 2023, accepted 17 July 2023, date of publication 20 July 2023, date of current version 26 July 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3297281

RESEARCH ARTICLE

Intracranial Haemorrhage Diagnosis Using Willow Catkin Optimization With Voting Ensemble Deep Learning on CT Brain Imaging

NOHA NEGM[1], GHADAH ALDEHIM[2], FAISAL MOHAMMED NAFIE[3], RADWA MARZOUK[1], MOHAMMED ASSIRI[4], MOHAMED IBRAHIM ALSAID[5], SUHANDA DRAR[5], AND SITELBANAT ABDELBAGI[5]

[1]Department of Computer Science, College of Science and Art at Mahayil, King Khalid University, Abha 62529, Saudi Arabia
[2]Department of Information Systems, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, Riyadh 11671, Saudi Arabia
[3]Department of Natural and Applied Sciences, Community College, Majmaah University, Al Majma'ah 11952, Saudi Arabia
[4]Department of Computer Science, Al-Aflaj College of Sciences and Humanities, Prince Sattam bin Abdulaziz University, Al-Aflaj 16273, Saudi Arabia
[5]Department of Computer and Self Development, Preparatory Year Deanship, Prince Sattam bin Abdulaziz University, Al-Kharj 16278, Saudi Arabia

Corresponding author: Mohammed Assiri (m.assiri@psau.edu.sa)

The authors extend their appreciation to the Deanship of Scientific Research at King Khalid University for funding this work through Large Groups Project under grant number (RGP2/96/44). Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2023R387), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia. The authors would like to thank the Deanship of Scientific Research at Majmaah University for supporting this work under Project No. R-2023-520. This study is supported via funding from Prince Sattam bin Abdulaziz University project number (PSAU/2023/R/1444).

ABSTRACT Intracranial haemorrhage (ICH) has become a critical healthcare emergency that needs accurate assessment and earlier diagnosis. Due to the high rates of mortality (about 40%), the early classification and detection of diseases through computed tomography (CT) images were needed to guarantee a better prognosis and control the occurrence of neurologic deficiencies. Generally, in the earlier diagnoses test for severe ICH, CT imaging of the brain was implemented in the emergency department. Meanwhile, manual diagnoses are labour-intensive, and automatic ICH recognition and classification techniques utilizing artificial intelligence (AI) models are needed. Therefore, the study presents an Intracranial Haemorrhage Diagnosis using Willow Catkin Optimization with Voting Ensemble (ICHD-WCOVE) Model on CT images. The presented ICHD-WCOVE technique exploits computer vision and ensemble learning techniques for automated ICH classification. The presented ICHD-WCOVE technique involves the design of a multi-head attention-based CNN (MAFNet) model for feature vector generation with optimal hyperparameter tuning using the WCO algorithm. For automated ICH detection and classification, the majority voting ensemble deep learning (MVEDL) technique is used, which comprises recurrent neural network (RNN), Bi-directional long short-term memory (BiLSTM), and extreme learning machine-stacked autoencoder (ELM-SAE). The experimental analysis of the ICHD-WCOVE approach can be tested by a medical dataset and the outcomes signified the betterment of the ICHD-WCOVE technique over other existing approaches.

INDEX TERMS Brain imaging, intracranial haemorrhage, deep learning, computer vision, ensemble learning, willow catkin optimization.

I. INTRODUCTION

Stroke is the main reason for a higher number of death across the world. It may occur if there is any interruption in the blood supply to brain parenchyma because of the rupture of blood vessels (hemorrhagic stroke) or occlusion (ischemic stroke) [1]. Hemorrhagic stroke or Intracerebral hemorrhage (ICH) occurs whenever bleeding takes place in the cerebral parenchyma because of ruptured blood vessels. Drug abuse, Weak blood vessels, trauma, and hypertension were usually inducing these medical conditions [2]. ICH can be considered a neurologic emergency and it has numerous

The associate editor coordinating the review of this manuscript and approving it for publication was Edith C.-H. Ngai.

subtypes namely pons, caudate nucleus, or basal ganglia. The haemorrhage types were usually liable on the anatomic location of the bleeding [3]. A timely and early ICH diagnosis is important since this condition can worsen the patient's health in the first few hours of occurrence [4]. The imaging modality named Noncontrast head computer tomography (CT) was leveraged for detecting haemorrhage due to its speed and wide availability. Such a modality has revealed a high specificity and sensitivity in the detection of acute haemorrhage [5].

Computer-aided diagnosis (CAD) refers to cutting-edge technology in the medical field that interfaces medicine and computer science [6]. CAD schemes are similar to the skilled human expert to do diagnostic decisions through diagnostic rules. The CAD system performance can be enhanced and advanced CAD can infer novel knowledge through analyses of the medical dataset [7]. For learning these capabilities, the mechanism should have a feedback system where the learning occurs through failures and successes. Over the past, dramatic enhancement is found in human skills and examination tools like CT, X-ray, ultrasound, and MRI [8]. The diagnosis seems to be more complex with the study and discovery of novel diseases and their development. Several aspects like complicated medical diagnosis, accessibility of massive data concerning diseases and conditions in the medical field, rising knowledge on diagnostic rules, and the appearance of innovative areas like data mining, AI, and ML in the computer science field have paved the way for the CAD development [9]. Quantitative analysis of pathology imageries has increased significantly among the research community in the arena of image analysis and pathology. There comes the demand for quantitative image-based assessment of pathological slides since the diagnosis solely depends on the pathologist's opinion. CAD decreases the pathologist's burden by filtering the benign cancer imageries thereby the diagnosticians can concentrate on more complex imageries which is tough to detect. Quantitative analysis of pathology imageries will be helpful in diagnosis and also in medical research. Lately, deep learning (DL) has risen effectively and rapidly [10]. DL-related networks have a great generalization ability that can be enforced to overcome challenging medical complexities like medical organ detection, medical image classification, disease detection, and medical image analysis. Convolutional neural network (CNN) is the most effective network than deep network.

This study presents an Intracranial Haemorrhage Diagnosis using Willow Catkin Optimization with Voting Ensemble (ICHD-WCOVE) Model on CT images. The presented ICHD-WCOVE technique exploits computer vision and ensemble learning techniques for automated ICH classification. The presented ICHD-WCOVE technique involves the design of a multi-head attention-based CNN (MAFNet) model for feature vector generation with optimal hyperparameter tuning using the WCO algorithm. For automated ICH detection and classification, the majority voting ensemble deep learning (MVEDL) technique is used, which comprises recurrent neural network (RNN), Bi-directional long short-term memory (BiLSTM), and extreme learning machine-stacked autoencoder (ELM-SAE). The experimental analysis of the ICHD-WCOVE algorithm can be tested by a medical dataset. In short, the key contribution of the paper is listed as follows.

- An automated ICHD-WCOVE technique comprising MAFNet feature extraction, WCO based hyperparameter tuning, and ensemble classification has been presented for ICH diagnosis. To the best of our knowledge, the ICHD-WCOVE technique for ICH classification never existed in the literature.
- Employ MAFNet model for feature extraction, which enables to capture various aspects of the input data, improving the network's ability in the extraction of useful features and enhance its performance.
- Hyperparmaeter optimization of MAFNet model using WCO algorithm helps to boost the overall ICH classification performance.
- Present ensemble model, comprising RNN, BiLSTM, ELM-SAE model for classification process. It leverages the strengths of multiple models, improving accuracy, robustness, and generalization capabilities.

II. LITERATURE REVIEW

Kothala et al. [11] developed a novel YOLOv5x-GCB architecture that is capable of detecting multiple haemorrhages with a lack of availability of resources by using the ghost convolution technique. It can be benefited from generating a similar amount of feature maps as vanilla convolution while applying lower-cost linear operation. An additional feature is that it employs a mosaic augmentation method during the training for improving the performance of mixed haemorrhage recognition. Kumaravel et al. [12] introduced a model that leverages the capability of pretrained DCNN for the recognition of ICH in CT scans. Initially, the presented method including the AlexNet, pretrained DCNN, was used for the classification and feature extraction using the TL approach. Then, an adapted AlexNet- SVM classification is considered, and lastly, a feature selection technique, PCA, was proposed in the AlexNet-SVM classification method, and also its efficiency is considered.

Anupama et al. [13] developed a DL-based ICH analysis with GrabCut-based segmentation using synergic DL (SDL), termed the GC-SDL approach. The presented technique uses GF for the removal of the noise, thus the image quality could be enhanced. Additionally, the GrabCut-based segmentation method is used for efficiently identifying the infected portion in the CT images. The SDL technique is exploited to implement the feature extraction method, and eventually, the softmax (SM) function is exploited as a classifier. Kirithika et al. [14] designed an ensemble of deep and handcrafted features for ICH and BT detection. The input image preprocessing can be performed in three different ways such as CLAHE based contrast enhancement, skull

stripping and bilateral filtering (BF). Furthermore, AlexNet and SIFT techniques are exploited for the process of feature extraction. To categorize the presence of ICH and BT, two classifier methods are performed namely random forest (RF) and Gaussian Naïve Bayes (GNB).

Le et al. [15] introduced a novel strategy for brain hemorrhage detection with object detection techniques such as R-FCN and Fast RCNN. The presented technique could identify various areas of a brain hemorrhage on CT images. The result shows that the R-FCN technique provides the best outcome than the Fast RCNN model on accuracy and time of detection. Patel et al. [16] introduced the recognition of ICH in 3D non-contrast CT. The technique integrates an RNN and CNN in the form of a bi-directional LSTM for ICH detection at the image level. A CNN was trained for the recognition of ICH. Zhang et al. [17] formulated a new technique to produce artificial lesions on non-lesion CT images to generate labelled training instances. Artificial masks in any shape, size, or location can be made by Artificial Mask Generator (AMG) and later transformed into hemorrhage lesions via Lesion Synthesis Network (LSN).

Khan et al. [18] presented an automated system for brain tumor recognition and classification utilizing a saliency map and DL feature optimizer in this study. During the primary stage of the presented structure, a fusion-based contrast enhancement system was presented. An optimum features can be selected in the last stage utilizing an enhanced dragonfly optimizer algorithm (DOA). At last, an optimum features can be classified utilizing an ELM. In [19], Harris Hawks optimized CNN (HHOCNN) was utilized in this effort to solve these problems. The brain MRI are pre-processing, and the noisy pixels can be removed for minimizing the false tumor detection rate. Afterward, the candidate region procedure was executed for identifying the tumor area. Odusami et al. [20] presented a multi-modal fusion-based system which utilizes a mathematical approach termed as DWT for analyzing the data, and the optimizer of this method was realized by TL utilizing a pre-training NN named as VGG16. The last fused image was reconstructed utilizing inverse DWT (IDWT).

In the framework of ICH analysis utilizing CT brain imaging, there exist a research gap in searching and optimizing the hyperparameters of method utilized for analyzing. It contains parameters like rate of learning, regularized strengths, network structures, and dropout rate. Exploring the outcomes of various hyperparameter settings on the efficiency of methods are support recognize the optimum configuration which affects to enhanced efficiency on ICH analysis. Besides, an ensemble learning was extremely discovered in many fields, there could still exist a research gap in its application specifically for ICH analysis utilizing CT brain imaging. Research can concentration on examining the efficiency of distinct ensemble approach like bagging, boosting, or stacking, for combining several approaches trained on various subsets of the data or utilizing distinct techniques. It is support enhance the entire efficiency and robustness of ICH analysis by leveraging the collective knowledge of varied methods.

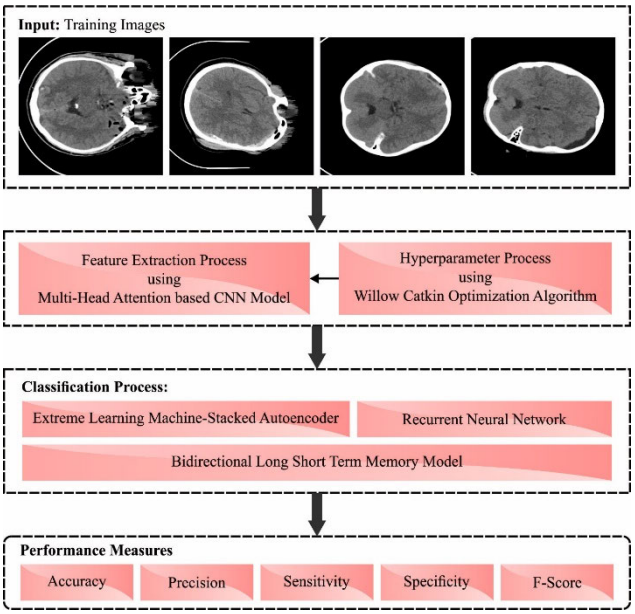

FIGURE 1. The overall flow of the ICHD-WCOVE approach.

III. THE PROPOSED MODEL

In this study, we have focused on the development of the ICHD-WCOVE model for ICH classification on the CT images. The presented ICHD-WCOVE technique exploits computer vision and ensemble learning techniques for automated ICH classification. The presented ICHD-WCOVE technique encompasses MAFNet feature extraction, WCO-based parameter tuning, and MVEDL-based classification. Fig. 1 signifies the overall flow of the ICHD-WCOVE approach.

A. FEATURE EXTRACTION: MAFNET MODEL

The presented ICHD-WCOVE technique comprises the MAFNet model for feature vector generation. The MAFNet is composed of $1 \times 1 \times 1$ convolutional layers [21]; in the middle are 4 convolution models with three convolution structures in every module, where the convolutional model was disseminated in a symmetrical structure of [2,2,2,2], and at last, in FC layer with the overall of 26 convolution modules. Next, the images are inputted, and the 1×1 convolutional process can be performed at first. Then, the 4 convolution models are passed over the 4 convolution models. The 3×3 convolutional operation was substituted with Contextual Transformer (CoT) blocks in the original ResNet convolution block. Then, the preceding convolution, excitation, and pooling operations, the feature extracted was inputted to the FC layers that performed as ''classification'' in the CNN.

The FC layers act as a ''classification'' in the CNN which incorporates the preceding, extremely abstracted

feature, maps the learned feature to sample space, employs *the Softmax* function to estimate the probability for the classification, and lastly output the classifier outcome:

$$Softmax\left(z_j\right) = \frac{e^{z_i}}{\Sigma_{i=1}^{n} e^{z_i}},$$

Eq. (1) illustrates the number of classes, z_j characterizes the output value of $j - th$nodes, z_i indicates the output value of $i - th$nodes, and e shows the natural constant and it is given in the following equation:

$$L_{CE} = -\sum_{i=1}^{N} l_i log p_i,$$

In Eq. (2), N signifies the number of tags) and l_i designates the thermal encoding of tag $i(i\in (0, \cdots, N-1)$ once the target label was i then $l_i = 1$, other labels were corresponding to zero , p_i represents the predictive probability of i^{th} labels, i.e., Softmax value.

Note that the smallest learning rate might lead to slower convergence, and the largest learning rate might lead to the constant oscillation of loss functions. A dynamic-learning-rate approach can be used for adjusting the learning rate per 30 epochs for making a model to find the optimal parameter accurately and more quickly trained as follows:

$$lr = lr_0 \times 0.1^{\frac{epoch}{30}},$$

where lr represents the current learning rate, lr_0 shows the primary learning rate, and *epoch* characterizes the overall amount of training rounds.

B. PARAMETER TUNING: WCO ALGORITHM

At this stage, the WCO technique is applied for the optimal hyperparameter selection of the MAFNet method. The WCO was presented as the novel meta-heuristic optimizer technique. Hybrid exploration and exploitation methods are presented in WCO [22]. The particles can be separated into 2 parts from the searching procedure, and exploitation and exploration can be started simultaneously. The particles from distinct procedures are various performances and parameters.

During this initialized stage, many random seeds can be created in the willow tree and disseminated from the solution space. Utilizing a matrix with D columns and N rows, the population was demonstrated where N implies the particle count and D represents the data dimensional of every particle. The algorithm utilizes Eq. (1) for generating random particles.

$$\chi_i = r \times (UB - LB) + LB, i = 1, 2, \ldots, N \tag{1}$$

In Eq. (1), r denotes the random number from the interval of 0 and 1.x_i signifies the solution of D dimensional. The lower and upper bounds of solution spaces are LB and UB.

In the search stage, a willow catkin falls in the willow trees and flutters from the wind which is all affected by 2 variables like wind speed and direction. Airflow from the environment takes either wind speed or wind direction. The wind was signified mathematical with a vector. Thus, the meteorological wind vector changed to "math" wind direction. It was general practice in meteorology for working with u and v elements. The vector components of winds, u and v are achieved when the wind direction and speed are measured, as:

$$\begin{cases} v = -ws \times cos(wd) \\ u = -ws \times sin(wd) \end{cases} \tag{2}$$

In Eq. (2), wd signifies the wind direction, andws refers to the wind speed. The transformation of u, and v elements as wind speed and direction on a 2D plane.

Afterwards decomposing the wind direction and speed for obtaining u and v, the particles are upgraded. The particle upgrade was mostly compared with the wind speed and direction. The particle upgrade was carried out in the exploration stage utilizing Eq. (3).

$$x = x + a \times (u \times v) + (2 - a)\left(P_{g^{-}\chi i}\right) \tag{3}$$

whereas x_i denotes the individual's present location and a refers to the variable which controls the change from exploration to exploitation. P_g denotes global optima under the present iteration. By balancing exploration and exploitation, a will change during the iteration process:

$$a = c \times e^{-(\frac{t}{1000})^2} \tag{4}$$

In Eq. (4), T denotes the maximal amount of iterations. c represents the constant with value 2. t shows the existing amount of iterations. If $t < 0.4T$, as was higher than 1, the particle update will be impacted by wind, which makes particle position further random.

The distance di between the global optimum and the current individual is evaluated. Based on the easier adhesion radius R, When $di <= R$, the two seeds were liable to stick together. When $di > R$, then two seeds are lesser likely to stick together.

$$\begin{cases} ws = r \times R \\ wd = r \times 2\pi \end{cases} \tag{5}$$

Now d_i indicates the distance of the particle from optimum global solutions. R indicates the adhesion radius of willow catkins. ws is produced by Eq. (5) if $d_i > R$. r denotes the random integer within [0, 1]. Bringing the wd and ws randomly generated is define the distance of movement and individuals' direction, such that every individual with a distance of more than R would be moved at random, enhancing the exploration capability of the model and prevent from getting trapped in local optima.

$$DW = 1 - \frac{\left|p_g - x_i\right|}{\left\|\chi_i - p_g\right\|} \tag{6}$$

$$K = \frac{DW}{\Sigma_{i=1}^{D} DW_i} \tag{7}$$

$$\begin{cases} ws = \mu \times (\Sigma_{i=1}^{D} K_i |p_g - x_i|) + (1-\mu) \times r_2 \times R \\ wd = \arccos(\frac{x_i \cdot p_g}{||x_i|| \times ||p_g||}) + r_3 \times \frac{\pi}{8} \end{cases} \tag{8}$$

When $d_i <= R$, the weight can be evaluated using Eqs (6) and (7), and the wind direction and wind speed are evaluated using Eq. (8). DW represents the weights of every dimensional distance in p_g and x_i to the overall distance and $||\cdot||$ represents the Euclidean distance between p_g and χ_j. Eq. (8) normalize $.\mu$, representing the random value from 0.4 to 0.6. r_2 and r_3 denote the random number within [0, 1].

Algorithm 1 Pseudocode of WCO Algorithm

```
Require Population size N, Fitness function, Dimension D, Max_iteration
T, Upper and Lower boundaries UB and LB,
Ensure: Global optimum value
Initializing the Gbest, GlobalBestPos and Pop of every
population
While (t <MaxIteration or satisfy the minimum
threshold value) d
        a = Equation
For = 1: N d
If R_i > Rz then
Produce ws and wdz based on Equation
Else
Produce ws and wdz based on Eqs (6)-(8)
End if
Update Pop
Evaluate the fitness value of the population
Upgrade GlobalFmin and GlobalBes
End for
End while
```

The WCO method has derived a fitness function to obtain maximum accuracy of the classification and also described a positive integer to characterize the superior efficacy of the solution candidate. The reduction of the classification error rate is taken as a fitness function.

$$fitness\,(x_i) = ClassifierErrorRate\,(x_i) = \frac{number\ of\ misclassified\ samples}{Total\ number\ of\ samples} * 100 \quad (9)$$

C. ENSEMBLE LEARNING-BASED CLASSIFICATION: MVEDL MODEL

For automated ICH detection and classification, the MVEDL technique is used, which comprises ELM-SAE, RNN, and BiLSTM models. The MVEDL exploits the confidence preservation mode for increasing the performance of ICH classification [23]. The class prediction of the MVEDL technique equivalent to the ICH classification can be assessed by Algorithm 1. Hard and Soft voting systems have been agglutinated owing to the number of base learners utilized, to resolve the probability of an even number of output predicted Ξ_j. The average weight confidence probability $\mu_{\overline{W}_j}$ of each Ξ_j is expressed as follows:

$$\mu_{\overline{\omega}_j}, = \frac{1}{n}\sum_{j}^{n} \Xi_j \quad (10)$$

1) ELM-SAE MODEL

The deep architecture captures abstract and high-level features. One of the direct ways to construct the deep

Algorithm 2: Pseudocode of MVEDL

```
Given an input image i
Transfer iz to the corresponding handler (b_1, b_2, ...b_n)z of the learner.
Calculate prediction for every handler based on distributed processing.
Compile responses from every handler of the learner.
Calculate ϖ_j per Ξ_j
Aggregate the outcomes of the handler with ϖ_j >= 0.25.
If accurate one class K_j has the maximum predictable outputs Ξ_j
        P(K_j) = K_j
Else
P(K_j) = K_j with the highest average weighted confidence μ_ϖj
        End
```

ELM (DELM) is to cascade the feature layer randomly. But the randomness requires guidance and suffers from unpredictability [24]. The ELM-SAE is used with l_1-norm constraint for finetuning the weight between layers of DWELM to accomplish meaningful and more compact representation. Fig. 2 showcases the framework of ELM-SAE. The input parameter of ELM-SAE is produced at random. An ELM-SAE is an FFNN whose input is similar to the output. By resolving the optimization problems, the output parameter can be upgraded:

$$\arg\min_{\hat{W}} ||Z\hat{W} - X||_2^2 + \lambda||\hat{W}||_1, \quad (11)$$

In Eq. (11), Z represents the sparse feature and X represents the input and output of $ELM - SAE$. $\hat{W}$ represents the output weight of $ELM - SAE$. The abovementioned formula is named as Lasso problem where a faster iterative shrinkage thresholding algorithm (FISTA) model and alternating direction technique of multipliers (ADMM) is used to effectively resolve the problem:

$$\arg\min_{\hat{W}} f\left(\hat{W}\right) + g\left(\hat{W}\right), \hat{W} \in \mathrm{R}^{n\times L}, \quad (12)$$

where $f(\hat{W}) = ||Z\hat{W} - X||$ and $g(\hat{W}) = \lambda||\hat{W}||_1$. By applying ADMM, the abovementioned problem is expressed by Eq. (13):

$$\arg\min_{\hat{W}} f(\hat{W}) + g(O), s.t. W - O = 0. \quad (13)$$

The proximal problem can be resolved by:

$$\begin{cases} \hat{W}_{j+1} := (Z^T Z + \rho I)^{-1}(Z^T X + \rho(O_j - V_j)) \\ O_{j+1} := S_{\frac{\lambda}{\rho}}(\hat{W}_{j+1} + V_j) \\ V_{j+1} := V_j + (\hat{W}_{j+1} - O_{j+1}) \end{cases} \quad (14)$$

where $\rho > 0$ represents the regularization parameter, I indicate the unit matrix, j denotes the iteration number, and $S_{\lambda/\rho}$ represent the soft thresholding operator, given as follows:

$$S_\kappa(\alpha) = \begin{cases} \alpha - \kappa, \alpha > \kappa \\ 0, |\alpha| \le \kappa \\ \alpha + \kappa, \alpha < -\kappa. \end{cases} \quad (15)$$

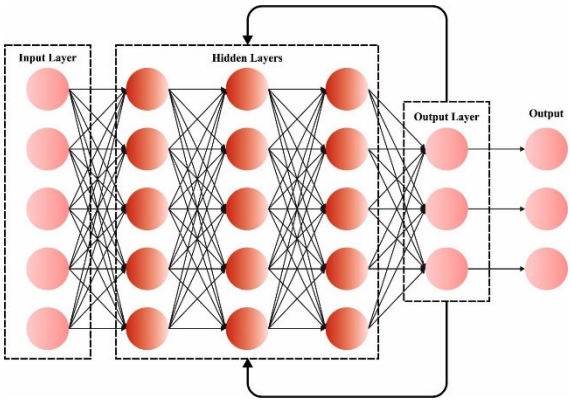

FIGURE 2. Structure of ELM-SAE.

2) RNN MODEL

Due to the effects of reducing gradients, training RNN to capture long-term dependency is more complicated, thus the frequently used RNN unit is the LSTM which does not prevent free gradient flow and provides the "constant error carousel" [25]. The most common LSTM architecture comprises a recurrent cell, output, input and forget gates. Generally, the LSTM cell is well-arranged in the form of a chain structure, with an output of the previous LSTM connected with the input of subsequent LSTM. GRU encompasses separate "update gates" instead of input and single forget gates, does not discriminate hidden and cell states, and frequently discloses the entire hidden layer (HL). The RNN is an NN model that includes cyclic connection which enables one to learn the temporal dynamics of successive data as follows:

$$h_t = F\left(W_h h_{t-1} + \mathrm{U}_h x_t + b_h\right) \tag{16}$$

$$y_t = F\left(W_y h_t + b_y\right) \tag{17}$$

An HL in RNN encompasses various nodes. Each node has a function which is used to generate the yt current output and ht hidden state using its preceding hidden state ht−1 and current input xt. From the expression, Wh, Uh, & Wy correspondingly indicate weight for hidden to hidden recurrent connection, input to hidden connection, and hidden to output connection. b_h and b_y represents bias and hidden output state. Furthermore, it has an activation function F connected with all the nodes. This is a component-wise nonlinearity function, usually chosen from different functions like ReLu, sigmoid, or hyperbolic tangent.

3) BILSTM MODEL

Here, the Bi-LSTM was proposed to execute the detection of ICH [26]. Bi-LSTM is particularly helpful to learn the sequence with a pattern of unknown length. Also, the stacked recurrent HL in NN is used to confine the structure of a time series. To develop the gradient vanishing problems as the LSTM utilizes the cell state of the memory for data transmission. A cell state was suitable to calculate a dataset that isn't used for a longer duration. Hence, LSTM comprises Reset, Forget and Update gates. Furthermore, some key components of LSTM are given below.

- Input Gate: A multiplicative unit that defends data saved in CEC from irregular inputs.
- Output Gate: A multiplicative unit defends another unit from interruption through content retained in CEC.
- Constant error carousel (CEC): A fundamental component uses unit weight and repeated connection. Then, the recurrent connection shows a time step 1 and a feedback loop.

At present, input and output gates manage access control for CEC. During the training phase, the input gate can be recognized and enables a new dataset in the CEC. The data is not connected when the input gate is zero. Likewise, the output gate was analysed and allowed the dataset to be obtained from CEC. Moreover, if the data get stuck in the memory cell, then the gate is closed. Without considering diminishing gradients, it permits the error signals to flow many times. It is a trivial data sequence. When the input stream is separated physically, then the LSTM state isn't organized and changed to the appropriate sequence. Especially, LSTM will learn to reset memory cells as a sequence is completed and obtain a new sequence. These problems are resolved by the LSTM using forget gate. As well, the Bi − LSTM was an extended model of LSTM where both LSTM models are used to input the dataset. The maximized learning of long-term dependency enhances the efficiency of the network in the application of LSTM.

TABLE 1. Details of the dataset.

Class	No. of Instances
Epidural	171
Intraventricular	24
Intraparenchymal	72
Subdural	56
Subarachnoid	18
Total Number of Instances	341

IV. RESULTS AND DISCUSSION

The proposed model is simulated using Python 3.6.5 tool. The proposed model is experimented on PC i5-8600k, GeForce 1050Ti 4GB, 16GB RAM, 250GB SSD, and 1TB HDD. The parameter settings are given as follows: learning rate: 0.01, dropout: 0.5, batch size: 5, epoch count: 50, and activation: ReLU. In this work, the ICH classification outcomes of the ICHD-WCOVE technique can be examined on the CT image dataset [27], comprising 341 samples with five classes as represented in Table 1. Fig. 3 showcases the sample images.

In Fig. 4, the ICH classifier outcomes of the ICHD-WCOVE technique can be examined in the form of a confusion matrix. The results indicate that the ICHD-WCOVE technique results in improvised performance in all classes.

In Table 2 and Fig. 5, the overall ICH classification results of the ICHD-WCOVE technique are studied on

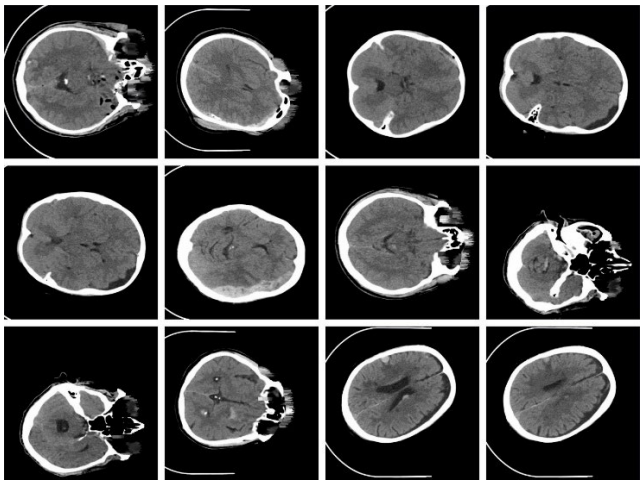

FIGURE 3. Sample images.

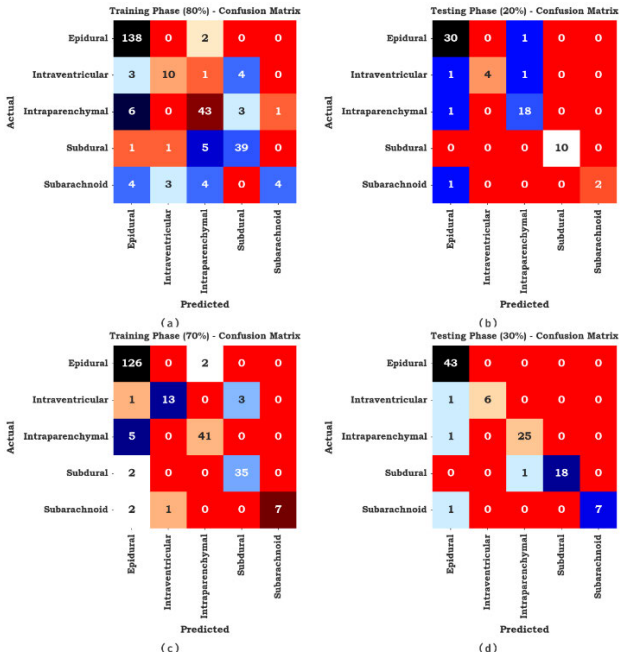

FIGURE 4. Confusion matrices of ICHD-WCOVE approach (a-b) TRS/TSS of 80:20 and (c-d) TRS/TSS of 70:30.

TABLE 2. ICH outcome of ICHD-WCOVE approach under 80:20 of TRS/TSS.

Class	Accuracy	Precision	Sensitivity	Specificity	F-Score
Training Phase (80%)					
Epidural	94.12	90.79	98.57	89.39	94.52
Intraventricular	95.59	71.43	55.56	98.43	62.50
Intraparenchymal	91.91	78.18	81.13	94.52	79.63
Subdural	94.85	84.78	84.78	96.90	84.78
Subarachnoid	95.59	80.00	26.67	99.61	40.00
Average	94.41	81.04	69.34	95.77	72.29
Testing Phase (20%)					
Epidural	94.20	90.91	96.77	92.11	93.75
Intraventricular	97.10	100.00	66.67	100.00	80.00
Intraparenchymal	95.65	90.00	94.74	96.00	92.31
Subdural	100.00	100.00	100.00	100.00	100.00
Subarachnoid	98.55	100.00	66.67	100.00	80.00
Average	97.10	96.18	84.97	97.62	89.21

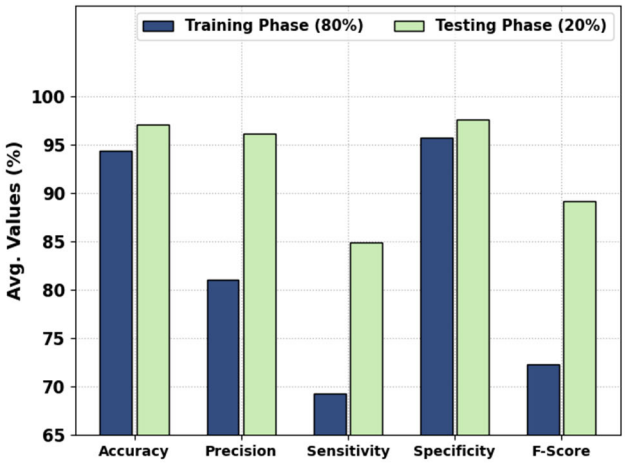

FIGURE 5. The average outcome of the ICHD-WCOVE method on 80:20 of TRS/TSS.

80:20 of TRS/TSS data. The results pointed out that the ICHD-WCOVE technique gains improvised performance under all class labels. For instance, with 80% of TRS, the ICHD-WCOVE technique obtains an average $accu_y$ of 94.41%, $prec_n$ of 81.04%, $sens_y$ of 69.34%, $spec_y$ of 95.77%, a F_{score} of 72.29%. On the other hand, with 20% of TSS, the ICHD-WCOVE method attains an average $accu_y$ of 97.10%, $prec_n$ of 96.18%, $sens_y$ of 84.97%, $spec_y$ of 97.62%, a F_{score} of 89.21%.

In Table 3 and Fig. 6, the overall ICH classification outcomes of the ICHD-WCOVE algorithm are studied on 70:30 of TRS/TSS data. The outcomes signified that the ICHD-WCOVE approach obtains improvised performance under all class labels.

For example, with 70% of TRS, the ICHD-WCOVE approach achieves an average $accu_y$ of 97.31%, $prec_n$ of 94.59%, $sens_y$ of 85.73%, $spec_y$ of 97.58%, and F_{score} of 89.43%. Instead, with 30% of TSS, the ICHD-WCOVE method gains an average $accu_y$ of 98.45%, $prec_n$ of 97.93%, $sens_y$ of 92.82%, $spec_y$ of 98.74%, and F_{score} of 95.14%.

The TACY and VACY of the ICHD-WCOVE method are investigated on ICH performance in Fig. 7. The figure designated that the ICHD-WCOVE technique has enriched performance with higher values of TACY and VACY. Visibly, the ICHD-WCOVE technique has higher TACY outcomes.

The TLOS and VLOS of the ICHD-WCOVE algorithm are tested on ICH performance in Fig. 8. The figure exhibits

TABLE 3. ICH outcome of ICHD-WCOVE approach under 70:30 of TRS/TSS.

Class	$Accu_y$	$Prec_n$	$Sens_y$	$Spec_y$	F_{score}
Training Phase (70%)					
Epidural	94.96	92.65	98.44	90.91	95.45
Intraventricular	97.90	92.86	76.47	99.55	83.87
Intraparenchymal	97.06	95.35	89.13	98.96	92.13
Subdural	97.90	92.11	94.59	98.51	93.33
Subarachnoid	98.74	100.00	70.00	100.00	82.35
Average	97.31	94.59	85.73	97.58	89.43
Testing Phase (30%)					
Epidural	97.09	93.48	100.00	95.00	96.63
Intraventricular	99.03	100.00	85.71	100.00	92.31
Intraparenchymal	98.06	96.15	96.15	98.70	96.15
Subdural	99.03	100.00	94.74	100.00	97.30
Subarachnoid	99.03	100.00	87.50	100.00	93.33
Average	98.45	97.93	92.82	98.74	95.14

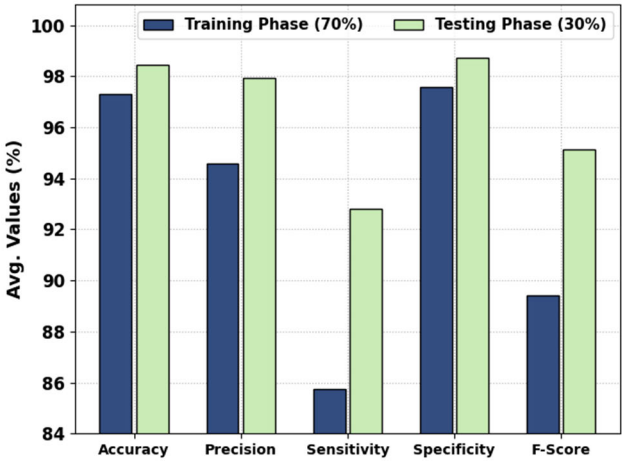

FIGURE 6. The average outcome of the ICHD-WCOVE approach on 70:30 of TRS/TSS.

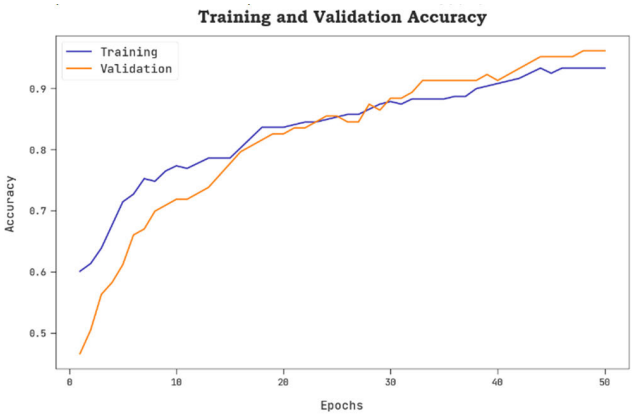

FIGURE 7. TACY and VACY outcome of the ICHD-WCOVE approach.

the ICHD-WCOVE method has superior performance with the minimal values of TLOS and VLOS. Notably, the ICHD-WCOVE algorithm has reduced VLOS outcomes.

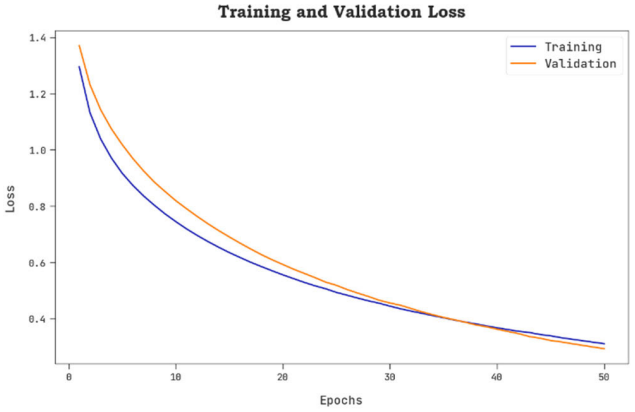

FIGURE 8. TLOS and VLOS outcome of the ICHD-WCOVE approach.

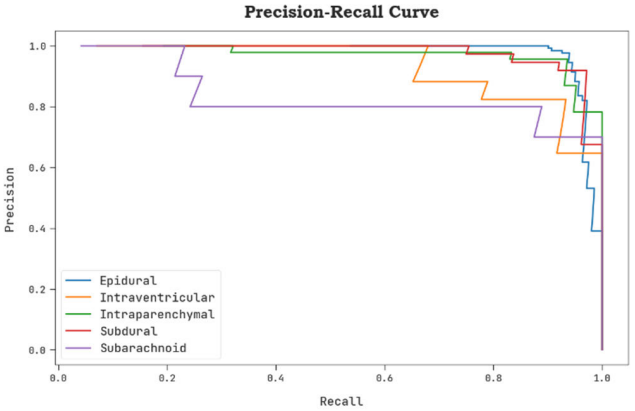

FIGURE 9. The precision-recall outcome of the ICHD-WCOVE method.

A brief precision-recall review of the ICHD-WCOVE method under the test database is shown in Fig. 9. The outcomes specified that the ICHD-WCOVE algorithm has enhanced values of precision-recall values in each class label.

The detailed ROC analysis of the ICHD-WCOVE process under the test database is given in Fig. 10. The outcomes signified the ICHD-WCOVE algorithm has the ability in classifying distinct.

To ensure the goodness of the ICHD-WCOVE algorithm, a comparative study is performed in Table 4 and Fig. 11 [28]. Based on $sens_y$, the ICHD-WCOVE technique reaches an increasing $sens_y$ of 92.82% while the AIMA-ICHDC, DL-ICH, AMG-LSN, RF, and SVM models obtain decreasing $sens_y$ of 92.05%, 91.22%, 91.37%, 90.68%, and 76.44% respectively. Concurrently, with $spec_y$, the ICHD-WCOVE method reaches an increasing $spec_y$ of 98.74% while the AIMA-ICHDC, DL-ICH, AMG-LSN, RF, and SVM methods obtain decreasing $spec_y$ of 91.83%, 91.60%, 90.57%, 89.63% and 79.63% correspondingly.

Simultaneously, with $accu_y$, the ICHD-WCOVE system reaches an increasing $accu_y$ of 98.45% while the AIMA-ICHDC, DL-ICH, AMG-LSN, RF, and SVM approaches gain decreasing $accu_y$ of 94.51%, 93.58%,

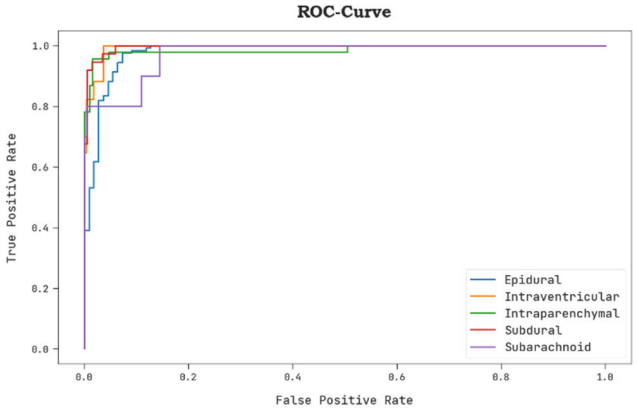

FIGURE 10. ROC curve outcome of ICHD-WCOVE approach.

TABLE 4. Comparative analysis of the ICHD-WCOVE method with other systems [28].

Methods	$Sens_y$	$Spec_y$	$Prec_n$	$Accu_y$
ICHD-WCOVE	92.82	98.74	97.93	98.45
AIMA-ICHDC	92.05	91.83	96.50	94.51
DL-ICH	91.22	91.60	95.57	93.58
AMG-LSN	91.37	90.57	93.77	92.95
Random Forest	90.68	89.63	89.30	89.99
SVM method	76.44	79.63	78.38	77.58

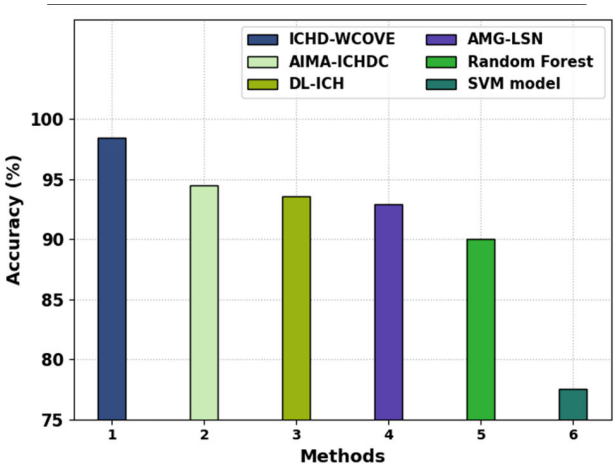

FIGURE 11. $Accu_y$ analysis of the ICHD-WCOVE approach with other systems.

92.95%, 89.99% and 77.58% correspondingly. These outcomes verified the improved performance of the ICHD-WCOVE method over other existing methods.

In summary, the ICHD-WCOVE approach demonstrates best outcome with maximal $accu_y$ of 98.45%. The improved performance of the ICHD-WCOVE is because of the integration of the WCO based hyperparameter tuning and ensemble learning method. Hyperparameters are settings which could not learned in the training, then need to be fixed prior to training. It could be an important effect on the performance of the model, and choosing the optimum values are cause to optimum accuracy. Ensemble learning is improve the entire accuracy of ICH classifier. With integrating the predictive of several approaches, ensemble approaches are mitigate the biases and errors of individual method, leads to further reliable and correct outcome. The ensemble is frequently demonstrate some single method or classification. It alleviate overfitting by decreasing the outcome of individual methods which can overfit certain features of trainied data. This leads to enhanced efficiency on unseen data that is vital to accurate ICH classifier. With integrating WCO based hyperparameter tuning and ensemble learning, the presented method was gain even optimum outcomes by choosing an optimum settings for this approach. These outcomes make sure the enhanced efficiency of the presented method over other existing approaches.

V. CONCLUSION

In this study, we have focused on the development of the ICHD-WCOVE model for ICH classification on the CT images. The presented ICHD-WCOVE technique exploits computer vision and ensemble learning techniques for automated ICH classification. The presented ICHD-WCOVE technique comprises a MAFNet model for feature vector generation with optimal hyperparameter tuning using the WCO algorithm. For automated ICH detection and classification, the MVEDL technique is used, which comprises ELM-SAE, RNN, and BiLSTM models. The experimental analysis of the ICHD-WCOVE technique is tested using a medical dataset and the outcomes signified the betterment of the ICHD-WCOVE technique over other existing approaches with maximum accuracy of 98.45%. In the future, the performance of the proposed method can be enhanced by feature fusion-based approaches. Besides, the computational complexity of the proposed model can be examined in future.

ACKNOWLEDGMENT

The authors extend their appreciation to the Deanship of Scientific Research at King Khalid University for funding this work through Large Groups Project under grant number (RGP2/96/44). Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2023R387), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia. The authors would like to thank the Deanship of Scientific Research at Majmaah University for supporting this work under Project No. R-2023-520. This study is supported via funding from Prince Sattam bin Abdulaziz University project number (PSAU/2023/R/1444).

REFERENCES

[1] L. El Mouna, H. Silkan, Y. Haynf, A. Tmiri, and A. Dahmouni, "Comparative study of deep learning models for detection and classification of intracranial hemorrhage," in *Proc. Int. Conf. Bus. Intell.* Cham, Switzerland: Springer, 2022, pp. 122–131.

[2] S. Kyung, K. Shin, H. Jeong, K. D. Kim, J. Park, K. Cho, J. H. Lee, G. Hong, and N. Kim, "Improved performance and robustness of multi-task representation learning with consistency loss between pretexts for intracranial hemorrhage identification in head CT," *Med. Image Anal.*, vol. 81, Oct. 2022, Art. no. 102489.

[3] X. Li et al., "The state-of-the-art 3D anisotropic intracranial hemorrhage segmentation on non-contrast head CT: The INSTANCE challenge," 2023, *arXiv:2301.03281*.

[4] M. Yeo, B. Tahayori, H. K. Kok, J. Maingard, N. Kutaiba, J. Russell, V. Thijs, A. Jhamb, R. V. Chandra, M. Brooks, and C. D. Barras, "Review of deep learning algorithms for the automatic detection of intracranial hemorrhages on computed tomography head imaging," *J. Neurointerventional Surg.*, vol. 13, no. 4, pp. 369–378, 2021.

[5] S. Matsoukas, J. Scaggiante, B. R. Schuldt, C. J. Smith, S. Chennareddy, R. Kalagara, S. Majidi, J. B. Bederson, J. T. Fifi, J. Mocco, and C. P. Kellner, "Accuracy of artificial intelligence for the detection of intracranial hemorrhage and chronic cerebral microbleeds: A systematic review and pooled analysis," *Radiologia Medica*, vol. 10, pp. 1106–1123, Oct. 2022.

[6] A. M. Dawud, K. Yurtkan, and H. Oztoprak, "Research article application of deep learning in neuroradiology: Brain haemorrhage classification using transfer learning," *Comput. Intell. Neurosci.*, vol. 2019, pp. 1–12, 2019, Art. no. 4629859, doi: 10.1155/2019/4629859.

[7] M. Altuve and A. Pérez, "Intracerebral hemorrhage detection on computed tomography images using a residual neural network," *Phys. Medica*, vol. 99, pp. 113–119, Jul. 2022.

[8] Q. Zhou, W. Zhu, F. Li, M. Yuan, L. Zheng, and X. Liu, "Transfer learning of the ResNet-18 and DenseNet-121 model used to diagnose intracranial hemorrhage in CT scanning," *Current Pharmaceutical Des.*, vol. 28, no. 4, pp. 287–295, 2022.

[9] H. Rane and K. Warhade, "A survey on deep learning for intracranial hemorrhage detection," in *Proc. Int. Conf. Emerg. Smart Comput. Informat. (ESCI)*, Mar. 2021, pp. 38–42.

[10] M. Burduja, R. T. Ionescu, and N. Verga, "Accurate and efficient intracranial hemorrhage detection and subtype classification in 3D CT scans with convolutional and long short-term memory neural networks," *Sensors*, vol. 20, no. 19, p. 5611, 2020.

[11] L. P. Kothala, P. Jonnala, and S. R. Guntur, "Localization of mixed intracranial hemorrhages by using a ghost convolution-based YOLO network," *Biomed. Signal Process. Control*, vol. 80, Feb. 2023, Art. no. 104378.

[12] P. Kumaravel, S. Mohan, J. Arivudaiyanambi, N. Shajil, andH. N. Venkatakrishnan, "A simplified framework for the detection of intracranial hemorrhage in CT brain images using deep learning," *Current Med. Image*, vol. 17, no. 10, pp. 1226–1236, 2021.

[13] C. S. S. Anupama, M. Sivaram, E. L. Lydia, D. Gupta, and K. Shankar, "Synergic deep learning model-based automated detection and classification of brain intracranial hemorrhage images in wearable networks," *Pers. Ubiquitous Comput.*, vol. 26, pp. 1–10, Nov. 2020.

[14] R. A. Kirithika, S. Sathiya, M. Balasubramanian, and P. Sivaraj, "Brain tumor and intracranial haemorrhage feature extraction and classification using conventional and deep learning methods," *Eur. J. Mol. Clin. Med.*, vol. 7, no. 7, pp. 237–258, 2020.

[15] T.-H.-Y. Le, A.-C. Phan, H.-P. Cao, and T.-C. Phan, "Automatic identification of intracranial hemorrhage on CT/MRI image using meta-architectures improved from region-based CNN," in *Proc. World Congr. Global Optim.* Cham, Switzerland: Springer, 2019, pp. 740–750.

[16] A. Patel, S. C. van de Leemput, M. Prokop, B. Van Ginneken, and R. Manniesing, "Image level training and prediction: Intracranial hemorrhage identification in 3D non-contrast CT," *IEEE Access*, vol. 7, pp. 92355–92364, 2019.

[17] G. Zhang, K. Chen, S. Xu, P. C. Cho, Y. Nan, X. Zhou, C. Lv, C. Li, and G. Xie, "Lesion synthesis to improve intracranial hemorrhage detection and classification for CT images," *Computerized Med. Image Graph.*, vol. 90, 2021, Art. no. 101929.

[18] M. A. Khan, A. Khan, M. Alhaisoni, A. Alqahtani, S. Alsubai, M. Alharbi, N. A. Malik, and R. Damaševičius, "Multimodal brain tumor detection and classification using deep saliency map and improved dragonfly optimization algorithm," *Int. J. Imag. Syst. Technol.*, vol. 33, no. 2, pp. 572–587, Mar. 2023.

[19] S. Z. Kurdi, M. H. Ali, M. M. Jaber, T. Saba, A. Rehman, and R. Damaševičius, "Brain tumor classification using meta-heuristic optimized convolutional neural networks," *J. Personalized Med.*, vol. 13, no. 2, p. 181, Jan. 2023.

[20] M. Odusami, R. Maskeliūnas, and R. Damaševičius, "Pixel-level fusion approach with vision transformer for early detection of Alzheimer's disease," *Electron.*, vol. 12, no. 5, p. 1218, 2023.

[21] F. Hasan and H. Huang, "MALS-Net: A multi-head attention-based LSTM sequence-to-sequence network for socio-temporal interaction modelling and trajectory prediction," *Sensors*, vol. 23, no. 1, p. 530, 2023.

[22] J. S. Pan, S. Q. Zhang, S. C. Chu, H. M. Yang, and B. Yan, "Willow Catkin Optimization algorithm applied in the TDOA-FDOA joint location problem," *Entropy*, vol. 25, no. 1, p. 171, 2023.

[23] D. A. Okuboyejo and O. O. Olugbara, "Classification of skin lesions using weighted majority voting ensemble deep learning," *Algorithms*, vol. 15, no. 12, p. 443, 2022.

[24] L. Wang, H. Cao, and Y. Fu, "A bearing prognosis framework based on deep wavelet extreme learning machine and particle filtering," *Appl. Soft Comput.*, vol. 131, 2022, Art. no. 109763.

[25] M. A. Alohali, F. N. Al-Wesabi, A. M. Hilal, S. Goel, D. Gupta, and A. Khanna, "Artificial intelligence enabled intrusion detection systems for cognitive cyber-physical systems in industry 4.0 environment," *Cogn. Neurodynamics*, vol. 16, pp. 1045–1057, Jan. 2022.

[26] S. Minaee, E. Azimi, and A. Abdolrashidi, "Deep-sentiment: Sentiment analysis using ensemble of CNN and bi-LSTM models," 2019, *arXiv:1904.04206*.

[27] M. Hssayeni. (Mar. 2021). *Computed Tomography Images for Intracranial Hemorrhage Detection and Segmentation*. [Online]. Available: https://physionet.org/content/ct-ich/1.3.1/

[28] F. Meng, J. Wang, H. Zhang, and W. Li, "Artificial intelligence-enabled medical analysis for intracranial cerebral hemorrhage detection and classification," *J. Healthcare Eng.*, vol. 2022, Mar. 2022, Art. no. 2017223.