

A NOVEL VITBiLSTM DEEP LEARNING FRAMEWORK FOR BRAIN HEMORRHAGE PREDICTION USING CT BRAIN IMAGES

Delveen Luqman Abd Alnabi^{*1}

¹College of Administration and Economic, University of Duhok, Duhok, Kurdistan Region, Iraq.

^{*}Corresponding Author email: delveen.luqman@uod.ac

Received: 22 Feb 2025

Accepted: 13 Apr 2025

Published: 19 Aug 2025

<https://doi.org/10.25271/sjuoz.2025.13.3.1488>

ABSTRACT:

Bleeding in the surrounding tissues of the human brain is called a brain hemorrhage. This problem can lead to stroke and even death. It requires fast intervention and accurate treatment to save a patient's life. Current state-of-the-art methodologies to detect this issue benefit from the development in the artificial intelligence field, especially its sub-field "deep learning". This study introduces a new deep learning-based framework to detect brain hemorrhage inside CT brain images. The proposed model is a novel hybrid model of vision transformer models and the bidirectional long short-term memory and is denoted as "ViTBiLSTM". The study utilizes two datasets, which are different in size and challenging. The first dataset consists of 6772 CT images, while the second one contains 2500 CT images. The study also compares the original vision transformer model with the proposed one. Besides that, the study utilizes different optimizers and compares the current research with the related work. Results show that the proposed ViTBiLSTM achieves its best performance when using the RMSProp optimizer with an accuracy of 100% and 96.94% on both datasets. Comparison with the current state of the art shows that the proposed methodology's performance exceeds the best study by 3.7% in accuracy.

KEYWORDS: Artificial Intelligence, Deep Learning, Vision Transformer, BiLSTM, Brain Hemorrhage

1. INTRODUCTION

When the blood vessels in the area around the brain burst, the surrounding tissues of the brain will bleed, leading to a critical case called "Brain Hemorrhage" (Grey, 2024; Rather *et al.*, 2024; Sheikh *et al.*, 2024). The increased pressure on the brain caused by such bleeding can cause severe damage to the brain cells and may even lead to death (Ahmed *et al.*, 2024; Neethi *et al.*, 2024; Schiariti *et al.*, 2024; Zhang *et al.*, 2024). A stroke problem that a brain hemorrhage can cause is considered the second leading cause of death around the world. However, the three main causes of such issues are metabolic risks, behavioural factors, and environmental risks (Feigin *et al.*, 2022; Neethi *et al.*, 2024). Mental changes, headaches, difficulty in speaking, weakness, lack of balance, and even vision issues are all symptoms of stroke caused by hemorrhage (Akmaljon *et al.*, 2024; Suryadi, 2024; Studer & Thompson, 2024;).

Deep learning is one of the most powerful artificial intelligence technologies that has been recently used for the aim of brain hemorrhage detection and prediction to help physicians correctly and effectively detect this problem, and further make the appropriate treatment (Del Gaizo *et al.*, 2024; Majeed *et al.*, 2024; Haldorai *et al.*, 2024; Hu *et al.*, 2024). Convolutional neural networks (CNN) and their newer versions, including VGG16, ResNets, MobileNet, DenseNet and Inception (Ibrahim & Mahmood, 2023). Deep learning-based classification models, especially the models that use the techniques of learning transfer such as, DenseNet121, have shown high accuracy the classification of the medical image, these classification models improve accuracy in diagnosis by utilizing pre-trained architectures which help to extract relevant features from the CT scans and differentiate between diseases and non-diseases cases. Ahmed *et al.* (2024) and Prasher *et al.* (2024) proposed very effective models to extract the image features that can be used in a classification framework to make an automatic brain hemorrhage detection system. On the other hand, results from Murad *et al.* (2023) show that CNNs have been

very effective in the automatic classification of medical images, specifically for CT scan images where disease detection is crucial. By automatically extracting relevant features from these images, CNNs improve accuracy by ensuring appropriate conditions are assigned promptly. However, some recent pieces of research started to make a fusion of the CNN-based model and other sequence processing models like the recurrent neural networks (RNN) (Datta & Rohilla, 2024; Sindhura *et al.*, 2024; Kothala & Guntur, 2024; Lafraxo *et al.*, 2024).

Deep learning-based models (CNN, auto-encoder AE, and stacked auto-encoder SAE) were introduced by Helwan *et al.* (2018) with the aim of brain hemorrhagic detection. They utilized the "Brain CT Hemorrhage Dataset" consisting of 6772 images of both normal and hemorrhagic cases. However, they utilized a part of the dataset consisting of 2527 images. They found that the SAE model was the best model with the highest accuracy. They registered an accuracy of 90.9%. Although the AE model is a good DL architecture, it still has its performance limitations.

In a study by (Hssayeni *et al.*, 2020), they utilized the "Brain CT Images with Intracranial Hemorrhage Masks" dataset for brain hemorrhagic segmentation using a fully convolutional network (FCN) "U-Net". They achieved an accuracy of 87%.

In another study, Altuve and Pérez (2022) utilized the well-known ResNet18 model in a transfer learning way for the aim of brain hemorrhage detection. They used a small dataset of only 100 normal and 100 hemorrhagic CT images. They got an accuracy of 96% and a precision of 97%. However, their methodology is already known and the utilized dataset is too small.

On the other hand, Kothala and Guntur (2024) proposed the stacked bidirectional GRU-LSTM model along with the traditional CNN model to detect possible hemorrhage in CT brain images. They utilized the "Brain CT Hemorrhage Dataset," consisting of 6772 images of both normal and hemorrhagic cases, and achieved training and test accuracies of 96.2% and 93.4%, respectively. Their approach also achieved precision, recall, and

^{*} Corresponding author

This is an open access under a CC BY-NC-SA 4.0 license (<https://creativecommons.org/licenses/by-nc-sa/4.0/>)

F1-score values of 62%, 68%, and 65%, respectively. However, the low values of precision and recall indicate a high percentage of false positive and false negative errors. They also tried many transfers learning methods, including LeNet, ResNet50, and AlexNet, and they registered a low performance compared to their proposed CNN-BiLSTM model.

The EfficientNetB0 model was utilized as a feature extraction and classification model for brain hemorrhage detection (Feng *et al.*, 2023). They utilized a CT brain dataset consisting of 561 images of the normal and spontaneous intracerebral hemorrhage issue. Their proposed methodology achieved an accuracy in a range between 70% and 86.6% with an Area Under the Curve (AUC) value of 0.71 to 0.83.

Like traditional forms of feature selection, machine learning models used for the classification of medical data can greatly benefit from feature selection methods for improved predictive outcomes in medical diagnoses (Siddiq Hassan, 2013).

The American College of Neuroradiology (ASNR) hemorrhagic dataset was used in a study by He *et al.* (2024). They applied the multiscale feature classification supported by the attention fusion method and the weakly supervised localization model. They evaluated their model using only the AUC value. However, their methodology achieved AUC values of 0.89 to 0.995.

In a different study, (Malik *et al.*, 2024) compared the performance of many deep learning models in the field of brain hemorrhage detection. They utilized 2500 images of the “Brain CT Images with Intracranial Hemorrhage Masks” dataset. Their methodology achieved accuracy values of 93.29%, 90%, 82.35%, and 39.45% using the EfficientNet, ResNet50, SEResNeXt, and ResNeXt models, respectively.

There are many problems in the current state-of-art methodologies, starting from the usage of small datasets, moving to the issue of using traditional DL architectures without any modifications, struggling with high computational time or low accuracy, and ending with the problem of a bad evaluation process. However, in this study, the key contributions are listed as follows:

1- This is the first study that introduces a hybrid model of vision transformer models (ViT) and the bidirectional long-short-term memory (LSTM) architectures to enhance the performance of the current brain hemorrhage detection systems. The proposed

model is denoted as “ViTBiLSTM”. The ability of the ViTBiLSTM model to capture a better feature representation of the images based on its self-attention mechanism will improve the accuracy and solve the low performance of the traditional CNNs.

2- Since the ViT is considered a lightweight model, the computational time required for the training and validation steps will be low compared to other more complex architectures (solve the high computational time problem).

3- This study takes into consideration the problem of generalization of the proposed model by evaluating it using two different CT brain datasets.

4- The study introduces a comprehensive analysis of the performance of the proposed ViTBiLSTM model using different optimizers. It compares the proposed methodology with the original ViT model to show its efficiency.

5- The study utilizes all possible performance evaluation metrics in order to make a comprehensive assessment of the proposed model.

The next paragraphs will be organized as follows. First, the materials and proposed methodologies with the detailed architecture of the ViTBiLSTM model will be introduced. Next, the main results and findings of the adopted model will be listed (for both utilized datasets). Then, a comprehensive discussion and ablation study will be presented. Finally, the conclusion, limitations, and future work will be given.

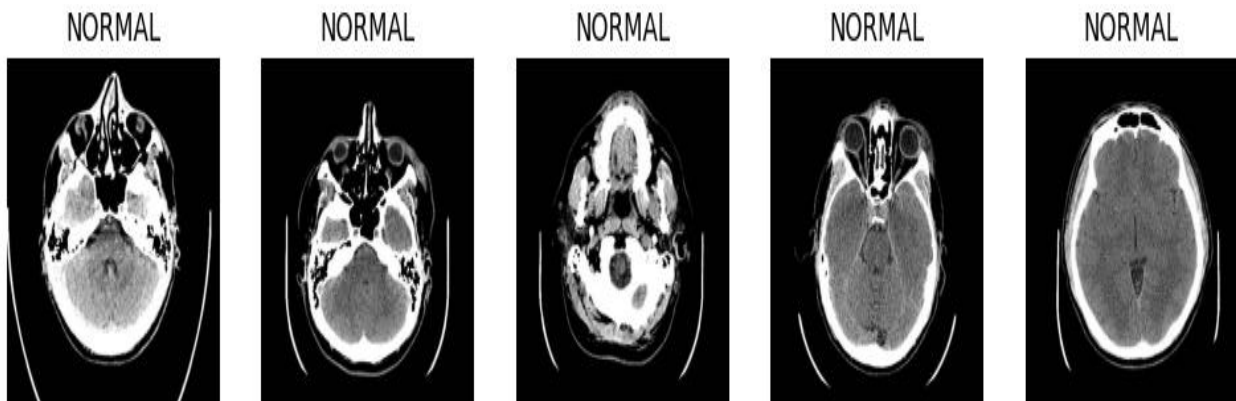
2. MATERIALS AND METHODS

CT Brain Datasets:

In this study, two datasets of both normal and hemorrhagic brain images are utilized. The first dataset, “Brain CT Hemorrhage Dataset,” consists of 6772 CT scans (4105 normal and 2667 hemorrhagic), which were originally collected from the Near East Hospital (Helwan *et al.*, 2018). The second dataset is the “Brain CT Images with Intracranial Hemorrhage Masks” (Hssayeni *et al.*, 2020), which consists of CT scans of both bones and the brain. However, in this study, the brain CT images (82 subjects: 2500 images; 2182 normal, and 318 hemorrhagic) will be utilized. Figure 1 shows some samples of these utilized datasets (the third row corresponds to the second dataset). Table 1 summarizes the characteristics of both utilized datasets.

Table 1: Datasets characteristics

Name	Num. of subjects	Num. of images	Class distribution	Source
Brain CT Hemorrhage Dataset	45 subjects	6772	4105 normal and 2667 hemorrhagic	(Helwan <i>et al.</i> , 2018)
Brain CT Images with Intracranial Hemorrhage Masks	82 subjects	2500	2182 normal, and 318 hemorrhagic	(Hssayeni <i>et al.</i> , 2020)



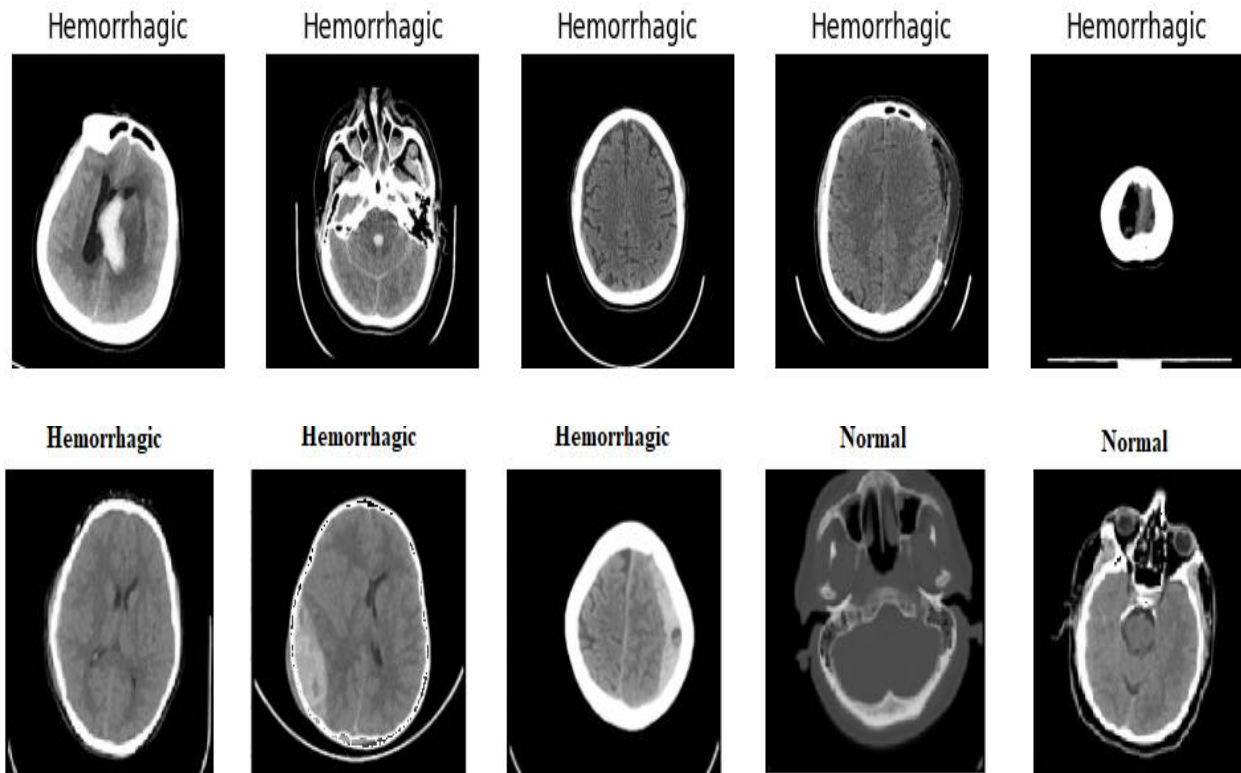


Figure 1: Examples of normal and hemorrhagic cases of the utilized datasets

The Adopted Methodology:

The proposed methodology is illustrated in Figure 2. In the first step of this study, the dataset is pre-processed using the following operations. First, the images and labels are read; then, the images are resized to a fixed size (224*224) to minimize computational time and memory-GPU-storage requirements, which are required during the training phase. Besides that, most pre-trained architectures require such input size. Then, the images are split into a training set (80%), a validation set (10%), and a test set (10%) to allow the model to learn as much as possible and ensure sufficient validation and test samples for

robustness and effectiveness. Then, all sets' images are rescaled to the range [0-1] as a normalization operation. After that, the data augmentation operations are applied to the training set. The data augmentation includes the following processes: random rotation (by the angle of 10°), zooming (percentage of 0.1), and horizontal flipping. The data augmentation operations help to avoid overfitting or overfitting and make the model more robust against changes in images; besides this, they increase the training size. The data augmentation that changes the structure of the image like zooming, is applied with a small percentage to avoid eliminating important parts that may include hemorrhagic parts.

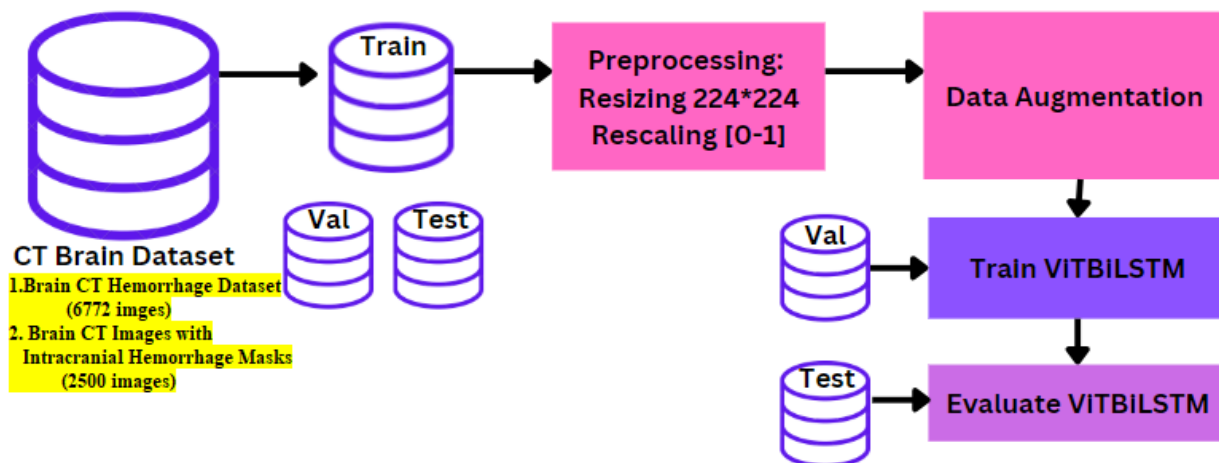
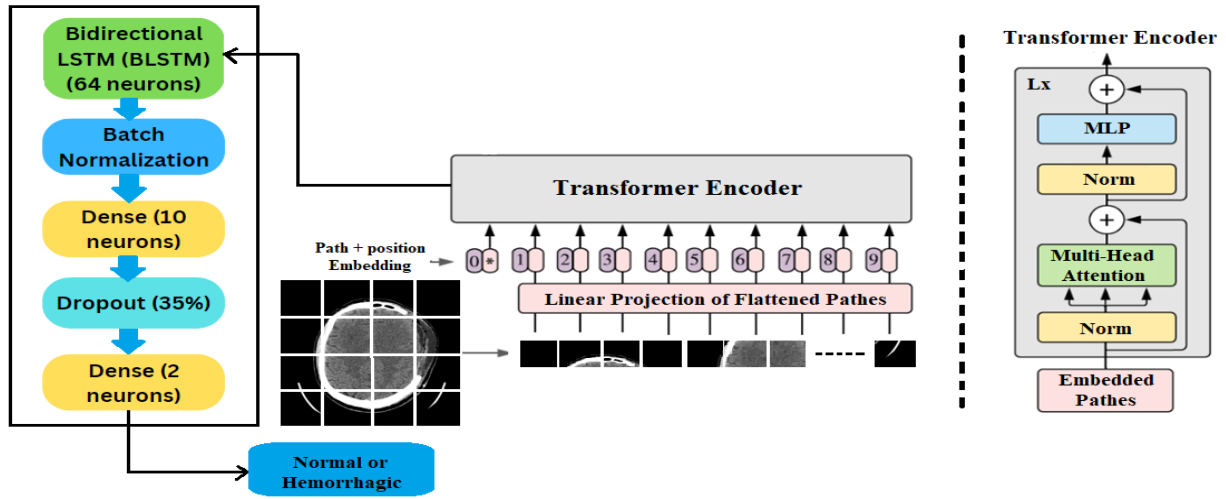


Figure 2: The general brain hemorrhagic detection system steps

After that, the training and validation sets are utilized to train and validate the utilized DL model. This study introduces a novel hybrid DL model consisting of two main architectures,

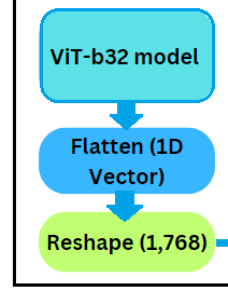
which are the vision transformer model (ViT) and the bi-directional long-short term memory (Bi-LSTM), so the developed new model is called "ViTBiLSTM" (Figure 3).



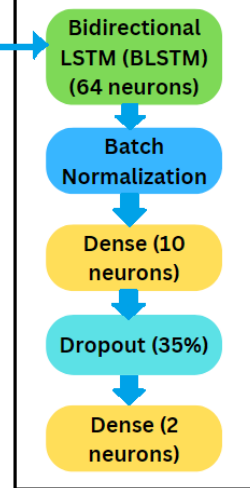
Proposed ViTBiLSTM model layers specification

Layer (type)	Output Shape	Param #
vit-b32 (Functional)	(None, 768)	87455232
flatten (Flatten)	(None, 768)	0
reshape_1 (Reshape)	(None, 1, 768)	0
bidirectional (Bidirectional)	(None, 128)	426496
batch_normalization (Batch Normalization)	(None, 128)	512
dense (Dense)	(None, 10)	1290
dropout (Dropout)	(None, 10)	0
dense_1 (Dense)	(None, 2)	22
Total params: 87883552 (335.25 MB)		
Trainable params: 87883296 (335.25 MB)		
Non-trainable params: 256 (1.00 KB)		

Feature Extraction Part

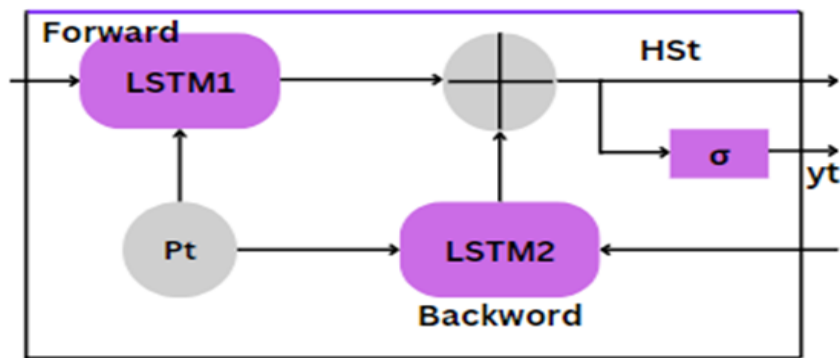


Feature Representation and classification part



(a)

(b)



(c)

Figure 3: The proposed ViTBiLSTM model: (a) The general architecture, (b) ViTBiLSMT architecture, (c) BiLSTM

While ViT models are powerful in capturing global spatial dependencies based on their self-attention mechanisms, they may lack the ability to model sequential dependencies across feature representations. For this reason, utilization of the BiLSTM model can guarantee capturing the sequential patterns in the input data (The input image of the ViT model is split into adjacent blocks that are spatially relevant). From another point of view, using ViT against traditional CNN improves the ability of the model to

understand the spatial structure of the input image, leading to a better feature representation compared to CNN.

The vision transformer model (ViT) was mainly introduced in a study by Dosovitskiy *et al.* (2020). The input image of the ViT model (in our case CT brain image) is divided into $S \times S$ patches (not overlapped). These $P \times P$ boxes (patches) are then linearly embedded using an embedding layer that transforms them into a sequence of tokens. Position-embedding information

is also added to each image patch to provide it with spatial information.

Let $X \in \mathbb{R}^{H \times W \times C}$ represent the input image, where H, W, and C are the height, width, and number of channels, respectively. Patch emedding of the ViT model divided input image into $P \times P$ patches with a total number of patches ($N=H*W/P^2$).

Each patch is flattened and linearly projected into a D-dimensional embedding space:

$$Z_0^i = E.Faltten(patch_i) + p_i \text{ where } i = 1, 2, \dots, N \quad (1)$$

Where $E \in \mathbb{R}^{(P^2.C) \times D}$ is the position matrix, while P_i is the positional encoding of the patch i , and z_0^i is the embedded representation of the patch i .

These positional-encoded embedding sequences are then fed into the main ViT backbone, which is a Transformer-based architecture. The main module in this architecture is the multi-head attention layers (Du *et al.*, 2024) that are responsible for extracting features based on the attention mechanism (i.e., the model updates its weights to concentrate on the most essential parts of the image). Each head is defined by Equation II.

$$head_i = softmax \left(\frac{Q_i K_i^T}{\sqrt{d_k}} \right) \cdot V_i \quad (2)$$

Where Q, k, and V are the Query, Key, and Value, which are the three different representations of the multi-head attention layer. The outputs of the transformer encoder model will be the final feature vector which will be delivered to the classification part. However, the original classification part of the ViT model is a multi-layer perceptron and a classification with a softmax layer. BiLSTM (Graves *et al.*, 2005) or Bidirectional LSTMs (see Figure 3-c) are a type of recurrent neural network (RNN) that maintains both past and future contexts in sequence data. In BiLSTMs, two LSTMs are utilized: one processes the sequence from start to end (called forward LSTM), while the other processes it from end to start (called backward LSTM). The final

Where, N is the number of patches, Z is the ViT output, W1 is the weight matrix of the first dense layer, W2 is the weight matrix of the second dense layer, ϕ is the ReLU activation function, b1 and b2 are the bias vectors.

Performance Evaluation Metrics:

The final step in this study is the evaluation process in which the trained ViTBiLSTM model will be evaluated to judge its performance and select the best parameters that achieve the best performance. For this reason, the main utilized computations, which are True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN), are computed. Using these raw computations, the precision, recall, F1-score, and accuracy metrics are also computed for the individual classes, and then both macro and weighted average calculations are derived. The weighted average is computed by assigning a weight for each class based on the number of samples, making the average score correspond to the class's percentage. On the other hand, macro average applies no weighting on the individual score and it's more suitable for performance calculations in case of unbalanced datasets to check any possible individual class's errors. The confusion matrix (CM) that shows the TP, TN, FP, and FN of each class is also drawn. CM shows how well the trained model is performing since it compares the original classes with the predicted ones. It gives detailed calculations of the individual classes' true and false predictions allowing to precisely define the classes with the best and worst performance. Figure 4 states the CM according to the probled addressed in this study. The receiver operating characteristics (ROC) curve, which represents the relationship between true positive rate (TPR) and false positive rate (FPR), is also drawn. The area under the curve metric is also computed using the ROC curve. The training time of all proposed scenarios is also computed. Equations 6, 7, 8, and 9 show the formula of precision, recall, score, and accuracy,

In this study, we removed this part and replaced it with a more powerful classification model consisting of Bi-LSTM architecture which fits the idea of a transformer model since the input image is decomposed into patches and the Bi-LSTM module can be very suitable to retain the local information of the gray levels of the adjacent patches in the original image, leading to a better classification task. To make the output of the transformer model suitable for the input of the Bi-LSTM model, the study suggests using a reshape layer, which is responsible for changing the output feature vector of the transformer model into a (1,768) shape, which is suitable to the input shape of the Bi-LSTM model. Moreover, a batch normalization layer is added to the output of the Bi-LSTM model to improve training stability and make a faster convergence (Rivoir *et al.*, 2024; Tin *et al.*, 2024; Fei *et al.*, 2024). Finally, two dense layers are added as a hidden fully connected layer with a softmax activation function in the final layer. A dropout layer is also inserted before the last layer with a drop percentage of 35% to regularize the model and prevent overfitting. Moreover, in the present study, the B32 version of the ViT model is utilized (Liu & Aldrich, 2024), which means that the input image is divided into 49 patches (so if the input image size is 224*224, then the image will be divided into 7*7 patches). The output of the transformer model and the reshaped version is given in equations III and IV.

$$Z_{reshaped} = Reshape(Z_{vit}) \quad (3)$$

Where, Z_{vit} is the output of the encoder part of the ViT model and is given as follows:

$$Z_{vit} = TransformerEncoder([z_0^1, z_0^2, \dots, z_0^{49}]) \quad (4)$$

output of the proposed ViTBiLSTM model is given in Equation (V).

$$y = Softmax \left(W_2 \cdot Dropout \left(\phi \left(W_1 \cdot BatchNorm \left(\frac{1}{N} \sum_{t=1}^N BiLSTM(Z_t) \right) + b1 \right) \right) + b2 \right) \quad (5)$$

respectively (Khozama & Mayya, 2022; Szabó *et al.*, 2024; Hoang *et al.*, 2024).

$$Precision = \frac{TP}{TP+FP} \quad (6)$$

$$Recall = \frac{TP}{TP+FN} \quad (7)$$

$$F1 - score = \frac{2 * Precision * Recall}{Precision + Recall} \quad (8)$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (9)$$

		Predicted Label	
		Hemorrhagic	NORMAL
True Label	Hemorrhagic	TP	FN
	NORMAL	FP	TN

Figure 4: CM description according to the brain hemorrhage prediction problem

Ethical Approval and Consent:

The study's design and procedures were reviewed and approved by the Ethics and Scientific Committee of the College of Medicine at the University of Zakho with the reference number (FEB2024/UOZE440).

3. RESULTS

The experiments in this study are applied to the utilized dataset using the training parameters illustrated in Table 2. In all training scenarios, the number of epochs is 50, the utilized batch size is 64, the loss function is the categorical cross entropy (two

classes in the classification layer), learning rate is 0.0001, image size is 224*224, and the early stop condition is enabled with a patience value of 5 (if the validation loss isn't improved for 5 epochs, the training will be stopped to prevent possible overfitting or redundant training time).

Table 2: Training parameters

Parameter	Description
Image Size	224*224*3
Batch size	64
Optimizer	Adam/ RMSProp/ NADAM/ ADAMX
Learning Rate	1e-4
Loss function	Categorical cross-entropy
Metrics	Accuracy
Epochs	50
Early stop condition	Patience=5
Save best only	Yes
Resources	Training resources: The model was trained using the COLAB environment [NVIDIA Tesla T4 GPU (16 GB VRAM) and 12 GB RAM] Required implementation resources: CPU with 4 to 8 GB of RAM (no need for GPU) since the test or evaluation will be on a single image.

First Dataset's Results:

The training and validation accuracy and loss curves of the trained ViTBiLSTM are illustrated in Figure 5. The model training is stopped at the 13th epoch due to the early stop

condition. The curves have no overfitting and the convergence between the training and validation curves is noticed. The confusion matrix (Figure 6) shows that the trained model achieves one false positive and one false negative error and an AUC value of 1

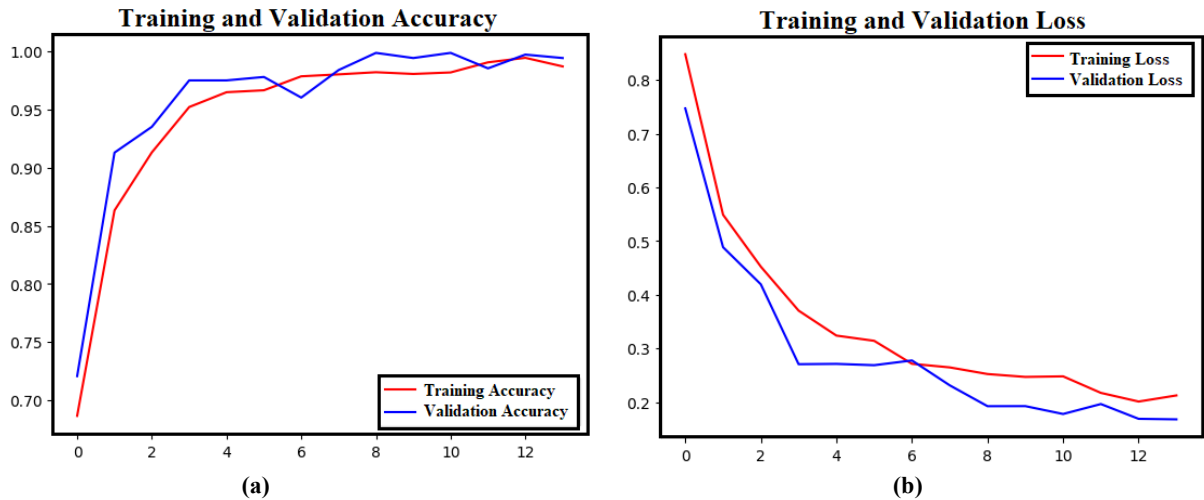


Figure 5: Training and validation metrics of the ViT model: (a) Accuracy, (B) Loss.

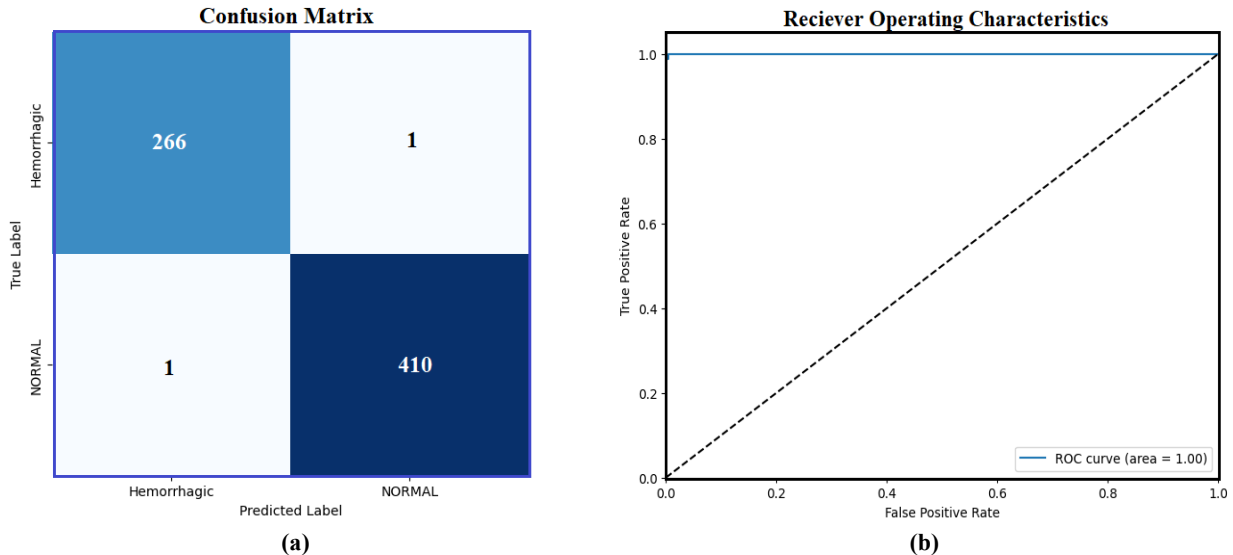


Figure 6: Confusion matrix and ROC plot of the proposed ViTBiLSTM model: (a) Confusion matrix, (B) ROC.

Table 3 includes the precision, recall, F1-score, and support (number of samples) for the normal and hemorrhagic classes. Besides that, the macro and weighted average metrics are also shown in Table 3. The average precision, recall, and F1-score of the proposed ViTBiLSTM model are 99.86% for all metrics, indicating a high performance. Results indicate that the model

not only achieves high accuracy but also maintains balanced and consistent predictions across both normal and hemorrhagic classes.

Figure 7 shows some test examples of the CT brain (of the test set) and their corresponding predictions.

Table 3: Assessment of the trained ViT model in terms of precision, recall, and F1-score using the test set.

	Precision (%)	Recall (%)	F1-score (%)	Support
Hemorrhagic	99.62	99.62	99.62	267
NORMAL	99.75	99.75	99.75	411
Macro avg	99.685	99.685	99.685	678
Weighted avg	99.698	99.698	99.698	678

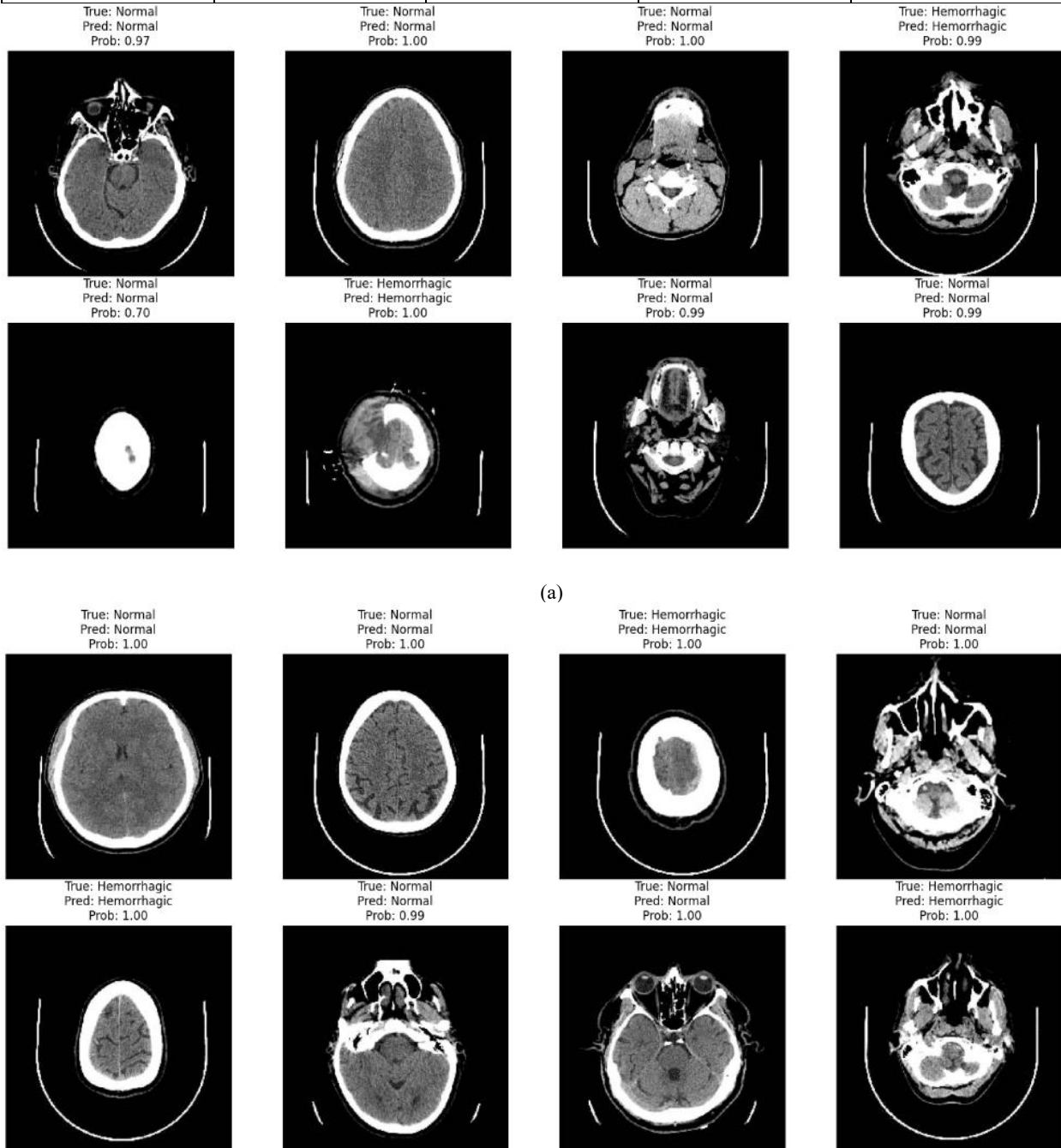


Figure 7: Results of predicting some samples using the trained ViTBiLSTM model: (a) Validation images, (b) Test images

Second Dataset's Results:

In this section, a new dataset with different specifications will be utilized to generalize the results. In this dataset, the number of normal cases is 2183, while the number of hemorrhage samples is only 318, revealing a problem of data imbalance. To address this issue, the dataset is split into training, validation, and test sets (with the same criteria as in the first dataset), and then the small class is oversampled to match the number of the majority class. The same data augmentation operations and the same training parameters are also utilized to preprocess the dataset and prepare the training operation. Figure 8 shows the training and validation accuracy and loss curves of the trained

ViTBiLSTM model using the second dataset (before and after the data balance).

The proposed balancing method includes three main steps: duplicating samples from the minority class (oversampling), balancing the class distribution, and shuffling the dataset to prevent order bias. These operations are applied to the training set only. Curves are more stable and have better convergence in the case of using data balance. These findings are also proved by the confusion matrix and AUC values, where the data-balance-based ViTBiLSTM model registered only 4 false negatives and 0.99 as the AUC value, while in the case of training ViTBiLSTM without data balance, the AUC value is only 0.84.

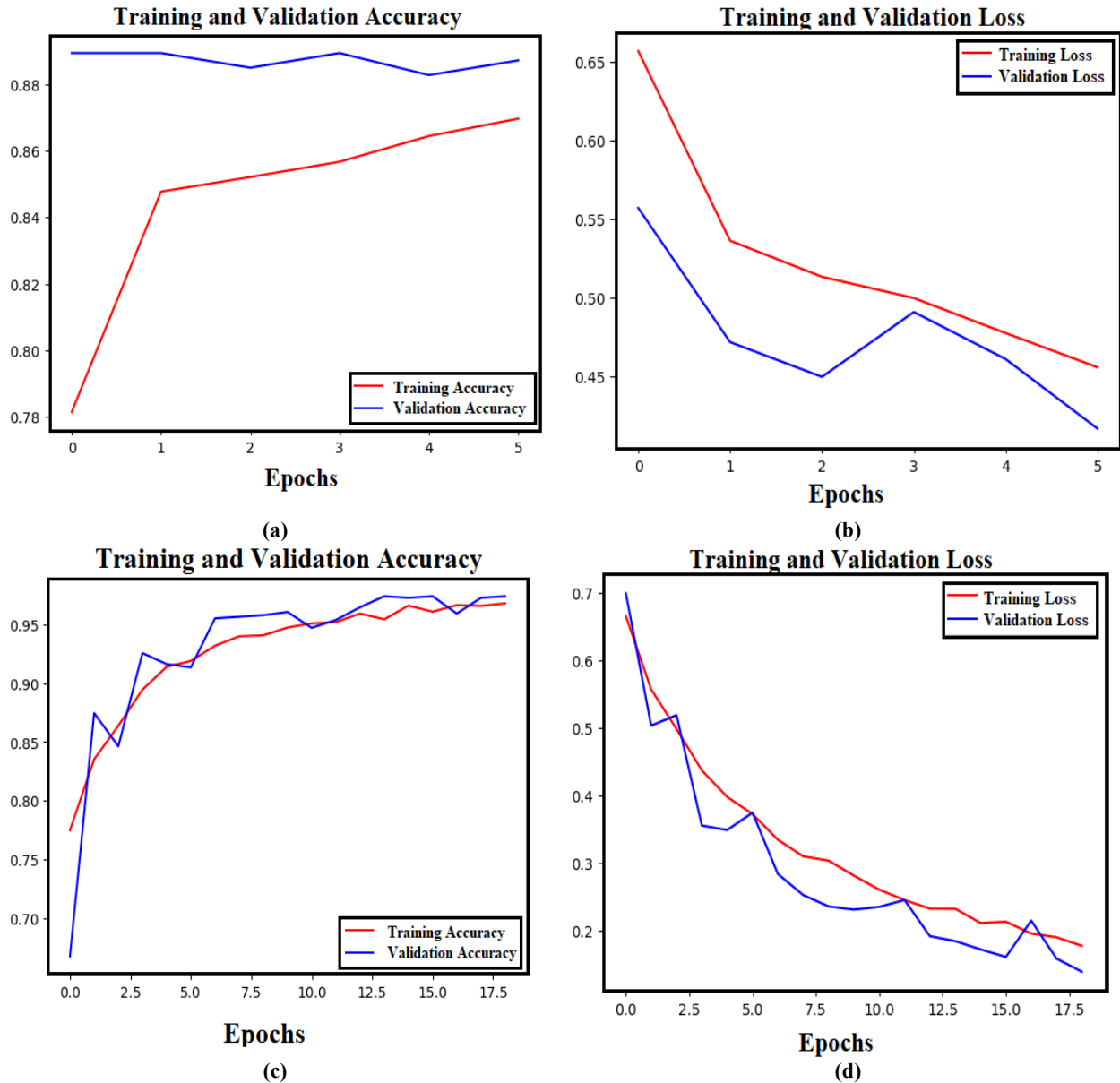


Figure 8: Performance curves of the trained ViTBiLSTM model: (a, b) before balance, (c, d) after balance

Table 4 states a performance comparison between two scenarios of the second dataset; one case is the training results of the ViTBLSM model without balance, while the other one is the same evaluation metrics after the data balance. Table 4 proves also shows that the data balance has improved the performance

by 10.29% in terms of test accuracy. Generally, the proposed ViTBiLSTM achieved a high accuracy using a new dataset with different specifications. These results indicate the possibility of generalizing the results and the robustness of the proposed ViTBiLSTM model.

Table 4: Training, validation, and test metrics of the second dataset.

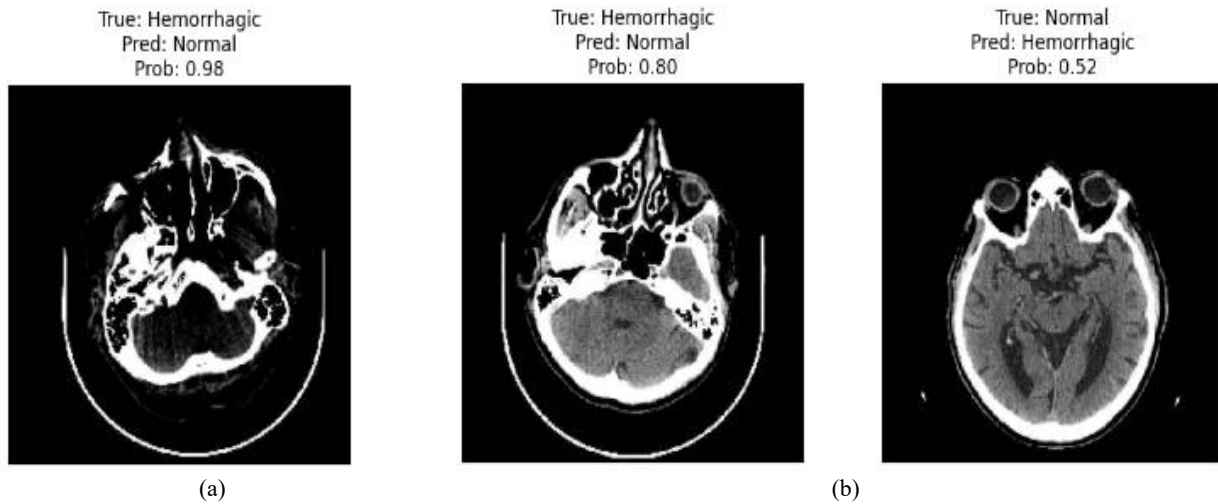
	Training Accuracy %	Validation Accuracy %	Test Accuracy %	Test Precision	Test Recall	Test F1-score	Training Time (s/epoch)	AUC
Without Balance	78.15	88.94	86.25	81	61	64	35.16	0.84
With Balance	95.47	97.44	96.94	97	97	97	58.73	0.99

4. DISCUSSION

Overall Discussion:

The proposed ViTBiLSTM model achieves high performance in terms of precision, recall, F1 score, and accuracy. The test samples also showed that the proposed model can detect the right class with a high confidence level. However, some test

and validation samples failed to be predicted, as shown in Figure 9-a (misclassified test sample) and Figure 9-b. There are three misclassified samples as shown in Figure 9, where two of them are of class 'Hemorrhagic' but are misclassified as 'Normal'. This may be due to the additional areas in the brain CT image, which is similar to the normal tissues. In contrast, the 'normal' case, which is misclassified as 'Hemorrhagic', contains gray levels that have a similar distribution to the hemorrhagic levels.

**Figure 9:** Misclassified samples (a) Validation, (b) Test

Ablation Study:

For a deeper discussion, the "Adam" optimizer of the ViTBiLSTM model will be changed, and new results will be derived and compared with the original ones. The comparison between the performance of these different scenarios is shown in Table 5 and Figure 10. Table 5 summarizes the macro average metrics of the trained ViTBiLSTM model under different optimizers and proves that the best case is the 'RMSProp'

optimizer, and the second best one is the 'ADAM' optimizer. This finding is normal since the ADAM and RMSProp optimizers both utilize the concept of adaptive learning rate, which is changed during the training, leading to a better convergence. Using RMSProp optimizer instead of ADAM optimizer enhances the performance by almost 0.3% for all metrics, while the utilization of other optimizers like ADAMX or NADAM leads to a lower performance.

Table 5: Comparison experiment of different optimizers with the same ViTBiLSTM model.

Model	Precision (%)	Recall (%)	F1-score (%)	Accuracy (%)
ViTBiLSTM (ADAM)	99.685	99.685	99.685	99.7
ViTBiLSTM (RMSProp)	100	100	100	100
ViTBiLSTM (NADAM)	99	99	99	98.96
ViTBiLSTM (ADAMX)	91	89	90	90.56

Confusion matrixes and ROC plots of the three models are shown in Figure 10. However, NADAM combines two optimizers, ADAM and Nesterov moments, making it more complex and may not align well with the utilized dataset. For ADAMX, the gradient (loss), as shown in Figure 10-C, is not

stable, and this is due to the complexity of the learning process, which does not fit the current problem. The confusion matrix of the RMSProp shows zeros false positives and false negative errors, while the NADAM and ADAMX optimizers show too many false positive and false negative errors.

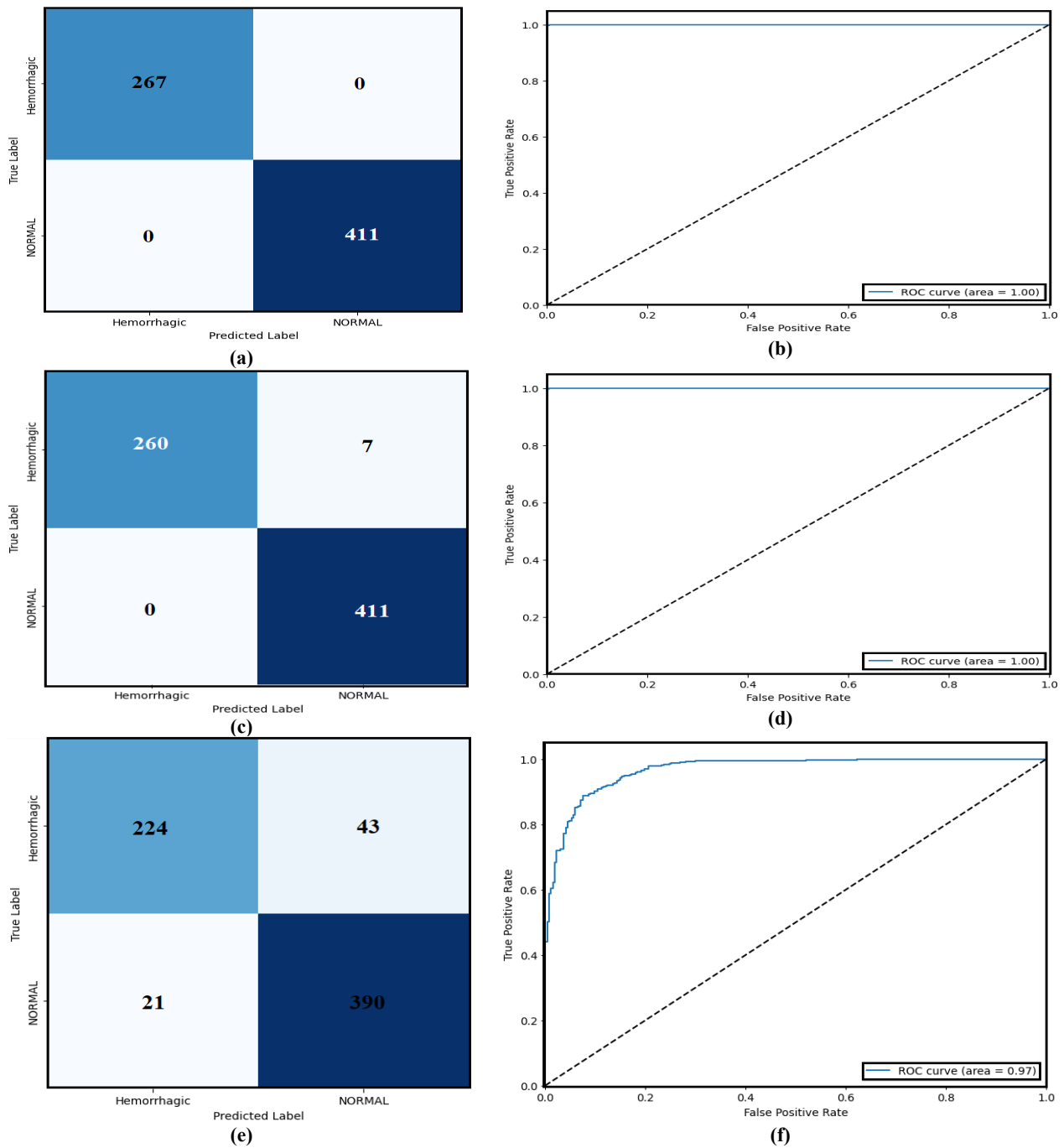


Figure 10: Confusion matrix and ROC plot of different optimizers (same ViTBiLSTM model): (a, b) CM and ROC of RMSProp, (c, d) CM and ROC of NADAM, (e, f) CM and ROC of ADAMX

The training, validation, and test evaluation metrics of different optimizer experiments are also shown in Table 6. The findings shown in Table 6 indicate that the Adam and RMSProp optimizers give the best performance with the ViTBiLSMT

model. ViTBiLSTM with RMSProp achieves a validation accuracy of 99.85% (similar to ADAM optimizer), a test accuracy of 100%, and an AUC score of 1. For training time, the RMSProp consumed the least computational time.

Table 6: Training, validation, and test metrics of different optimizers.

	Training Accuracy %	Training Loss	Validation Accuracy %	Validation Loss	Test Accuracy %	Test Loss	Training Time (s/epoch)	AUC (Test)
ViTBiLSTM (ADAM)	98.19	0.2468	99.85	0.1923	99.7	0.19	99.85	1.0
ViTBiLSTM (RMSProp)	98.95	0.2208	99.85	0.1799	100	0.169	98.22	1.0
ViTBiLSTM (NADAM)	96.2	0.3413	99.7	0.3111	98.96	0.313	108.33	1.0
ViTBiLSTM (ADAMX)	87.39	0.3451	93.06	0.2232	90.56	0.257	103.96	0.97

External Validation:

In order to evaluate the trained model on a different dataset, we utilized the validation and test sections of the second dataset as an external validation set (without training the train set) of the

trained ViTBiLSTM model of the first dataset. Table 7 illustrates the results of external validation. The proposed ViTBiLSTM model revealed a test accuracy of 93%, which proves the ability of the model to present correct predictions even under different conditions.

Table 7: External validation of the ViTBiLSTM model using the validation and test set of the second dataset.

Precision %	Recall %	F1-score %	Accuracy %
92	93	92	93

Comparison with CNN and Other Vit-Based Architectures

To show the main benefit of the proposed ViTBiLSTM model, a comparison of the performance between this model and other architectures (CNN, ViT (original), and ViT with LSTM) is shown in Table 8. The comparison shows that the ViT model registered 5 false negatives and one false positive, compared with

an accuracy of 99%, while the proposed ViTBiLSTM model achieved an accuracy of 99.7%. Table 8 states that the ViTBiLSTM model outperforms the original ViT model (with an MLP classification layer) by almost 0.7% for all metrics. Similarly, ViT with LSTM achieved the closest scores to the proposed methodology, while the CNN model has a lower performance by almost 5.6%.

Table 8: Comparison of the original ViT and the proposed ViTBiLSTM model.

	Precision (%)	Recall (%)	F1-score (%)	Accuracy (%)
ViTBiLSTM	99.685	99.685	99.685	99.7
ViTLSTM	99.35	99.35	99.35	99
ViT	99	99	99	99
CNN	94	94	94	94

Error Analysis Study:

The proposed ViTBiLSTM model misclassified only a few samples (an estimated ~2–3 in total) out of 678 test images using the first dataset. These errors were nearly balanced across the two classes: Hemorrhagic and NORMAL, as both achieved almost identical precision and recall (99.62% vs. 99.75%). For the second dataset, and without balancing, the model's recall dropped significantly to 61%, while precision remained at 81%, suggesting a tendency to under-predict the positive class (likely hemorrhagic). After applying data balancing, performance improved to 97% across all metrics. These findings indicate that class imbalance was a major cause of misclassification, and addressing it during training was crucial for reliable predictions. Moreover, the ADAMX optimizer showed the lowest performance, with a test Accuracy of 90.56% and an F1-score of 90%. Misclassifications were more frequent here, possibly due to slower convergence or instability during training (as seen from its higher loss) and incomplete optimization of deeper layers in

the ViTBiLSTM model. However, the ADAM and RMSProp optimizers achieved the best performance and lowest error rates.

Comparison with Related Work:

The comparison with the state-of-the-art studies is essential to focus on the contribution of this study (Table 9). The studies that utilized the same datasets achieved a lower performance compared to the current study. (Helwan *et al.*, 2018) registered an accuracy of 90.0%, while Kothala and Guntur (2024) achieved an accuracy of 93.4% on the same dataset. The current study outperforms all previous studies that utilized the same dataset. Using another dataset (Brain CT Images with Intracranial Hemorrhage Masks (2500 images)), the current study also outperformed Hssayeni *et al.* (2020), which utilized the same dataset, in terms of accuracy by 9.94%. Similarly, the current study outperforms Malik *et al.* (2024) by 3.65% in terms of accuracy. The proposed hybrid model of ViT and BiLSTM unified the high accuracy and low computational time of both architectures to build a robust model that achieves state-of-the-art performance.

Table 9: Comparison with related work

Study	Methodology	Dataset	Dataset size	Results & Notes
(Helwan <i>et al.</i> , 2018)	SAE	Brain CT Hemorrhage Dataset	6772 images	Accuracy=90.9%
(Hssayeni <i>et al.</i> , 2020)	FCN-U-Net	Brain CT Images with Intracranial Hemorrhage Masks	2500 images	Accuracy=87%
(Altuve & Pérez, 2022)	ResNet18 (Transfer Learning)	A small dataset	200 images	Accuracy=96%
(Kothala & Guntur, 2024)	Stacked bidirectional GRU-LSTM and CNN	Brain CT Hemorrhage Dataset	6772 images	Accuracy=93.4%
(Feng <i>et al.</i> , 2023)	EfficientNetB0	CT brain dataset	561 images	Accuracy=70%-86.6%
(He <i>et al.</i> , 2024)	Multiscale feature classification supported by the attention	American College of Neuroradiology (ASNR)	-	AUC = 0.89-0.995
(Malik <i>et al.</i> , 2024)	EfficientNet	Brain CT Images with Intracranial Hemorrhage Masks (2500 images)	2500 images	Best Accuracy=93.29%
Current Study	Novel ViTBiLSTM model	Brain CT Hemorrhage Dataset	6772 images	Accuracy: 99.7% (ADAM), 100% (RMSProp)

		Brain CT Images with Intracranial Hemorrhage Masks	2500 images	Accuracy: 96.94%
--	--	--	-------------	------------------

CONCLUSION

In this study, a novel deep learning framework called the ViTBiLSTM model consists of two main parts; the feature extraction part which is the ViT model responsible for extracting features of the images, while the second part is the classification part in which the BiLSTM model is utilized to make the classification instead of the original transformer classification part. The BiLSTM is chosen since it fits the idea of decomposing input images into patches and maintaining the information of the adjacent pixels of the patches to improve the ability to extract better information. The study utilizes two different CT image datasets; one contains the problem of data balance, and the other one contains a higher number of samples. Both datasets are pre-processed and data augmentation operations are applied for a better training process. Many experiments are applied: one with the original ViT model and other experiments with the ViTBiLSTM model with different optimizers (ADAM, NADAM, ADAMX, and RMSProp). Results showed that the best case is the usage of the ViTBiLSTM model with RMSProp optimizer with an accuracy of 100% of the first dataset. The second dataset got the best performance using the data balance operation and the ViTBiLSTM model with an accuracy of 96.94%. A comparison of the current study with the previous studies in the same field was also performed and proved the robustness and high performance of the proposed ViTBiLSTM against the traditional ViT models, CNN, transfer learning-based, and CNN-LSTM models. Future studies can focus on the utilization of other different datasets and the fusion of some feature extraction and classification DL-based architectures for better enhancement. Moreover, the current research focused on the binary classification of brain hemorrhage; however, future studies can focus on the multi-class classification problem.

Acknowledgements:

I would like to express my gratitude to the University of Duhok and the University of Zakho for their time and consideration in supporting my academic endeavours.

Statements and Declarations:

Ethical Approval :

All authors gave verbal informed consent for their participation. The study's design and procedures were reviewed and approved by the Research Ethics Committee of the College of Medicine, UOZ, in compliance with ethical standards (Code UOZE448; 2024).

Conflict of Interest: The author declared that no potential conflict of interest.

Author Contributions: The author has reviewed the final version to be published and agreed to be accountable for all aspects of the work.

Consent to Participate: The author has consented to submit this article to this journal.

Consent to Publish: The author has consented to publish this article in this journal.

Funding: The study has not receive any specific funding from public, commercial, or any non-profit organizations.

Concept and Design: Delveen Luqman Abd Alnabi.

Acquisition, Analysis, or Interpretation of Data: Delveen Luqman Abd Alnabi

Drafting of the Manuscript: Delveen Luqman Abd Alnabi.

REFERENCES

Ahmed, S., Esha, J. F., Rahman, M. S., Kaiser, M. S., Hosen, A. S. M. S., Ghimire, D., & Park, M. J. (2024). Exploring

deep learning and machine learning approaches for brain hemorrhage detection. *IEEE Access*. <https://doi.org/https://doi.org/10.1109/ACCESS.2024.3376438>

Akmaljon o'g, M. A., Abdullajon o'g'li, M. S., & Tolmasovich, T. R. (2024). Acute disturbance of blood circulation in the head. *Western European Journal of Medicine and Medical Science*, 2(4), 27–31.

Altuve, M., & Pérez, A. (2022). Intracerebral hemorrhage detection on computed tomography images using a residual neural network. *Physica Medica*, 99, 113–119. <https://doi.org/10.1016/j.ejmp.2022.05.015>

Datta, P., & Rohilla, R. (2024). An autonomous and intelligent hybrid CNN-RNN-LSTM-based approach for the detection and classification of abnormalities in the brain. *Multimedia Tools and Applications*, 1–27. <https://doi.org/https://doi.org/10.1007/s11042-023-17877-3>

Del Gaizo, A. J., Osborne, T. F., Shahoumian, T., & Sherrier, R. (2024). Deep learning to detect intracranial hemorrhage in a national teleradiology program and the impact on interpretation time. *Radiology: Artificial Intelligence*, 6(5), e240067. <https://doi.org/https://doi.org/10.1148/ryai.240067>

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houslsby, N. (2020). An image is worth 16x16 Words: Transformers for image recognition at scale. *ArXiv, abs/2010.11929*. <https://doi.org/10.1016/j.procs.2024.04.157>

Du, Y., Lang, W., Hu, X., Yu, L., Zhang, H., Zhang, L., & Wu, Y. (2024). Quality assessment of light field images based on adaptive attention in ViT. *Electronics*, 13(15), 2985. <https://doi.org/10.3390/electronics13152985>

Fei, W., Dai, W., Li, C., Zou, J., & Xiong, H. (2024). On centralization and unitization of batch normalization for deep ReLU neural networks. *IEEE Transactions on Signal Processing*, 72, 2827–2841. <https://doi.org/10.1109/TSP.2024.3410291>

Feigin, V. L., Brainin, M., Norrving, B., Martins, S., Sacco, R. L., Hacke, W., Fisher, M., Pandian, J., & Lindsay, P. (2022). World stroke organization (WSO): Global stroke fact sheet 2022. *International Journal of Stroke*, 17(1), 18–29. <https://doi.org/https://doi.org/10.1177/17474930211065917>

Feng, C., Ding, Z., Lao, Q., Zhen, T., Ruan, M., Han, J., He, L., & Shen, Q. (2023). Prediction of early hematoma expansion of spontaneous intracerebral hemorrhage based on deep learning radiomics features of noncontrast computed tomography. *European Radiology*, 34(5), 2908–2920. <https://doi.org/10.1007/s00330-023-10410-y>

Graves, A., Fernández, S., & Schmidhuber, J. (2005). Bidirectional LSTM networks for improved phoneme classification and recognition. *International Conference on Artificial Neural Networks*, 799–804. https://doi.org/http://dx.doi.org/10.1007/11550907_163

Grey, M. T. (2024). *White matter lesions: Development, imaging, effect on brain function* [Doctoral dissertation, Masaryk University, Faculty of Medicine]. Theses.cz.

Haldorai, A., Murugan, S., & Balakrishnan, M. (2024). Hemorrhage Detection from Whole-Body CT Images Using Deep Learning. In *Artificial Intelligence for Sustainable Development* (pp. 139–151). Cham: Springer Nature Switzerland. <https://doi.org/https://doi.org/10.1007/9>

- He, B., Xu, Z., Zhou, D., & Zhang, L. (2024). Deep multiscale convolutional feature learning for intracranial hemorrhage classification and weakly supervised localization. *Heliyon*, 10(9), e30270. <https://doi.org/10.1016/j.heliyon.2024.e30270>
- Helwan, A., El-Fakhri, G., Sasani, H., & Uzun Ozsahin, D. (2018). Deep networks in identifying CT brain hemorrhage. *Journal of Intelligent & Fuzzy Systems*, 35(2), 2215–2228. <https://doi.org/10.3233/JIFS-172261>
- Hoang, Q. T., Pham, X. H., Trinh, X. T., Le, A. V., Bui, M. V., & Bui, T. T. (2024). An efficient CNN-based method for intracranial hemorrhage segmentation from computerized tomography imaging. *Journal of Imaging*, 10(4), 77. <https://doi.org/10.3390/jimaging10040077>
- Hsayeni, M., Croock, M., Salman, A., Al-khafaji, H., Yahya, Z., & Ghoraani, B. (2020). Computed tomography images for intracranial hemorrhage detection and segmentation. *Intracranial Hemorrhage Segmentation Using a Deep Convolutional Model*. *Data*, 5(1), 14. <https://doi.org/10.13026/w8q8-ky94>
- Hu, P., Yan, T., Xiao, B., Shu, H., Sheng, Y., Wu, Y., Shu, L., Lv, S., Ye, M., & Gong, Y. (2024). Deep learning-assisted detection and segmentation of intracranial hemorrhage in noncontrast computed tomography scans of acute stroke patients: A systematic review and meta-analysis. *International Journal of Surgery*, 110(6), 3839–3847. <https://doi.org/10.1097/JS9.0000000000001266>
- Ibrahim, W. R., & Mahmood, M. R. (2023). Classified covid-19 by densenet121-based deep transfer learning from ct-scan images. *Science Journal of University of Zakho*, 11(4), 571–580. <https://doi.org/10.25271/sjuoz.2023.11.4.1166>
- Khozama, S., & Mayya, A. M. (2022). A new range-based breast cancer prediction model using the Bayes' theorem and ensemble learning. *Information Technology and Control*, 51(4), 757–770. <https://doi.org/10.5755/j01.itc.51.4.31347>
- Kothala, L. P., & Guntur, S. R. (2024). An efficient stacked bidirectional GRU-LSTM network for intracranial hemorrhage detection. *International Journal of Imaging Systems and Technology*, 34(1), e22958. <https://doi.org/https://doi.org/10.1002/ima.22958>
- Lafraxo, S., El Ansari, M., & Koutti, L. (2024). Computer-aided system for bleeding detection in wce images based on CNN-GRU network. *Multimedia Tools and Applications*, 83(7), 21081–21106. <https://doi.org/https://doi.org/10.1007/s11042-023-16305-w>
- Liu, X., & Aldrich, C. (2024). Multivariate image processing in minerals engineering with vision transformers. *Minerals Engineering*, 208, 108599. <https://doi.org/10.1016/j.mineng.2024.108599>
- Majeed, M. A. A., Alrawi, A. T., & Okashi, O. M. Al. (2024). Survey on machine and deep learning methods used in CT scan brain diseases diagnosis. *AIP Conference Proceedings*, 3009(1). <https://doi.org/https://doi.org/10.1063/5.0190368>
- Malik, P., Dureja, A., Dureja, A., Rathore, R. S., & Malhotra, N. (2024). Enhancing intracranial hemorrhage diagnosis through deep learning models. *Procedia Computer Science*, 235, 1664–1673. <https://doi.org/10.1016/j.procs.2024.04.157>
- Murad, S. H., Awlla, A. H., & Moahmmed, B. T. (2023). Prediction lung cancer based critical factors using machine learning. *Science Journal of University of Zakho*, 11(3), 447–452. <https://doi.org/10.25271/sjuoz.2023.11.3.1105>
- Neethi, A. S., Kannath, S. K., Kumar, A. A., Mathew, J., & Rajan, J. (2024). A comprehensive review and experimental comparison of deep learning methods for automated hemorrhage detection. *Engineering Applications of Artificial Intelligence*, 133, 108192. <https://doi.org/https://doi.org/10.1016/j.engappai.2024.108192>
- Prasher, S., Nelson, L., & Arumugam, D. (2024). Sequential CNN model for hemorrhage prediction using brain CT images. *2024 International Conference on Advances in Modern Age Technologies for Health and Engineering Science (AMATHE)*, 1–4. <https://doi.org/https://doi.org/10.1109/AMATHE61652.2024.10582253>
- Rather, M. A., Khan, A., Javed, H., Jahan, S., Tabassum, R., & Begum, R. (2024). Neuropathology of neurological disorders. In *Mechanism and Genetic Susceptibility of Neurological Disorders* (pp. 1–33). Singapore: Springer Nature Singapore. https://doi.org/https://doi.org/10.1007/978-981-99-9404-5_1
- Rivoir, D., Funke, I., & Speidel, S. (2024). On the pitfalls of Batch Normalization for end-to-end video learning: A study on surgical workflow analysis. *Medical Image Analysis*, 94, 103126. <https://doi.org/10.1016/j.media.2024.103126>
- Schiariti, V., Shierk, A., Stashinko, E. E., Sukal-Moulton, T., Feldman, R. S., Aman, C., Mendoza-Puccini, M. C., Brandenburg, J. E., & Committee, N. I. of N. D. and S. C. P. C. D. E. O. (2024). Cerebral palsy pain instruments: Recommended tools for clinical research studies by the National Institute of Neurological Disorders and Stroke Cerebral Palsy Common Data Elements project. *Developmental Medicine & Child Neurology*, 66(5), 610–622. <https://doi.org/https://doi.org/10.1111/dmcn.15743>
- Sheikh, A. M., Hossain, S., & Tabassum, S. (2024). Advances in stem cell therapy for stroke: mechanisms, challenges, and future directions. *Regenerative Medicine Reports*, 10–4103. <https://doi.org/0.4103/RMR.REGENMED-D-23-00002>
- Siddiq Hassan, D. (2013). The effect of feature selection methods on machine learning model performance: A comparative study for breast cancer prediction. *Science Journal of University of Zakho*, 13(1), 101–112. <https://doi.org/10.25271/sjuoz.2024.12.3.1429>
- Sindhura, C., Al Fahim, M., Yalavarthy, P. K., & Gorthi, S. (2024). Fully automated sinogram-based deep learning model for detection and classification of intracranial hemorrhage. *Medical Physics*, 51(3), 1944–1956. <https://doi.org/https://doi.org/10.1002/mp.16714>
- Studer, M., & Thompson, C. R. (2024). Prevention practice for neurological conditions. In *Prevention Practice and Health Promotion* (pp. 241–265). Routledge. <https://doi.org/https://doi.org/10.4324/9781003525882>
- Suryadi, B. (2024). Methods for detecting early symptoms of stroke: A literature review. *Jurnal Ilmiah Ilmu Keperawatan Indonesia*, 14(01), 32–43. <https://doi.org/https://doi.org/10.33221/jiiki.v14i01.3165>
- Szabó, S., Holb, I. J., Abriha-Molnár, V. É., Szatmári, G., Singh, S. K., & Abriha, D. (2024). Classification assessment tool: A program to measure the uncertainty of classification models in terms of class-level metrics. *Applied Soft Computing*, 155, 111468. <https://doi.org/10.1016/j.asoc.2024.111468>
- Tin, T. A., Aye, M. M., Khin, E. E., Oo, T., Tun, H. M., & Pradhan, D. (2024). Performance optimization of brain tumor detection and classification based MRI by using batch normalization algorithms in deep convolution neural network. *Journal of Novel Engineering Science and Technology*, 3(03), 66–72. <https://doi.org/10.56741/jnest.v3i03.567>
- Zhang, R., Ding, R., Wang, Q., Zhang, L., Fan, X., Guo, F., Chen, X., Jiang, C., Cao, J., & Wang, J. (2024). Inflammation in intracerebral hemorrhage: A bibliometric perspective. *Brain Hemorrhages*, 5(3), 107–116. <https://doi.org/https://doi.org/10.1016/j.hest.2024.01.003>