Classification of Brain Hemorrhage Using Deep Learning from CT Scan Images

Nipa Anjum, Abu Noman Md. Sakib, and Sk. Md. Masudul Ahsan

Department of Computer Science and Engineering
Khulna University of Engineering & Technology, Khulna-9203, Bangladesh
nipaanjum1998@gmail.com, abunomanmd.sakib@gmail.com,
smahsan@cse.kuet.ac.bd

Abstract. Brain Hemorrhage is a life-threatening problem that happens by bleeding inside human head. In this study, Computed Tomography (CT) scan images have been used to classify whether the case is hemorrhage or non-hemorrhage. Different Convolutional Neural Network (CNN) models have been observed along with some pre-trained deep learning models such as VGG16, VGG19, ResNet150, ResNet152 and InceptionV3. Pre-trained models have performed well on the dataset but all of them are heavyweight architectures in terms of number of total parameters. But the proposed model is a lightweight architecture as well as a well performing one. After evaluating the model performance, it has been observed that the proposed model gave 96.67% accuracy, 97.08% sensitivity and 96.25% specificity which is the best among other custom CNN models.

Keywords: Brain Hemorrhage Classification, Deep Learning, Convolutional Neural Network, InceptionV3, VGG16, VGG19, ResNet50, ResNet152.

1 Introduction

Brain hemorrhage is a potentially fatal condition that can be caused by physical trauma or a variety of medical problems such as high blood pressure [1]. When the artery in human brain bursts and starts bleeding inside the brain tissue, it causes the damage in human brain which leads to hemorrhage. Due to high blood pressure, aneurysm, bleeding disorder, brain tumor or any kind of trauma, brain hemorrhage can occur. If any person suddenly feels severe headache, weakness in arm or leg that makes him/her numb, starts vomiting and losing consciousness, these can be an augury of brain hemorrhage [2]. For diagnosing a person who is surmised of brain hemorrhage, physical symptoms are measured then brain images are observed. These images are obtained by Computed Tomography (CT) Scan, Magnetic Resonance Imaging (MRI) and Magnetic Resonance Angiogram (MRA) techniques.

Among the world population, the incidence rate of brain hemorrhage in black, white, Asian, Hispanic people were 22.9, 24.2, 51.8 and 19.6 respectively. This survey was per 100,000 per-years. They also showed that the rate of hemorrhage

increases in people who are more than 45 years old [3]. People who have had hemorrhage, has a chance to survive if proper treatment is provided instantly. For this, faster classification of brain hemorrhage is important. After surviving Intracerebral Hemorrhage, among 72,432 people, almost 18% died due to infection, cardiac disease, respiratory failure, ischemic stroke, recurrent Intracranial Hemorrhage within 4 years [4]. The percentage of fatality for a person diagnosed with brain hemorrhage might be as high as 57 percent depending on the type of hemorrhage [5]. Automated brain hemorrhage classification systems help to diagnose hemorrhage in human brain that helps to increase survival rate in human life. Brain hemorrhage can occur in older and younger people. Sometimes due to birth injury or force, infants can also have brain hemorrhage.
People are doing research works for finding better solutions to classify Brain Hemorrhage for doctors to give better and reliable treatments to the patients. Researchers [6, 7, 8, 9, 10] used classical image processing, machine learning algorithms or combines both algorithms while some of them worked with deep learning approaches. Working with medical dataset is very challenging task as the resource is very scarce. Image processing techniques require huge time to classify images and if the size of the dataset increases, it becomes very tedious. In medical image classification, identification of the problem requires high domain knowledge but the performance of such methods may not be appeasable. Machine learning techniques at first extracts the features with the help of experts in specific domain. In this feature extraction procedure, different image processing algorithms can help to better understand the data by making patterns more evident and reduce the complexity of data. On the contrary, deep learning works well when the dataset size increases and deep learning algorithms extracts the high-level features by learning the dataset on its own. This eliminates the knowledge of a specific domain which helps to solve any problem with limited domain knowledge. This has motivated to implement a deep learning-based approach to build a convolutional neural network that can classify brain hemorrhage. The focus in this research is to work on a small dataset and build a solution that can diagnose Brain Hemorrhage with an efficient result which can run on minimal system requirements. Brain hemorrhage can be of different types, the proposed method classifies whether there is a hemorrhage or not, overlooking the type of hemorrhage. The contribution of this research work can be stated as:

- Designed a light-weight automated system that would classify hemorrhage and non-hemorrhage accurately.
- Introduced a novel approach for classifying brain hemorrhage.
- Conducted a comparative study on brain hemorrhage classification.
- delivered an accurate prediction within a short time.

In section 2, some previous researches relative to the work have been reviewed, section 3 describes work methods that have been followed and information about the dataset, section 4 contains the evaluation matrix, experimental result and analysis of the proposed method. At last in section 5, the research work has been concluded.

2 Literature Review

Some remarkable works previously done on brain hemorrhage classification have been discussed in this section.

Napier et al. [6] proposed a CAD system that used different image processing techniques using different filters such as the Gaussian Filter, the Median Filter, the Bilateral Filter and the Wiener Filter and Morphological operations have been used to detect brain hemorrhage from CT scan. 36 head CT scan images were used to execute the method in this study. The proposal achieved an accuracy of 88.89%, a precision of 91.259%, a specificity of 94.4% and sensitivity of 94.4%. Srivastava et al. [7] proposed a way to classify hematomas in brain CT images by using Support Vector Machine using 150 brain CT scan images. An average accuracy of 88% is gained from O-V-A SVM while O-V-O SVM gained 97% accuracy. Their method SVM (O-V-O) outperforms other classical machine learning approaches by some metrics.

Vrbancic et al. [8] they proposed a concept that was to tune transfer learning model based on Grey Wolf Optimizer algorithm. The dataset had a total of 200 images from Head CT-Hemorrhage [11] dataset. They have used a CNN based model, VGG-16 as their model architecture. After applying their GWOTLT method, they were able to secure 91% accuracy. Mushtaq et al. [9] proposed BHCNet to classify the existence of hemorrhage in human brain. During preprocessing, the images were resized to 128x128 pixels and different image augmentations were applied to increase the size of the dataset. The images were split into train and test set by maintaining a 90% to 10% ratio. The performance of the proposed architecture was evaluated using metrics such as accuracy, precision, sensitivity, specificity, F1-score. The best model was found at 24 epochs with an accuracy of 95%.

Patel et al. [10] proposed an architecture with combination of VGG-like CNN with bidirectional LSTM. The CNN was initially trained using the dataset. The pre-trained CNN model was later used to initialize full model and two bidirectional LSTM layers were added. They have achieved 98% sensitivity and 78% specificity. End to end training without using pre-trained CNN model gives inferior result. Training the CNN and then training again after adding layers is a costly process.

3 Research Methodology

Image classification refers to the task of identifying the actual class of an image. Figure 1 shows the work flow of the classification task. Images in the Head CT – Hemorrhage [11] dataset have been resized and split into training set, test set and validation set. Multiple augmentation techniques have been applied for the classification of brain hemorrhage. Different deep learning models have been used to perform the classification task. The trained models are evaluated with the help of test dataset. The predicted results have been compared with the actual outcomes using multiple evaluation metrics to test the performance.

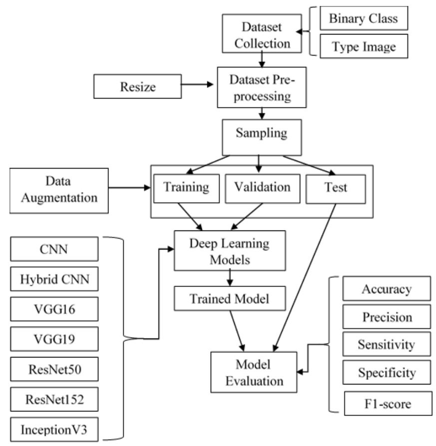
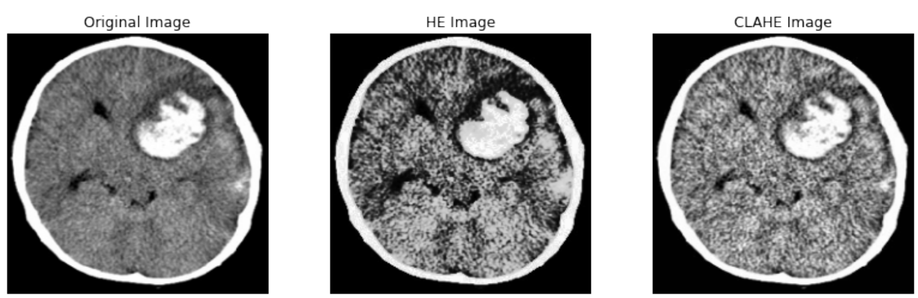


Fig. 1: Flow chart of hemorrhage classification process.

3.1 Data Preprocessing and Augmentation

Images have been resized to 224x224 for the model requirement. The size of the dataset is artificially increased by using augmentation. After augmenting the dataset, variations in the dataset can be found. Rotation of 0 to 10 degrees clockwise, vertical and horizontal with a range of 0 to 0.2, zoom range of 0.2 and horizontal flip are used for the augmentation procedure. After augmentation, a total of 2400 images have been acquired where training, validation and testing each set has 1536 (64%), 384 (16%) and 480 (20%) images, respectively. As model learns in training period, most of the features are used in training which is the reason for using large portion of dataset in training set.



(a) Original image (b) HE image (c) CLAHE image

Fig. 2: Hemorrhagic image and their output after histogram equalization and contrast limited adaptive histogram equalization.

3.2 Histogram Equalization (HE)

Frequency of pixel intensity levels in an image can be represented by histogram. The contrast of an image can be tuned by using histogram equalization. It improves the image's contrast by stretching out the intensity range or spreading out pixel intensity that are most frequent which results in more contrast in low contrast sections of the image [12]. In Figure 2b, equalized hemorrhagic image is shown.

3.3 Contrast Limited Adaptive Histogram Equalization (CLAHE)

The image is split into equal sized rectangular blocks, with each block undergoing histogram correction. Histogram formation, clipping, and redistribution are all parts of histogram adjustment. The difference in CLAHE is that it uses a clip point to chop off the peak value in each block's histogram by limiting the contrast which are redistributed to each gray level [13]. Figure 2c shows how the original image is enhanced after applying contrast limited adaptive histogram equalization on it.

3.4 Proposed CNN Model

To better anticipate the hemorrhage, different CNN architectures have been designed with variation in convolution layers, max pooling layers, flatten layers, and fully connected layers in all approaches. Proposed CNN model has five convolution layers with 3x3 sized filters followed by relu activation function. After each convolution layer, the output is downsampled using max pooling layer having a 2x2 pool. the output of last pooling layer has been flattened and then passed through the fully connected layers, followed by relu activation function. 40% of the neurons have been dropped out after each dense layer. In the output layer, sigmoid activation function has been used for the binary classification problem.

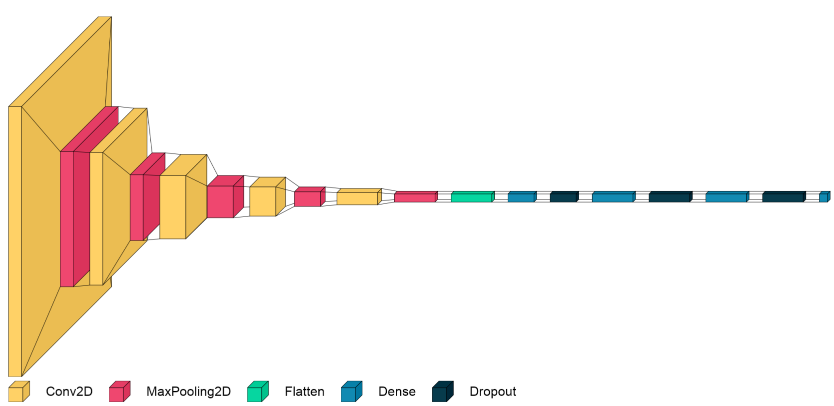


Fig. 3: Proposed CNN architecture.

3.5 Hybrid CNN Model

A hybrid CNN architecture has been built to experiment the models learning capability. The architecture has 6 convolution layers with same padding followed by ReLU activation function, 3 concatenation layers, 3 max pooling layers, 3 fully connected layers followed by ReLU activation function and 3 dropout layers. The concatenation layer concatenates the immediate output of the convolution layer with the previous convolution layer output. A max pooling layer is added after each concatenation. Sigmoid activation function is used in the output layer. Figure 4 illustrates the model architecture of hybrid CNN.

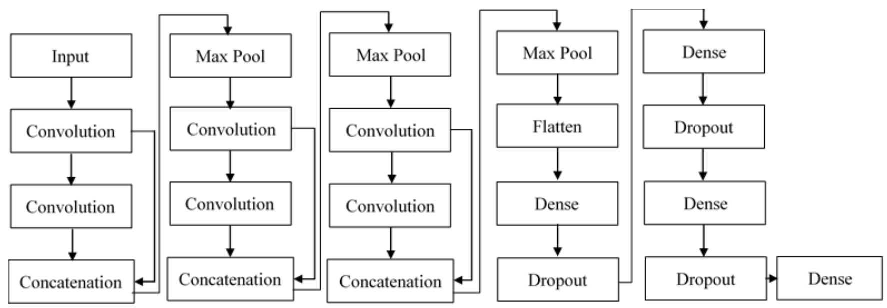


Fig. 4: Hybrid CNN model.

3.6 Pre-trained Models

Transfer learning is the process of using feature representations from a previously trained model to avoid training a new model from the ground up. This previously trained models, known as pre-trained models are typically trained on large datasets [14] and these are used as benchmark models as they are used in various applications of computer vision with great performance. These models reduce the training time and less erroneous. When there is a small training dataset, transfer learning works well by initializing the weights from pre-trained models. Previously acquired knowledge which is weights, is used in training new dataset. VGG16 [15], VGG19 [15], ResNet50 [16], ResNet152 [16], InceptionV3 [17] pre-trained models from keras have been used to classify brain hemorrhage. These models are trained on ImageNet dataset and works well on image classification as it contains 1000 classes in the dataset.

4 Results and Discussion

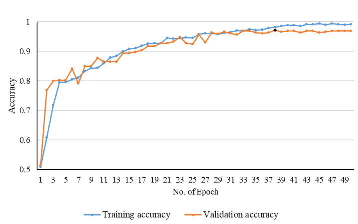
Building this brain hemorrhage classification system and training models, a computer having 2.5GHz duel-core Intel Core i5-7200U processor with 8GB RAM and NVIDIA GeForce 940MX has been used. Python and tensorflow have been used. Google Colaboratory with 12.7 GB RAM and 12 GB NVIDIA TESLA K80 GPU and Kaggle with 16 GB RAM and 13 GB NVIDIA TESLA P100 GPU are primary platform for model training and evaluation of the system.

4.1 Dataset

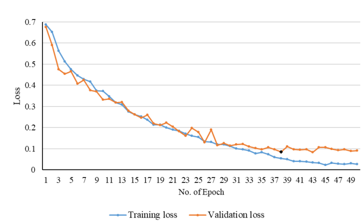
Dataset plays an important role in any research work. It is tough to conduct medical research work as there is scarcity in finding appropriate data. In this study, the dataset, Head CT – Hemorrhage [11] is used that contains 200 images in which 100 images are of hemorrhagic brain and 100 images are of non-hemorrhagic brain. Deep learning model requires a large number of data with all possible variations to train the model. A small sized dataset is sometimes beneficial as it requires less time to train and the likelihood of overfitting reduces. But another risk in that the accuracy is not appeasable. The problem in the size of the dataset has been slightly solved by using data augmentation techniques.

4.2 Brain Hemorrhage Classification

The results determine the performance of those models. In brain hemorrhage classification problem, confusion matrix, accuracy, precision, recall, F-1 score and specificity have been used as evaluation metrics. Figure 5a shows the model accuracy where the orange line is for validation and blue line is for training. Figure 5b shows the binary cross entropy loss graph where the blue line indicates training and orange line indicates the validation loss. The black dot has been added to show the point where the best model has been saved which is at 38th epoch by observing the validation accuracy criteria where the accuracy is maximum. In CNN approach, multiple CNN configurations have been observed to find out the best combination for the outcome. Table 1 shows different configurations of CNN models that have been tried out to find the best one. Number of hidden layer has been changed in each model where the number of filters also varied and in fully connected layers, number of neurons have also been changed.



(a) Training and validation accuracy graph.



(b) Training and validation loss graph.

Fig. 5: Training and validation performance of proposed CNN.

Table 1: Model descriptions of different CNN models.

Model	No. of convolution layer	No. of dense layer	Dropout	No. of parameters	filter size	Pool size	Activation function
CNN-1	3	3	0.40	697,057			
CNN-2	5	3	0.40	336,865	3x3	2x2	Relu and
Proposed CNN	4	3	0.40	385,889			sigmoid

Figure 6 shows the accuracy, precision, sensitivity, F1-score and specificity of all models. Comparing between the proposed CNN and InceptionV3, the pre-trained InceptionV3 model has achieved the higher accuracy. Observing the other metrics of all nine models, proposed CNN predicts that someone has brain hemorrhage, which is correct 96.28% of the time. It is the precision of the model which is higher than other custom models and sensitivity of the proposed model which means actual hemorrhagic cases that are predicted as hemorrhage is 97.08%. The proposed CNN achieved 96.68% F1-score. Specificity of the proposed model is 96.25% which determines the non-hemorrhagic cases are predicted to be non-hemorrhage 96.25% of the time. Figure 7 visualizes the ROC Curve Analysis of the proposed model along with the other models. It can be seen that InceptionV3 model architecture functions admirably, exceeding the performances than custom trained model. But the pre-trained models are all heavyweight models in terms of total parameters. The number of parameters of the InceptionV3 model makes it computationally very expensive. On the other hand, the architecture proposed in this study is both a lightweight and a high-performing one for practical use.

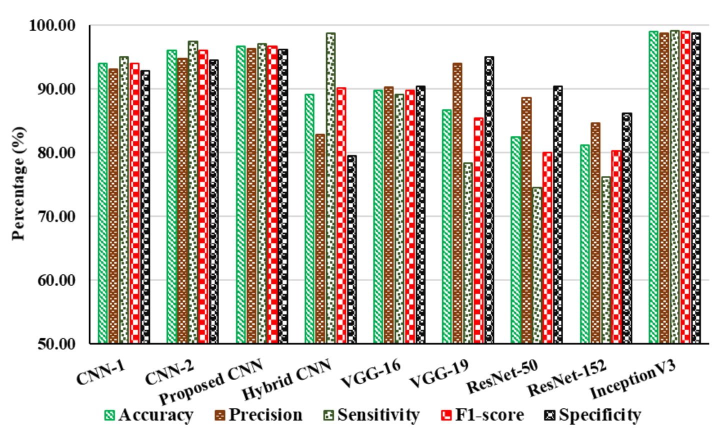


Fig. 6: Comparison with respect to different evaluation metrics of CNN models on hemorrhage data.

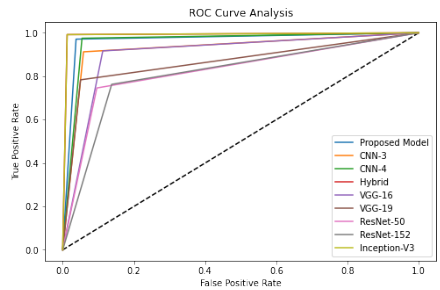


Fig. 7: ROC curve analysis.

Table 2: Performance comparison with respect to parameters.

Models	**Parameters**
VGG16	17,926,209
VGG19	23,235,905
ResNet50	126,350,209
ResNet152	161,133,441
InceptionV3	21,853,985
Hybrid CNN	13,265,889
Proposed CNN	**385,889**

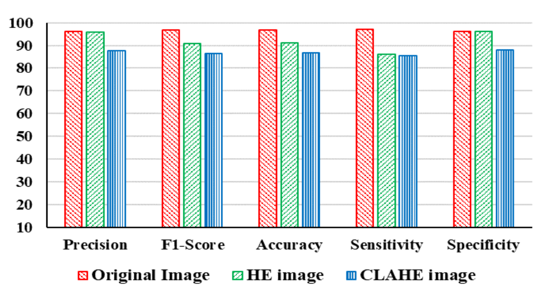


Fig. 8: Performance comparison with original, histogram equalized and contrast limited adaptive histogram equalized dataset.

As, it is seen from Table 2 that the number of parameters of the custom CNN is reasonably small in comparison to other models. Pre-trained models are trained with a large number of parameters and Inception v3 model has the lowest parameters among the other pre-trained models. To see whether the model performs better after increasing the number of parameters, the hybrid CNN

model has been evaluated and the performance has not fulfilled the expected result. The models with a smaller number of parameters work better on the dataset. Inception V3 has less parameters than VGG16, VGG19, ResNet50 and ResNet152 and the model is trained with a large dataset, it performed better but in terms of parameters, the proposed model has got less parameters than this pre-trained one. Two variations on the dataset has been made after applying HE and CLAHE and fed to the proposed CNN. The result of both datasets could not outperform the previous result as there is a large area of low-intensity. Figure 8 shows the performance of proposed model on these datasets.
In Figure 9, among 480 test images, four images are shown with the actual and predicted labels. The model predicted 233 hemorrhagic and 231 non-hemorrhagic images correctly. Total 16 images have been misclassified, 9 hemorrhagic images are predicted as non-hemorrhagic and 7 non-hemorrhagic images are predicted as hemorrhagic. Due to the complexities in CT scan images where head slices are taken from different direction, there are some images where white portions increases that lead the model toward misclassification. Also, in some images, the hemorrhagic part is diminutive to detect by the model.
Table 3 shows the validation accuracy of each fold. The average accuracy after 5-fold cross validation is 96.46% which doesn't differ much from the models accuracy proving that the model has not been overfitted using this dataset. The comparison with other state of the art works with the same dataset is shown in Table 4. It shows the similarities and differences among the previously proposed methods and the proposed method. 5-fold cross validation has been executed to check the performance.

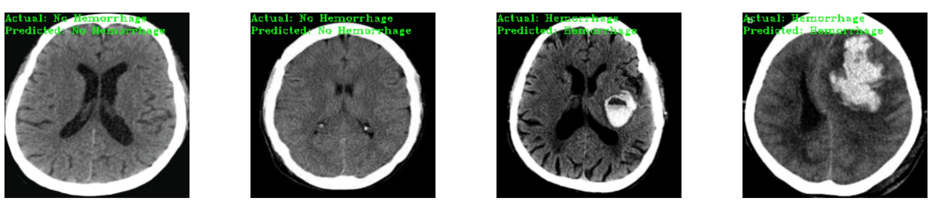


Fig. 9: Sample of some predictions.

Table 3: Cross validation results of proposed model.

Iteration	**1**	**2**	**3**	**4**	**5**	**Average**
Accuracy (%)	96.88	96.25	95.83	95.63	97.71	96.46

Table 4: Comparison with previous works with same dataset.

Literature	Year	Technique	Accuracy (%)	F1-score (%)	No. of parameters
Greywolf [8]	2019	VGG16 + Grey wolf optimizer algorithm	91.00	91.45	-
BHCNet [9]	2021	CNN	95.00	95.23	-
Transfer learning	2022	InceptionV3	**98.96**	**98.96**	21,853,985
Proposed Model	2022	CNN	96.67	96.68	**385,889**

5 Conclusion

In this study, different CNN models with variations in dataset and a comparison among the performances of pre-trained deep learning models and previous works with the proposed CNN model have been done. The most difficult part of this was working with very little dataset whereas a deep learning model requires a large number of images. It is not easy to get medical image dataset easily and requires knowledge to understand the data which sometimes can be expensive. Data augmentation have added the solution to this problem by increasing the amount and variation of the dataset. The proposed method is capable of solving the brain hemorrhage classification task faster than other existing works. In future, the dataset can be enhanced with more images. Multi-class classification can be done by using different types of hemorrhage. Also the localization of brain hemorrhage can be another inclusion. For improving the model performance, ensemble learning, Long Short-Term Memory (LSTM), Gated Recurrent Units (GRU) with CNN can be tried out.

REFERENCES

[1] Jeremy J Heit, Michael Iv, and Max Wintermark. "Imaging of intracranial hemorrhage". In: *Journal of stroke* 19.1 (2017), p. 11.

[2] Mohammed Ammar et al. "Deep Learning Models for Intracranial Hemorrhage Recognition: A comparative study". In: *Procedia Computer Science* 196 (2022), pp. 418–425.

[3] Charlotte JJ Van Asch et al. "Incidence, case fatality, and functional outcome of intracerebral haemorrhage over time, according to age, sex, and ethnic origin: a systematic review and meta-analysis". In: *The Lancet Neurology* 9.2 (2010), pp. 167–176.

[4] Lindsey R Kuohn et al. “Cause of death in spontaneous intracerebral hemorrhage survivors: multistate longitudinal study”. In: *Neurology* 95.20 (2020), e2736–e2745.

[5] K Haselsberger, R Pucher, and LM Auer. “Prognosis after acute subdural or epidural haemorrhage”. In: *Acta neurochirurgica* 90.3 (1988), pp. 111–116.

[6] John Napier et al. “A CAD System for Brain Haemorrhage Detection in Head CT Scans”. In: *IEEE EUROCON 2019-18th International Conference on Smart Technologies*. IEEE. 2019, pp. 1–6.

[7] Devesh Kumar Srivastava, Bhavna Sharma, and Ayush Singh. “Classification of hematomas in brain ct images using support vector machine”. In: *Information and Communication Technology for Sustainable Development*. Springer, 2018, pp. 375–385.

[8] Grega Vrbancic, Milan Zorman, and Vili Podgorelec. “Transfer learning tuning utilizing grey wolf optimizer for identification of brain hemorrhage from head ct images”. In: *StuCoSReC: proceedings of the 2019 6th student computer science research conference*. 2019, pp. 61–66.

[9] Muhammad Faheem Mushtaq et al. “BHCNet: neural network-based brain hemorrhage classification using head CT Scan”. In: *IEEE Access* 9 (2021), pp. 113901–113916.

[10] Ajay Patel et al. “Image level training and prediction: intracranial hemorrhage identification in 3D non-contrast CT”. In: *Ieee Access* 7 (2019), pp. 92355–92364.

[11] Felipe Kitamura. *Head CT - hemorrhage*. https://www.kaggle.com/felipekitamura/head-ct-hemorrhage. [Online; accessed 21-March-2022].

[12] Sunanda Das, OFM Riaz Rahman Aranya, and Nishat Nayla Labiba. “Brain tumor classification using convolutional neural network”. In: *2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT)*. IEEE. 2019, pp. 1–5.

[13] Yakun Chang et al. “Automatic contrast-limited adaptive histogram equalization with dual gamma correction”. In: *Ieee Access* 6 (2018), pp. 11782–11792.

[14] Sajja Tulasi Krishna and Hemantha Kumar Kalluri. “Deep learning and transfer learning approaches for image classification”. In: *International Journal of Recent Technology and Engineering (IJRTE)* 7.5S4 (2019), pp. 427–432.

[15] Karen Simonyan and Andrew Zisserman. “Very deep convolutional networks for large-scale image recognition”. In: *arXiv preprint arXiv:1409.1556* (2014).

[16] Kaiming He et al. “Deep residual learning for image recognition”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 770–778.

[17] Christian Szegedy et al. “Rethinking the inception architecture for computer vision”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 2818–2826.