# BHCNet: Neural Network-Based Brain Hemorrhage Classification Using Head CT Scan

**MUHAMMAD FAHEEM MUSHTAQ**[1], **MOBEEN SHAHROZ**[2], **ALI M. ASEERE**[3], **HABIB SHAH**[3], **RIZWAN MAJEED**[4], **DANISH SHEHZAD**[5], **AND ALI SAMAD**[6]

[1]Department of Artificial Intelligence, The Islamia University of Bahawalpur, Bahawalpur, Punjab 63100, Pakistan
[2]Department of Computer Science, The Islamia University of Bahawalpur, Bahawalpur, Punjab 63100, Pakistan
[3]Department of Computer Science, King Khalid University, Abha 62529, Saudi Arabia
[4]Directorate of IT, The Islamia University of Bahawalpur, Bahawalpur, Punjab 63100, Pakistan
[5]Department of Computer Science, National University of Computer and Emerging Sciences, Chiniot-Faislabad Campus, Islamabad 35400, Pakistan
[6]Department of Data Science, The Islamia University of Bahawalpur, Bahawalpur, Punjab 63100, Pakistan

Corresponding author: Muhammad Faheem Mushtaq (faheem.mushtaq@iub.edu.pk)

**ABSTRACT** Brain Hemorrhage is the eruption of the brain arteries due to high blood pressure or blood clotting that could be a cause of traumatic injury or death. It is the medical emergency in which a doctor also need years of experience to immediately diagnose the region of the internal bleeding before starting the treatment. In this study, the deep learning models Convolutional Neural Network (CNN), hybrid models CNN + LSTM and CNN + GRU are proposed for the Brain Hemorrhage classification. The 200 head CT scan images dataset is used to boost the accuracy rate and computational power of the deep learning models. The major aim of this study is to use the abstraction power of deep learning on a set of fewer images because in most crucial cases extensive datasets are not available on the spot. The image augmentation and imbalancing the dataset methods are adopted with CNN model to design a unique architecture and named as Brain Hemorrhage Classification based on Neural Network (BHCNet). The performance of the proposed approach are analyzed in terms of accuracy, precision, sensitivity, specificity and F1-score. Further, the experimental results are evaluated by comparative analyses of the balanced and imbalanced dataset with CNN, CNN + LSTM and CNN + GRU models. The promising results are achieved with CNN by imbalancing the dataset and gain highest accuracy that outperforms the hybrid CNN + LSTM and CNN + GRU models. The results reveals the effectiveness of the proposed model for accurate prediction to save the life of the patient in the meantime and fast employment in the real life scenario.
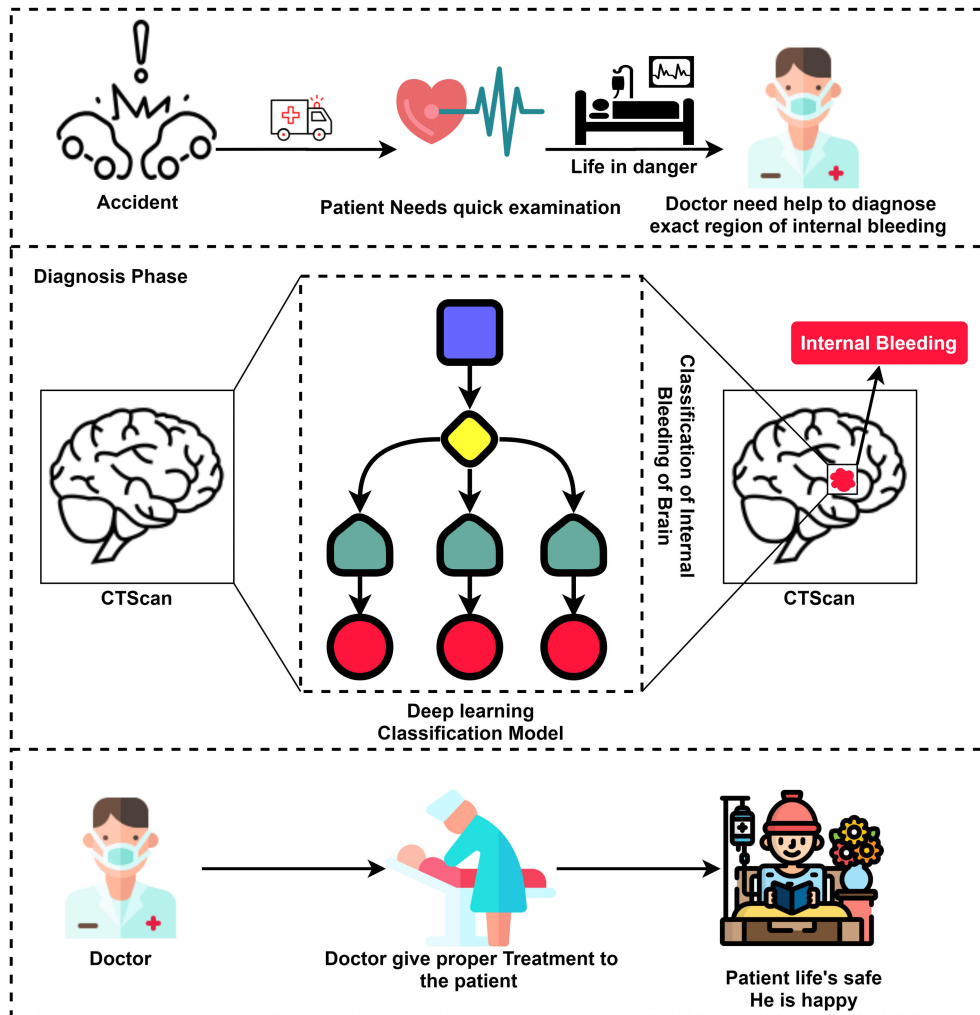
**INDEX TERMS** Brain hemorrhage, image classification, deep learning, CNN, image augmentation, CT scan.

## I. INTRODUCTION

Hemorrhage [1], [2] is a medical term referring to bleeding within or out of the body. Internal bleeding of the brain is known as Brain Hemorrhage [3]. This is caused by a sudden blood clot [4] in arteries that supply blood to the brain or internal bleeding in the surrounding tissues of the brain due to rupturing of the arteries [5]–[8]. The brain cells got damage by this bleeding and the most common causes [9] are

The associate editor coordinating the review of this manuscript and approving it for publication was Guitao Cao.

trauma [1], high blood pressure [10], aneurysm, blood vessel abnormalities [11], amyloid angiopathy, bleeding disorders and brain tumors [12], [13]. Those are the major causes of death and severe disability. The brain hemorrhage is the cause of the 30% of deaths in the united states in 2013 [14] by the ratio of 100,000:7 in the west and 100,000:200 people in Asia. Moreover, women are affected more than men by the ratio of 3:2 and 80% of people born with weak spots in brain arteries [15]. According to the World Health Organization (WHO) report of 2009 patients suffer from stroke was 15 million, in which 5 million died and 5 million were disabled [16].

**FIGURE 1.** Brain hemorrhage classification using the CNN model to diagnose the region of the internal bleeding in the CT scan images of the Brain.

Medical experts [17] refers to the fast diagnoses and effective initial treatment in these cases to prevent disability and death [18]. The Computed Tomography(CT) scan [19] and Magnetic Resonance Imaging (MRI) [20] are used to visualize the internal structure of skull and brain. The medical imaging technique such as CT scan preferred by the medical experts than the MRI to analyze the internal structure of the human body include the brain because it provides the wider availability, lower cost and sensitiveness to early diagnoses of the brain hemorrhage. The CT scan is a collective combination of different X-ray images in cross-sectional views generated by computer with different angles using 3D imaging [21]. The X-ray [6], [22] beams are transmitted by the CT scanner in an arc that allows to capture tissues with different intensity levels depend on the X-ray absorbency level of the tissues. The CT scan provides detailed information about the internal body structure and tissues of the solid organs that helps the medical experts to diagnose the internal bleeding and blood accumulation in the brain [6]. The CT scan is mostly used in emergencies to diagnose infection, tumor,

traumatic injuries [23] and hemorrhagic strokes inside the body that are difficult to identify by the person [14]. The CT scan gain preference instead of MRI because it provides the fast acquisition for initial treatment. The identification of the brain hemorrhage is a challenging phase because it is caused by internal bleeding in the head [8] and the medical experts also need years of experience to identify the region of bleeding in the CT scan. There is a life of the patients is at risk and every single life is important. In previous studies, the researchers put their efforts and knowledge to make the perfect system to diagnose the exact region of brain hemorrhage but failed because of deficiencies like diagnosis processes took too much time and performance evaluations illustrates not enough results to save the life of every single patient.

The Brain Hemorrhage Classification Using NN (BHCNet) system is proposed to distinguish the brain hemorrhage using head CT scan image based on Convolutional Neural Network (CNN) as shown in Figure 1. The unique design of the BHCNet system that is based on newly designed

layered architecture of CNN with image augmentation and imbalancing methods by using small dataset of 200 head CT scan images to increase the abstraction power, prediction speed and accuracy. The contributions of this study are as follows:

- The BHCNet diagnosis system for the Brain Hemorrhage patients proposed in the presence of small dataset that would be helpful in critical times where extensive dataset are not available.
- The image preprocessing methods such as resize the image, flipping of image and Image augmentation is adopted to enhance the efficiency of training process to gain the highest accuracy rate and performance of the CNN.
- The experiments are carried out with deep learning models CNN, hybrid CNN + LSTM and CNN + GRU are proposed based on layered architecture using head CT scan images. The structure of layers are consists of a convolutional layer, max-pooling layer, global average pooling layer and the dense layer as shown in Figure 5.
- The performance evaluation matrices i.e. accuracy, precision, sensitivity, specificity and F1 score is used to evaluate the effectiveness of the proposed study.
- Further, the proposed approach and experimental results are compared by carried out different experiments using hybrid deep learning models with balanced and imbalanced dataset.

The structure of the preceding paper is organized as follow: Section II presents the existing and previous efforts of researchers as a related work. Section III presents the dataset. Section IV discusses the methods and methodology of the proposed study. Section V is the results and evaluation analyses of the performance. Section VI is the conclusion of this study.

## II. RELATED WORK

In recent times, neural networks are increasingly leveraging at the provisioning of smart systems for health diagnostics and treatment. The machine learning [24] and deep learning [25]–[28] approaches were developed in previous works for the automated diagnoses of a brain hemorrhage [29], [30]. The brain hemorrhage classification proposed by different authors using different methods to diagnose specific types. For this purpose, MRI [31] and CT scan [32], [33] were used to classify the patient has brain hemorrhage or not. The Naive Bayes [34], K-Means clustering [35], Image Segmentation [36], Multi-class classification [37], Recurrent Neural Network (RNN) [38], Long Short-Term Memory (LSTM) [39], CNN [40] and hybrid models [41] were used to classifies the brain hemorrhage. The small dataset and large dataset are used to validate their work by different authors.

Deep learning methods were applied [42] for brain hemorrhage classification using a Computer Tomography (CT) [19] scan. This work consists of deep convolutional neural networks and auto-encoders method that depends on three hidden layers and achieved better recognition rate of 89.6%
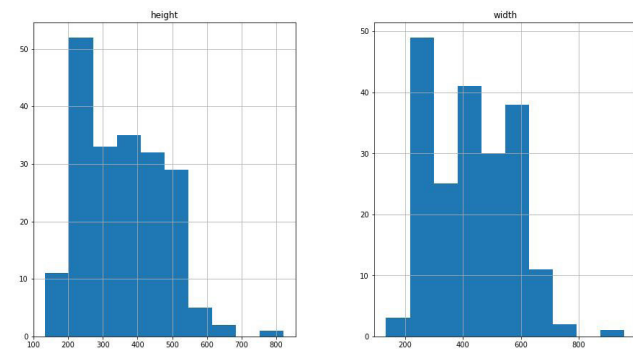
with CNN, 90.9% while using stack Autoencoder and lower Mean Squared Error (MSE) [40], [43] rate of 0.0021 with stacked auto encoder and 0.099 with CNN using the characterization curve. This approach takes longer time with 12000 iterations which make it more complex and time consuming as well as using larger dataset. The CNN model has been implemented [5] with Gray Level Co-occurrence Matrix(GLCM) [44] to assists the Computer Aided Diagnoses (CADx) [45] system for the classification of all types of medical images like MRI and CT. The proposed method was used to extract and convert the information of irregular segmentation regions of the image into fixed-size GLCM input for CNN. Furthermore, the 3D CT scan images are used to classify ICH based on the combination of CNN and LSTM [41]. CNN helps to identify the ICH in axial slices and LSTM help to analyze the obtained information from the classification of slices. They enhanced the classification model of ICH using 3D non-contrast CT scan without using preprocessing. K-Mean algorithm [46] is the unsupervised learning technique that was use for the segmentation of the brain MRI images with the integration of the dual-tree complex wavelet transform. Expectation Maximization Segmentation (EMS) software [47] is used to extract the infected regions of the brain by segmentation and corrected the intensity of inhomogeneities. The performance of the sector-based segmentation method is used to handle the intensity inhomogeneities. For this purpose, Local region-based level set method hybrid approach with a variation of fuzzy clustering was proposed. Identification of ICH is difficult when the bleeding occurs in a small area and when the medical person is not experienced. This Approach used to detect small region based ICH and higher attenuation signal caused by SAH [48].

The CNN model [49]–[51] improved to classify medical color images by automatically selecting misclassified negative samples during the training phase by tuning the hyper-parameters to compare Selective Sampling (SeS) [52] and Non-Selective Sampling (NSeS). The SeS with CNN outperforms NSeS by achieving 0.89 and 0.97 on two data sets. Hssayeni *et al.* [21] present two folds contribution. Firstly, the data set collects from CT scans of traumatic brain injury that manually outlined by two experienced radiologists [53], [54]. Second, apply the CNN on these image segmentation also known as U-Net and performance is evaluated by 5 fold cross-validation with a dice coefficient of 0.31. A novel approach NB-PKC [55] used to apply an image mask on MRI images for the detection of brain hemorrhage. The binary thresholding process used to get minimal local binary pattern and GLCM for segmentation but 13% results improved than the usual Support Vector Machine (SVM) algorithm, still not able to detect very small subarachnoid hemorrhage. The CNN model with the Gray Level Co-occurrence Matrix(GLCM) [5] presented to assists the Computer-Aided Diagnoses (CADx) system for the classification of all types of medical images like MRI and CT. The proposed method was used to extract and convert the information of irregular segmentation regions of the image into fixed-size GLCM input

for CNN. The Recurrent Attention DenseNet (RADnet) [56] combine with Recurrent Neural Network (RNN) layers [57] to perform slice level prediction and achieved better results than the benchmarked RADnet performance on an analysis of 77 CT scans by three senior radiologists. One dimensional CNN [58] is used to extract semantically co-located features, LSTM for sequential features and logistic function used to classify ICH from the radiologist reports. There were 12,852 CT radiologist reports used for training and testing purpose. The performance is evaluated by using a receiver operator curve (ROC) [59], [60] and achieved 0.94 area under the curve (AUC).
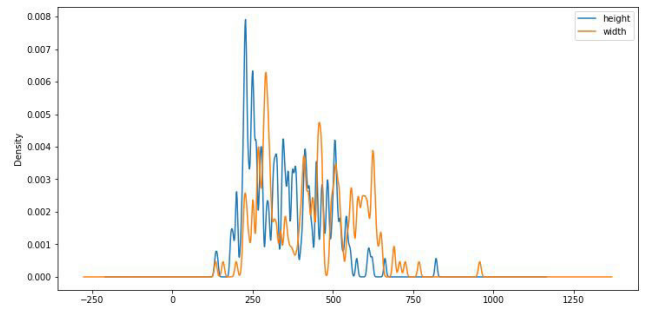
## III. HEAD CT SCAN IMAGES DATASET

The CT scan is the easiest and efficient way to diagnose the internal bleeding, traumatic injuries and disrupted of veins in the body that cause strokes or death. The dataset is consists of the equally balanced 200 CT scan images of the brain hemorrhage and non-brain hemorrhage patients.



**FIGURE 2.** The CT scan image dataset visualization based on variation in the height and width of the images.

There are 100 CT scan images of the Hemorrhage patients and 100 are non-brain hemorrhage. This is the labelled dataset that having two classes such that hemorrhage and non-brain hemorrhage. The images consists of different height and width as shown in Figure 2 that presents the different variation in images. This is the small dataset with some benefits that increases the speed of the training process of the CNN model and reduces the chances of the overfitting but it also reduces the accuracy rate as well which is the challenging part. This is the reason that the selection of this dataset is efficient for this study. First this dataset needs some concentration to enhance the accuracy rate by applying preprocessing as discussed in section IV-A.

The visualization of the CT scan image dataset based on the variation of height and width of images shown in Figure 2 and the visualization based on the variation in the density of images is shown in Figure 3. This disrupts the learning of the CNN model because CNN models need a fixed size of input data. The line chart 3 shows the density-based distortions in the images referred to as a noise in the images according to height and width.



**FIGURE 3.** The CT scan image dataset visualization is based on variation in the density of the images.

## IV. METHODOLOGY

The CT Scan of the brain hemorrhage as an input image data needs to be preprocessed first for further deep learning processing. First, resized the images into a fixed size because images are in a different size and deep learning model does not get input of different sizes. That's why the suitable size $128 \times 128$ is selected.

After applying the train-test split, Image augmentation is applied to increase the number of training images from 180 to 1000 images to boost the performance. The CNN, CNN + LSTM and CNN + GRU are applied to the preprocessed images. The layered architecture of deep learning models are consists of three convolutional 2D, two max-pooling, one global average pooling and fully connected dense layers as presented in Figure 5. The layer are all connected that extracts the features from the image data and pass it to the next layer. In the end, the dense layer drop out the links of the neuron to identify the brain hemorrhage in the CT scan images of the brain. The proposed methodology is presented in Figure 4.

### A. PREPROCESSING

The learning phase of the CNN model is the major part of the proposed study because in the medical field, the diagnosis of the brain hemorrhage is the most vital part. The CT scan image dataset needs more concentration based on some filtration and enhancements to enhance the learning efficiency of the training phase of the CNN model. The preprocessing is the phase of filtering, reshaping and enhance the dataset quality to increases the performance of the deep learning models. There are different methods used in this study like resizing, flipping and augmentation of images are applied as a preprocessing to enhance the quality and quantity of the images data.

### 1) RESIZE THE IMAGES DATA

The training of the deep learning models needs to train on the same size of the images. The resizing of the images into a same sizes speed up the learning process and reduces the chance of overfitting. The performance of the model and the accuracy rate also reduces due to the loss of data during image resizing that is one of the challenging parts to resize the images data. In the proposed study $128 \times 128$ dimensions
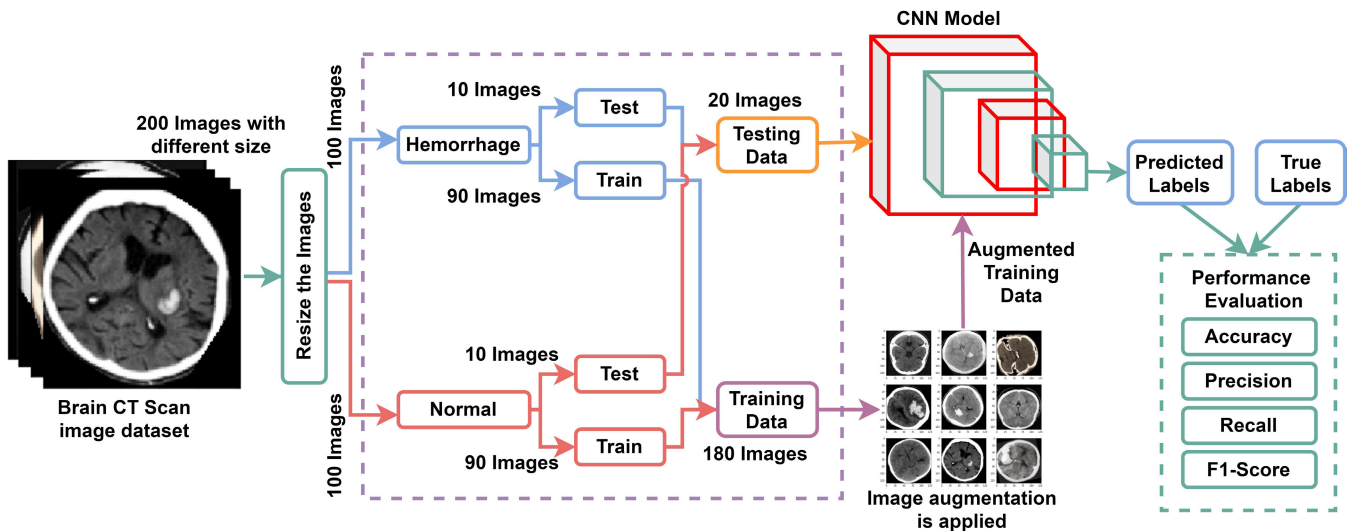
**FIGURE 4.** The proposed methodology of the BHCNet system.

are selected to resize the images in a fixed size. It is efficient to overcome both overfitting and fast learning rate issues with the best accuracy rate.

### 2) TRAINING AND TESTING DATA SPLIT

The train-test split is the process in which the data split into a fixed ratio for the train and the testing of the deep learning models. The brain hemorrhage is the medical emergency in which the most significant part is to diagnose correctly. For this purpose, the deep learning models need to train with maximum data and then perform accurate predictions. The 90% data selected for the training phase and 10% data for the testing phase. There are 200 images in which 100 images of Hemorrhage and 100 are of non-brain hemorrhage patients. The 180 images used for the training, from which 90 random images selected from Hemorrhage and 90 random images from non-brain hemorrhage. The remaining 20 images used for testing purpose from which 10 from Hemorrhage and the 10 images from non-brain hemorrhage as shown in Figure 4. Then Image augmentation used on the training data of 180 images to enhance the learning of the deep learning models.

### 3) IMAGE AUGMENTATION

The image augmentation [61] is the process of artificially create the data for the efficient learning and to increase the prediction accuracy. In the concepts of deep learning a small dataset is the major obstacle in the process of learning [62]. That is why image augmentation is used to artificially increase the training and validation dataset by flipping horizontally or vertically, rescaling, shearing, by increasing or decreasing the zoom range, by rotating image at different angles, by increasing or decreasing the width or height ranges and by using fill mode. The image dataset has different pixel values and by rescaling the pixel value of all the images

transform into the range of [0, 255] to [0, 1] to treat all the images in equal manners. By zooming at the range of 0.05, shearing counter-clockwise direction, by shifting range of image height and width at the range of 0.05 with the filling mode at constant, the image augmentation process enhances the small dataset for the training process.

### B. PROPOSED DEEP LEARNING MODELS

The image classification took years of efforts and experienced persons to construct algorithms [63], [64]. Deep learning reduces the efforts of years into hours or minutes that consists of neural networks. In the field of neurology, image classification is used on a large scale. Image input consists of pixel values in the form of numerics that assigned to neuron. Each neuron contains a single numeric value and connection between neuron contain weights that represent the strength between the neurons of different layers. In this study, deep learning CNN, hybrid of CNN + LSTM and LSTM + GRU models are proposed to diagnose the brain hemorrhage.

### 1) CONVOLUTIONAL NEURAL NETWORK (CNN) MODEL

Deep learning method CNN [49], [50] is consists of a network of layered architecture. CNN is mostly used for image classification purposes. The raw pixels are extracted by the layered architecture of CNN from the image in the form of features. These layers are the input layer, 2D convolutional layer, max-pooling layer, global average pooling layer and dense layer are used in this study as shown in Figure 5. The features are extracted by each layer and pass those to the next layer. These features are classified within the model.

The core of the CNN is convolutional layer that perform the most complex computational tasks than the other layers. The convolutional layer depend on the learnable filters. This layer convolves over the input image based on the receptive field size that is equivalent to the filter size known as the
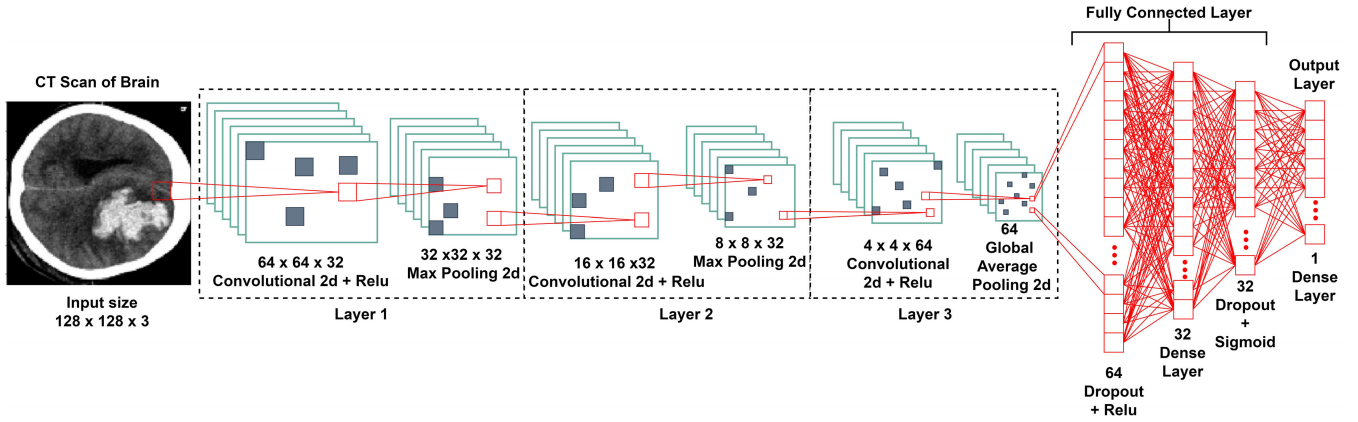
**FIGURE 5.** BHCNet using the deep layered architecture of the CNN model.

kernel [65], [66]. The kernel size is spatially small like the image size is $128 \times 128 \times 3$ in pixels (width $\times$ height $\times$ dimensions) then the kernel size is $3 \times 3$. The dimensions of the kernel are $3 \times 3 \times 3$ where 3 is the depth of the original image. This kernel first starts to convolve over the image from the left top corner then move forward unit by unit. The element-wise multiplication applies between the kernel numeric values and intensity values of the image. All the multiplied values are summed up and give a single value. After this, the kernel moves 2 unit forward on the image and repeat this process again and again until it reaches to the last unit of the image. Two main parameters padding and strides are responsible for improving the behavior of the CNN model. Further presentation of layers of CNN model in Figure 5. The mathematical computations are shown in equations 1 and 2.

$$\mu_{ij} = \sum_{p=0}^{H-1} \sum_{q=0}^{H-1} x_{si+p,sj+q} h_{pq} \qquad (1)$$

where $h$ presents the pixel value, $h-1$ shows the padding, $s$ presents the strides and $h_{pq}$ is weight sharing.

$$\mu_{ijm} = = \sum_{k=0}^{K-1} \sum_{p=0}^{H-1} \sum_{q=0}^{H-1} z_{i+p,j+q,k}^{l-1} h_{pqkm} + b_{ijm} \qquad (2)$$

where $\mu_{ijm}$ presents the output feature map, $b_{ijm}$ presents the bias and $h_{pqkm}$ presents weight sharing also known as weight tying.

*a: CNN LAYERED ARCHITECTURE*

The preprocessing phase enhances the CT scan image dataset for the CNN model to classify brain hemorrhage in efficient manners. After preprocessing CNN model applied on the dataset for the training and learning process. First of all convolutional 2D layer extract the features from the input image by convolving the $3 \times 3$ kernel on the image. By performing matrix multiplication operations, the kernel values multiplied by the pixel values of the image then summed up the values

**TABLE 1.** Summery of the model and hyperparameter.

| Hyperparameter | Values |
|---|---|
| Batch size | 32 |
| Pooling | 2 x 2 |
| Dropout rate | 0.4 |
| Optimizer | RMSprop |
| Function | Binary cross entropy |

and moving 2 strides forward then again summed up the value with the previous one and use this single summed up value as a feature map.

The convolutional 2D layer depends on the activation function ReLU [50], Strides and Dropout rate. The strides control the movement of the kernel that is equal to 2. The activation function ReLU dealing with the gradient descent by performs the thresholding process on the matrix multiplied the summed value of the convolutional layer. It converts the values into zero if it is lesser then zero. The strides and ReLU boost the computation and learning speed of the model but as well increase the chances of overfitting. There is dropout used to reduce the chances of the overfitting and improve the accuracy rate according to the accurate predictions. Then by using these parameters, the convolutional 2D layer produces the output feature map of $64 \times 64 \times 32$. The feature map that is generated by the convolutional 2D layer passes to the next Max Pooling layer that used for the down-sampling by convolving its own kernel around the feature map and extract the maximum value. It uses the dimensionality reduction technique to reduce the spatial size to increase computational power and also dealing with the overfitting problem. It down-samples the feature map of the $64 \times 64 \times 32$ to the $32 \times 32 \times 32$ by using the pool size $= 2$. It means the pooling window is $2 \times 2$ to convolving around the feature map and select the maximum value from it.

There are two sets of layers implemented consists of convolutional 2D and Max pooling 2D layer [53], [67] as shown in Figure 5 which extracts the features and down-samples the features map to $4 \times 4 \times 64$. Then the Global average pooling
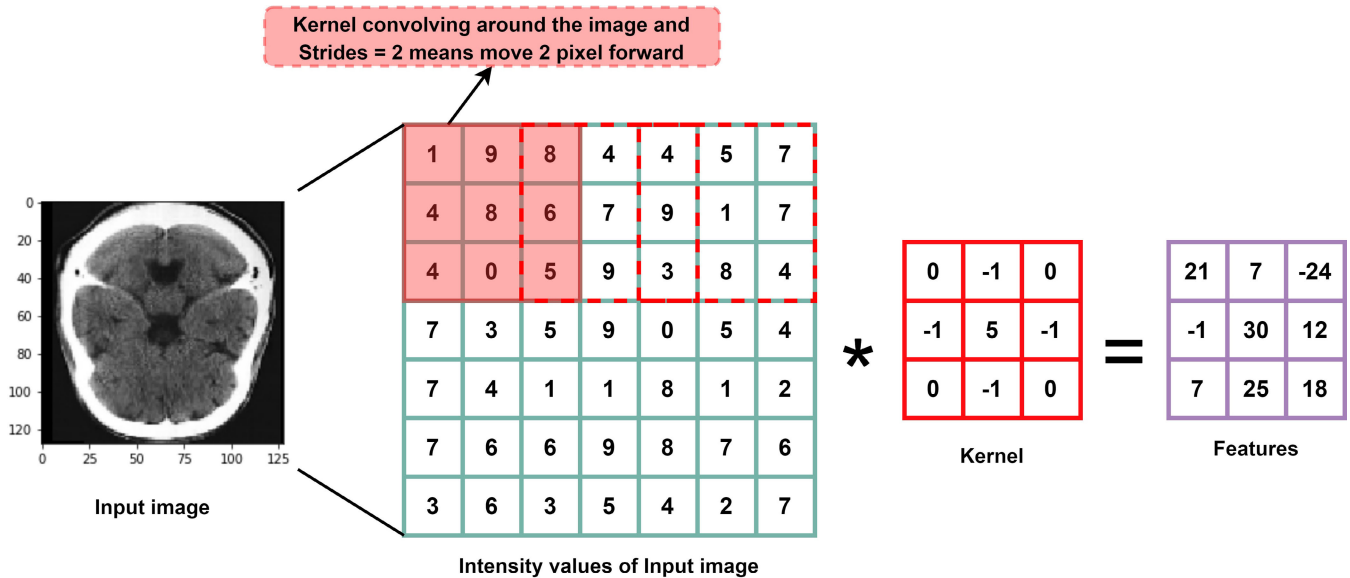
**FIGURE 6.** The convolutional layer convolving around the input image and extracting feature map as a output by using strides and padding.

layer [68] comes next. It also a dimensionality reduction technique that generates the one feature vector by extracting one feature from each feature map corresponding to the classification category. Then this feature vector directly passes to the dense layer. The dense layer by using dropout the link of the neurons performs the classification task efficiently.

### b: PADDING

The convolutional layers kernel is convolving over the input image in every channel that will absolutely reduce the dimension and spatial size. This cause the loss of information which increases the computation speed of the model but reduces the accuracy of the results. For preserving the information and to achieve higher accuracy results in the classification process, it is necessary to get the feature map after convolving the input image in its same size without any information loss. By padding zero around the matrix of the input image intensity values will save the information [69]. This padding of zero helps the kernel to convolve around image and produce feature of exact spatial size of original input matrix.

### c: STRIDES

Strides are handling the movement of the kernel convolving over input image. If stride is equal to 1 then it means kernel move one pixel at a time. If it is 2 then it means move 2 pixels forward at a time over the image as shown in Figure 6. The kernel is in the red shaded box that was floating over intensity values of the input image and red dotted lines shows the movement of the kernel according to the strides.

### d: ACTIVATION LAYER

Neural Network used activation functions [50], [70] to handle the given data through gradient processing by using gradient descent in which output produced for neural networks from data that contain parameters. The activation functions in the neural network used for the computation of the weighted sum biases and inputs. neuron has to be fired or not, decided based on this computation. The major purpose of the activation function is to convert input linear signals into non-linear output signals that are easily differentiable. Otherwise, during the process of backpropagation of neural networks, linear functions cannot work. The transformation of input vectors x is shown in equation 3

$$f(x) = w^T x + b \tag{3}$$

where w = weights, x = input, b = biases.

$$y = (w_1 x_1 + w_2 x_2 + \ldots + w_n x_n + b) \tag{4}$$
$$y = \alpha(w_1 x_1 + w_2 x_2 + \ldots + w_n x_n + b) \tag{5}$$

where $\alpha$ is the activation function.

The non-linear output is shown in equation 5 after apply the activation function.

*Rectified Linear Unit(ReLU):* The ReLU is the activation function in neural network models that is conventionally used for the hidden layers of the model. It is a non-linear function that was nearly presented as a linear function. The properties of linear function made this activation function simple to optimize among gradient descent. The ReLU activation made models to learn faster and gives better performance by overcomes the problems of vanishing gradient [70], [71]. The ReLU classification layer learning on the weight parameters through backpropagation that set each element value to zero if it is less than zero by performing the threshold process. The major advantage of ReLU activation function is to boost the computational power of the deep learning models but increase the risks of overfitting. The dropout technique is used with

ReLU activation function to reduce the effects of overfitting that improves the performance of the CNN model.

$$f(x) = max(0, x) = \begin{cases} x_i, & if \ x_i \geq 0 \\ 0, & if \ x_i < 0 \end{cases} \quad (6)$$

### e: MAX POOLING LAYER

The most common feature of CNN is pooling. The main aim of pooling is to accumulate the features from the maps generated by convolving the kernel over the input image. The discretization process based on sampling known as Max pooling [53], [67] and used as a downsampling technique in CNN. The max pooling downsample the input image by applying a dimensionality reduction process. It reduces the spatial size of the representation to reduce the computational cost and dealing with overfitting process by reducing the number of parameters and presenting an abstract form of representation. Max pooling kernel convolve over the subregions of the maps and gives the maximum value as an output. The process of max pooling is shown in Figure 7. There are four subregions, as an output maximum number is selected from every subregion and reduces the dimensionality.

$$\mu_{ijm} = max_{p,q \in P_{i,j}} z_{pqk} \quad (7)$$

where $z_{pqk}$ presents the pooling size and *max* function gets the maximum value by convolving the pool around the feature map.
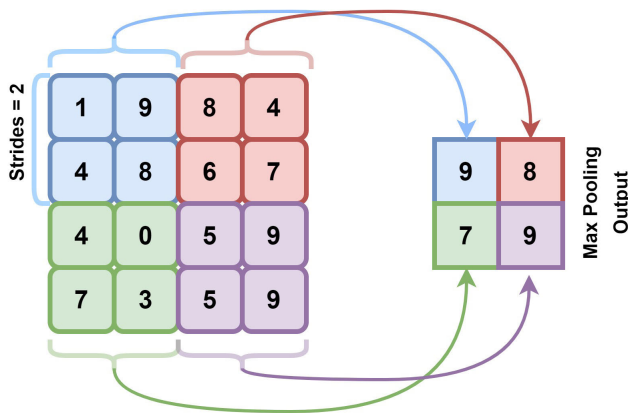


**FIGURE 7.** Pool of the max pooling layer with stride = 2 convolving around the feature map and perform down sampling operation by extracting maximum value.

### f: GLOBAL AVERAGE POOLING 2D LAYER

The global average pooling layer is the dimensionality reduction method that decreases the overfitting possibilities by reducing the representation of parameters and computation in the model [68]. The main purpose of this layer is to create one feature map for each corresponding category of the classification task than get the aggregate of each feature map in the form of resulting vector that directly fed into the

dense layer.

$$h_j^n(p, q) = \frac{1}{H} \sum_{p' \in N(p), q' \in N(q)} h_j^{n-1}(p', q') \quad (8)$$

where $h_j^n(x, y)$ is the output vector, $\frac{1}{H}$ is presents the averaging function, $H$ is the pixel value and the summation $\sum$ of the pooling convolving around the feature map.

### g: DENSE LAYER

The architecture of the dense layer is also known as a fully connected layer because neurons are fully connected in this layer to the activation of the previous layers [41]. The convolutional and max-pooling layers produce numerous features that are used by the dense layer for the classification of the input from the various class. The learning and classification process enhanced when the combination of features gives better results. The equation 9 showing the matrix multiplication:

$$Y = X.W + b \quad (9)$$

where Y is the prediction that was perform with the dot product of X output of previous layer with W the weight and biases b.

### h: DROPOUT LAYER

Dropout is generally a regularization function that prevents the process of overfitting that occur in the process of training of the model. When multiple neurons of the deep learning model detect the same feature is called co-adaption [50]. To reduce the effects of co-adaption, it drops the nodes randomly to disconnect the connections at the fully connected layer by setting the corresponding activation function to 0 value. The dropout rate is set to 0.4 in this research with ReLU activation function to reduce the overfitting effects and to enhance the performance of CNN model.

$$Z_i^L = \omega_i^L y^L + b_i^L \quad (10)$$

where $L \in l^1, l^2, l^3, \ldots, l^n$ are the hidden layers, $z^L$ presents input layer and $y^L$ presents the output vector [72].

### 2) LONG SHORT TERM MEMORY (LSTM) MODEL

The LSTM is combined with the CNN because it is capable of learning long term dependencies based on 50 units. The units are composed of memory cells that are used to keep the knowledge of previous states for an arbitrary time interval. When CNN combined with LSTM, it provides the facility to store the previous states of the CNN model layers and improve the learning of the model by using stored weights. The three gates input gate $i_t$, an output gate $o_t$ and a forget gate $f_t$ are used to regulates the flow of cell $c_k$ information based on current input $x_t$, current hidden state $h_t$ and old hidden state $h_{t-1}$ at each time step.

$$i_t = \sigma\left(W_i.[h_{t-1}, x_t] + b_i\right) \quad (11)$$
$$f_t = \sigma\left(W_f.[h_{t-1}, x_t] + b_f\right) \quad (12)$$
$$q_t = \tanh\left(W_q.[h_{t-1}, x_t] + b_q\right) \quad (13)$$

**TABLE 2.** Designed layered architecture for neural network models.

| Sr. no. | CNN Layers | CNN+LSTM Layers | CNN+GRU |
|---|---|---|---|
| 1 | Convolution 2D (64 x 64 x 32), ReLU, Kernel(3 x 3), Strides(2 x 2) | Convolution 2D (64 x 64 x 32), ReLU, Kernel(3 x 3), Strides(2 x 2) | Convolution 2D (64 x 64 x 32), ReLU, Kernel(3 x 3), Strides(2 x 2) |
| 2 | Max Pooling 2D (32 x 32 x 32), Pooling size(2 X 2) | Max Pooling 2D (32 x 32 x 32), Pooling size(2 X 2) | Max Pooling 2D (32 x 32 x 32), Pooling size(2 X 2) |
| 3 | Convolution 2D (16 x 16 x 32), ReLU, Kernel(3 x 3), Strides(2 x 2) | Convolution 2D (16 x 16 x 32), ReLU, Kernel(3 x 3), Strides(2 x 2) | Convolution 2D (16 x 16 x 32), ReLU, Kernel(3 x 3), Strides(2 x 2) |
| 4 | Max Pooling 2D (8 x 8 x 32), Pooling size(2 X 2) | Max Pooling 2D (8 x 8 x 32), Pooling size(2 X 2) | Max Pooling 2D (8 x 8 x 32), Pooling size(2 X 2) |
| 5 | Convolution 2D (4 x 4 x 64), ReLU, Kernel(3 x 3), Strides(2 x 2) | Convolution 2D (4 x 4 x 64), ReLU, Kernel(3 x 3), Strides(2 x 2) | Convolution 2D (4 x 4 x 64), ReLU, Kernel(3 x 3), Strides(2 x 2) |
| 6 | Global Average Pooling 2D (64) | Global Average Pooling 2D (64) | Global Average Pooling 2D (64) |
| 7 | Dropout (64), Dropout rate (0.4) | Dropout (64), Dropout rate (0.4) | Dropout (64), Dropout rate (0.4) |
| 9 | Dense (32), ReLU | LSTM (64 x 50), ReLU | GRU (64 x 50), ReLU |
| 8 | Dropout (32), Dropout rate (0.4) | Dropout (64 x 50), Dropout rate (0.4) | Dropout (64 x 50), Dropout rate (0.4) |
| 9 | Dense (1) , Sigmoid | Flatten (3200) | Flatten (3200) |
| 10 | | Dense (32), ReLU | Dense (32), ReLU |
| 11 | | Dropout (32), Dropout rate (0.4) | Dropout (32), Dropout rate (0.4) |
| 12 | | Dense (1) , Sigmoid | Dense (1) , Sigmoid |

$$o_t = \sigma \left( W_o . \left[ h_{t-1}, x_t \right] + b_o \right) \quad (14)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot q_t \quad (15)$$

$$h_t = o_t \odot \tanh c_t \quad (16)$$

where $\odot$ represents the element wise multiplication. $\sigma$ is the sigmoid function that gives output in [0, 1].

### 3) GATED RECURRENT UNIT (GRU) MODEL

The GRU model contain the feature of RNN and LSTM because it is dependent on the gating mechanism of the RNN model and perform like LSTM. The GRU is consists of update and reset gate where LSTM does not contain these gates. The LSTM is the sequential model which show gradient issue and working better with the large datasets where GRU is introduced to solve these issues and it performs better than the LSTM with small dataset.

$$z_t = \sigma_g \left( W_z x_t + U_z h_{t-1} + b_z \right) \quad (17)$$

$$r_t = \sigma_g \left( W_r x_t + U_r h_{t-1} + b_r \right) \quad (18)$$

$$h'_t = \phi_h \left( W_h x_t + U_h \left( r_t \odot h_{t-1} \right) + b_h \right) \quad (19)$$

$$h_t = \left( 1 - z_t \right) \odot h_{t-1} + z_t \odot h'_t \quad (20)$$

where $x_t$ is input vector, $h_t$ is output vector, $h'_t$ is the activation vector, $z_t$ is update gate vector, $r_t$ is reset gate vector, $W$, $U$ and $b$ is parameter matrices and vector.
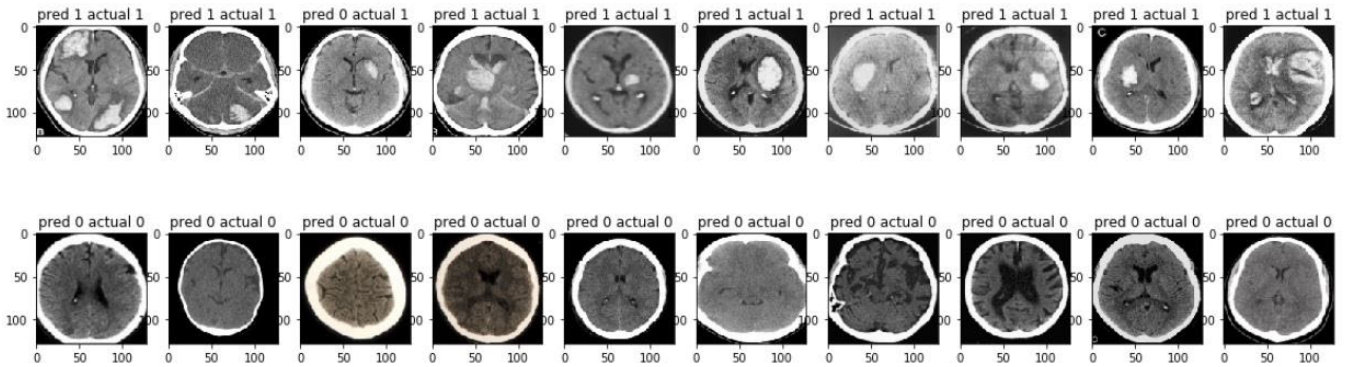
## V. RESULTS AND DISCUSSION

The main objectives are to enhance the training of deep learning models, prediction speed and performance of the classification process. It detects the patient has the brain hemorrhage or not by using the 200 CT scan images dataset.

The image augmentation methods are used to increase the dataset from 180 training images to 1000 images, deep learning model CNN, hybrid models CNN + LSTM and CNN + GRU are used for identification of brain hemorrhage and proposed approach BHCnet overcome all challenges.

The experiments are carried out by using Dell PowerEdge T430 graphical processing unit with installed 8 cores, 16 logical processors, DDR4 32 GB Random Access Memory (RAM). The training of the CNN with the 1000 augmented head CT scan dataset took 21 minutes to run the epochs where hybrid models CNN + LSTM and CNN + GRU took average 22.4 minutes. The experimental results illustrate the performance of the classification of the CNN model using balanced dataset in Table 3. There are different evaluation matrices adopted like accuracy, precision, sensitivity, specificity and F1-score. After the implementation, the analyses show the results in the form of TP, TN, FP and FN that shows the true and false positive and negative predictions of the model. In Table 3, there are 6 different experiments performed based on epochs. The highest result is obtained with 24 epochs and got 95% accuracy rate, 90.90% precision, 100% sensitivity, 90% specificity, 95.23% F1-score.

The true positive (TP) rate is 10. It means 10 correct predictions that CNN model predicts that the 10 patients has brain hemorrhage and in actual those patients also have a brain hemorrhage. There true negative (TN) rate is 9, means CNN model predicts 9 non-hemorrhage patients and in actual those are also non-brain hemorrhage patients out of 20 test cases. Now, false positive (FP) is 0, means CNN model predicts the 0 patient has brain hemorrhage and in actual is also 0.

In Last false negative (FN) case is the major aim of this study, there FN is 1. It means the 1 patient has the brain hemorrhage or internal bleeding in the head but the CNN model does not identify it and predicts the patient is non-brain hemorrhage. It would be a cause of death because doctor give different treatment to patient according to the false prediction. In this case, there are FN cases are 1 it means due to wrong prediction 1 patients are in severe danger. The 95% accuracy of the CNN model is not enough to save every single patient. That is why this study proposed the imbalanced concept to the dataset.

**FIGURE 8.** Comparative results of the hemorrhage and non-brain hemorrhage CT scan image classification using balanced dataset with proposed CNN model. The same results of prediction (pred) and actual shows accurate prediction and difference show false predictions.

**TABLE 3.** Experimental results of CNN model with balanced dataset.

| Ep. | TP | TN | FP | FN | Accu. | Prec. | Sensi. | Speci. | F1-Score |
|-----|----|----|----|----|-------|-------|--------|--------|----------|
| 04 | 7 | 9 | 1 | 3 | 80.0 | 87.5 | 70.0 | 90.0 | 77.77 |
| 08 | 8 | 8 | 2 | 2 | 80.0 | 80.0 | 80.0 | 80.0 | 80.0 |
| 12 | 1 | 10 | 0 | 9 | 55.0 | 100 | 10 | 100 | 18.18 |
| 16 | 10 | 8 | 2 | 0 | 90.0 | 83.33 | 100 | 80.0 | 90.90 |
| 20 | 10 | 8 | 2 | 0 | 90.0 | 83.33 | 100 | 80.0 | 90.90 |
| 24 | 10 | 9 | 0 | 1 | **95** | **90.90** | **100** | **90.0** | **95.23** |

where Ep. is number of Epochs, Accu. is Accuracy, Prec. is precision and Sensi. is Sensitivity Speci. is Specificity

The comparative analyses of the results are visualized in Figure 8 that shows the average results with the 24 epochs are best. There the improvement of the CNN model shows step by step with each epoch. The classification of the brain hemorrhage and non-brain hemorrhage patients in actual and in predict results shown in Figure 8. There are 20 testing CT scan images of the brain are labelled with Pred (Predicted) and Actual where (1) presents the brain hemorrhage patient and (0) presents the non-brain hemorrhage Patient. The similar results of the predicted and the actual shows that the results are accurate. The internal bleeding is also shows some gery-white spreading in the CT scan in predicted and actual (1) case. In the case of (0) the CT scan images have no sign of internal bleeding that shows the patient is non-brain hemorrhage. In the 3rd result of the first row, there is actual and predicted results are different that shows the patient has brain hemorrhage but CNN predicts the patient is non-brain hemorrhage. Here, the results show that in the medical field 95% accuracy means the death of one patient.

The medical field needs more concentration on the diagnoses because one false diagnose could be the death of the patient. Therefore in the proposed study, the case in which a patient has brain hemorrhage in actual but CNN model predicts it non-brain hemorrhage that needs more concentration. The proposed CNN model with the layered architecture and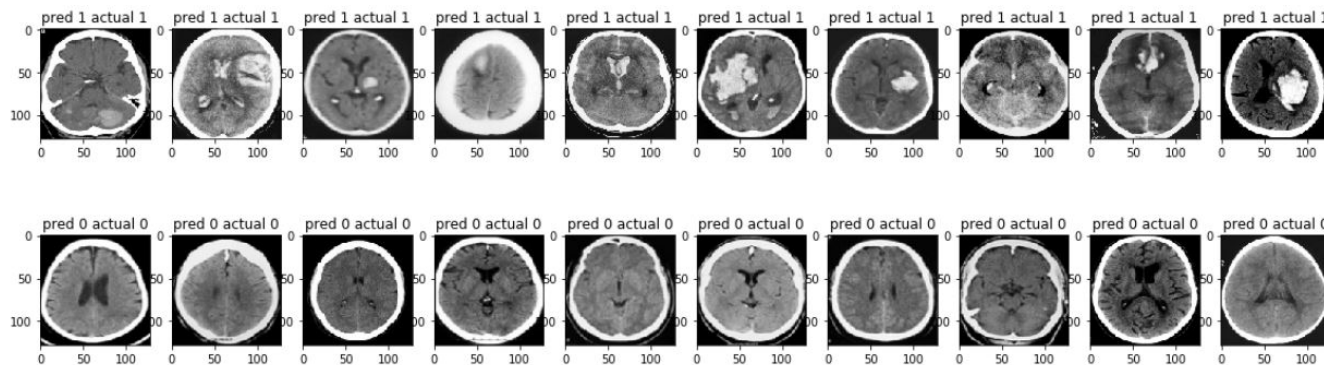 image augmentation achieves 95% accuracy but increase the number of sample images by using. In this study, two methods are adopted to overcome this issue. First, imbalancing the dataset that increase the number of positive cases in the dataset that concentrate the training of the CNN model on FN cases more. Second, save the class weights from the previous training of the model for the next iteration.

The imbalancing of the dataset get the concentration of the CNN model towards false predictions that improve the prediction and accuracy ratE. After training and test data splitting, the training data consists of 180 CT scan images and after imbalance the training data, it consists of 243 images.
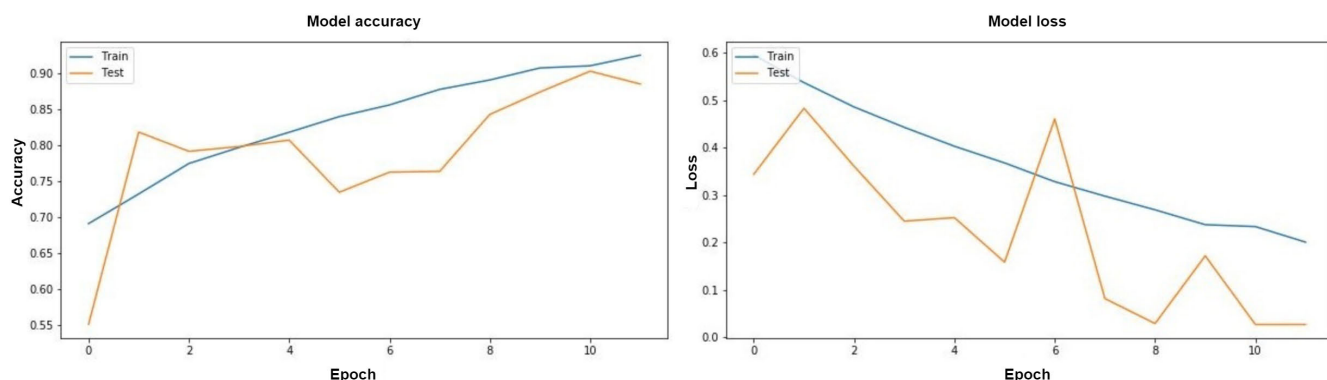
Further, the experimental analyses are shown in Table 4 that presents the highest results obtained with 12 epochs i.e. 100% accuracy, 95.54% precision, 100% sensitivity, 95.5% specificity and 95% F1-score. These results based on TP is 10, TN is 10, FP is 0 and FN is 0. The highest accuracy of 100% achieved by killing all false predictions and improve the learning of the CNN model.

In the experiments by imbalancing the dataset and saving the class weight for further processing. Table 2 presents the 10 layers of the CNN model that is implemented in the first experimental phase. In the second experimental phase, the same sequence of layers with the same hyper-parameters are added to the CNN model again. The previous class weights of the training phase based on epochs is saved that are uploaded again to increase the training efficiency to improve the training and accuracy rate. The imbalance the dataset provides the best support in the training phase of the CNN model. The main focus of this experimental phase is to eliminate the false results like FP and FN because the life of the patient is at risk.

The resultant Figure 9 shows the classification results of the CT scan images where each image is labelled with actual and predicted results. There is no dissimilar result, it means the predictions are matched with actual labels that show the proposed study are 100% accurately predicts in between the brain hemorrhage and non-brain hemorrhage patients. It successfully eliminates all false prediction and gives highest results as shown in Table 5.

**FIGURE 9.** Comparative results of the hemorrhage and non-brain hemorrhage CT scan image classification using proposed BHCNet with imbalanced dataset and proposed layered architecture of CNN model. Where the same results of the prediction (pred) and actual values shows accurate predictions and difference shows false predictions.



**FIGURE 10.** The validation of the training phase of the CNN model in terms of accuracy and loss.

**TABLE 4.** Proposed BHCNet experimental results using CNN model with imbalanced dataset.

| Ep. | TP | TN | FP | FN | Accu. | Prec. | Sensi. | Speci. | F1-Score |
|-----|----|----|----|----|-------|-------|--------|--------|----------|
| 04 | 10 | 7 | 3 | 0 | 85.0 | 76.92 | 100 | 70.0 | 85.0 |
| 08 | 10 | 9 | 1 | 0 | 95.0 | 90.90 | 100 | 90.0 | 95.23 |
| 12 | 10 | 10 | 0 | 0 | **100** | **95.54** | **100** | **100** | **95.0** |
| 16 | 9 | 10 | 0 | 1 | 95.0 | 100 | 90.0 | 100 | 94.73 |
| 20 | 9 | 6 | 4 | 1 | 75.0 | 69.23 | 90.0 | 60.0 | 78.28 |
| 24 | 10 | 9 | 1 | 0 | 95.5 | 90.90 | 100 | 90.0 | 95.23 |

where Ep. is number of Epochs, Accu. is Accuracy, Prec. is precision and Sensi. is Sensitivity Speci. is Specificity

The training processes of the CNN model is evaluated by the model loss and accuracy validation. The epochs are the number of iterations used during training to extract features and pass it to next layers for learning. If the model accuracy decreases and loss increase during the training phase that means the model is not learning and shows the overfitting. When the accuracy increases and loss decrease means the model is learning. Figure 10 presents the accuracy and loss of the proposed CNN model in terms of the train and test data. The highest training accuracy is achieved with the minimum data loss during the training process. This means the speed and computational power of the proposed model is enhanced just in 12 epochs by using image augmentation methods and minimum layers of the CNN model.

Further, the results are discussed to illustrate the effectiveness of the proposed study. The experimental results of the CNN and hybrid CNN + LSTM model with the balanced and imbalanced dataset is discussed comparatively.

## A. COMPARATIVE ANALYSES OF RESULTS OF CNN AND HYBRID CNN + LSTM MODEL

The CNN model is merged with LSTM and GRU models because it consists of memory cells and gated mechanism. The learning efficiency of the CNN model is enhanced with LSTM because the LSTM model can handle the sequential data and CNN does not have memory cell structure. The LSTM handle the flow of information in memory cell using input output and forget gate. The GRU model contain fewer parameter than the LSTM which makes it faster than LSTM. The GRU contains update and reset gates but LSTM out performs the GRU after combining it with CNN model. The CNN + LSTM shows 95% highest accuracy where CNN + GRU shows 90% accuracy. Table 5 presents the comparison results of the proposed BHCNet, CNN, hybrid CNN + LSTM and CNN + GRU models that clearly illustrates that the highest results are achieved with BHCNet in which CNN model
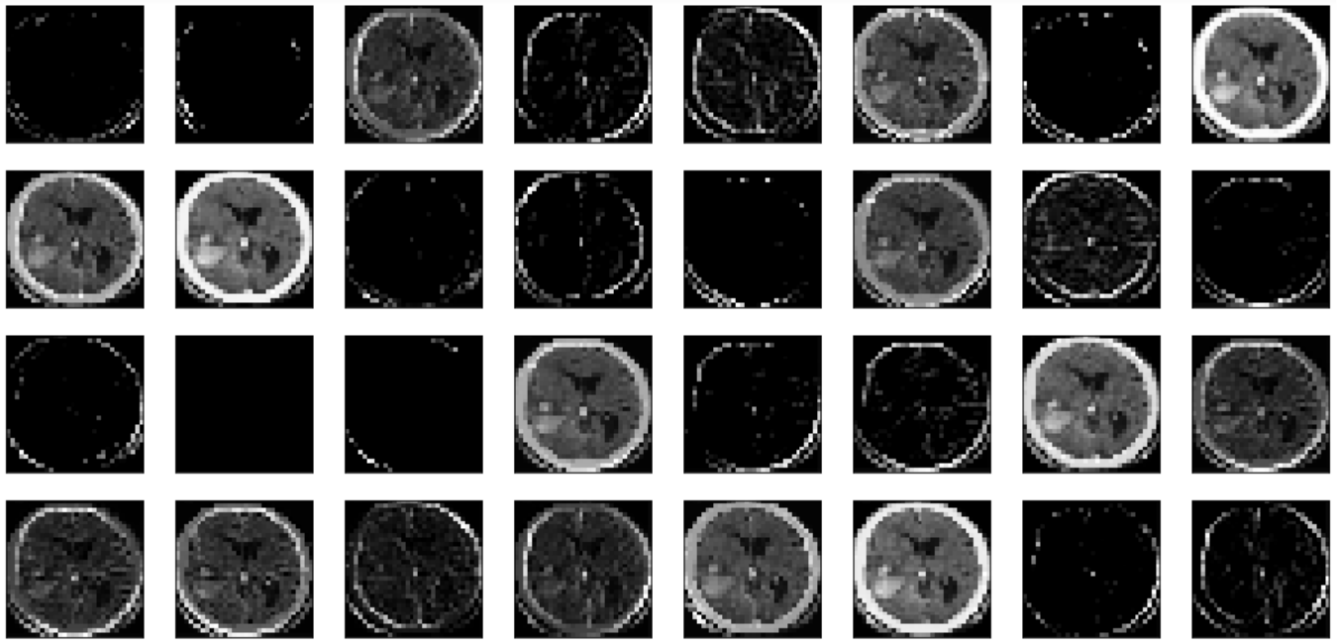
**TABLE 5.** Comparative representation of results of proposed BHCNet, CNN, hybrid model CNN + LSTM and CNN + GRU models with balanced and imbalanced dataset.

| Deep Learning Models | Accuracy | Precision | Sensitivity | Specificity | F1-Score |
|---|---|---|---|---|---|
| Proposed CNN with balanced dataset | 95 | 90.90 | 100 | 90 | 95.23 |
| **BHCNet: Proposed CNN with imbalanced dataset** | **100** | **95.54** | **100** | **100** | **95.0** |
| CNN+LSTM with balanced dataset | 90 | 90 | 90 | 90 | 90 |
| CNN+LSTM with imbalanced dataset | 95 | 90 | 100 | 90 | 95.23 |
| CNN+GRU with balanced dataset | 90 | 83.33 | 100 | 80 | 90.90 |
| CNN+GRU with imbalanced dataset | 90 | 85 | 100 | 90 | 95.23 |

**TABLE 6.** Analyses of previously used approaches with proposed BHCNet approach.

| Literature | year | Dataset | Technique | Accuracy | Specificity | Precision | Sensitivity | F1-Score |
|---|---|---|---|---|---|---|---|---|
| [73] | 2019 | Large non contrast CT scan dataset | AlexNet-SVM | 93% | 90% | - | 95% | - |
| [74] | 2019 | Large Head CT scan dataset | hybrid 3D/2D mask ROI-based CNN architecture | 97.5% | 97.5% | 79.3% | 97.1% | - |
| [75] | 2019 | Large dataset of non contrast CT scan | Aidoc (Tel Aviv, Israel) neural network software | Overall 93% in emergency cases 96% | 94.2% | 73.7% | 88.7% | - |
| [76] | 2019 | Small dataset of Head CT imaging | CNN | 91% | - | 90.89% | 93% | 91.45% |
| [77] | 2020 | Large dataset of CQ500 head CT dataset | 3D CNN | 97% | 94% | 97% | 98% | 94% |
| **Proposed BHCNet** | **2021** | **Head CT scan Small dataset** | **CNN** | **100%** | **100%** | **95.54%** | **100%** | **95%** |

using data imbalancing technique. The BHCNet approach outperforms the hybrid CNN + LSTM and CNN + GRU models by gaining 100% accuracy, 95.54% precision, 100% sensitivity, 100% specificity and 95.0% F1-score.

Further, Table 6 shows the comparison of the proposed approach with previous studies. The comparative analyses display two major objectives that are successfully achieved. **1:** The major aim is to propose an approach that will

be used in emergency cases in which large datasets are not available in the initial stages and requires training the models with a small dataset. Majority of the authors and researchers suggests to use large datasets, that's why small dataset is utilized to perform experiments. Table 6 presents the different and latest approaches based on deep learning. Where [62]–[64] and [66] used large datasets. In a comparison of the large dataset and complex methodology, we are proposing a less complex methodology with a small dataset that also requires less time to train models. 2: The second aim is to save every single life that is suffering from inadequate diagnoses and the carelessness of doctors. This requires improving prediction results. Table 6 also displays the results and evaluations in comprehensive manners that illustrates the effectiveness of the proposed approach over other approaches.

The main aim of the researchers is to use neural networks in the best way in practice with its most simple and fundamental elements. The proposed BHCNet also depends on these fundamental elements that show its importance. Convolutional Neural Network (CNN) is the most fundamental approach of its advanced application like LeNet-5, AlexNet, VGG, GoogleNet, and Residual Network or ResNet that have its specification, benefits and limitations. The first application of the CNN is LeNet-5 [78] used to recognize the handwritten characters and shows 99.2% accuracy. The key point of the LeNet-5 is, it gets fixed-size input of $5 \times 5$ pixels by using 16 filters and with the depth of the network, the filters are also increased. AlexNet was proposed by Krizhevsky *et al.* [79] which is created to ReLU activation function with every convolutional layer instead of other common activation functions. In AlexNet, half of the LeNet-5 layers are replaced by the max-pooling layers which are used to reduce the number of features by reducing the filter size and to address overfitting strides are introduced same as the size of pooling. In the last, the fully connected layers depend on the dropout rate. Needs a large data set for the training purpose and consume time for complex computations. VGG is proposed by group Visual Geometry Group (VGG) [80] at Oxford in 2014. VGG has consisted of a large network of layers of about 16 and 19 layers with dramatic repetition of the block of large convolutional layers set and max-pooling layers. But it depends on the small size of filters with the stride one. Due to its depth and heavy pre-trained model, the VGG used as a starting point in transfer learning. GoogleNet was proposed by Szegedy *et al.* [81], consists of the parallel convolutional layered architecture of different size filters of $1 \times 1$, $3 \times 3$ and $5 \times 5$ and $3 \times 3$ max-pooling layer. The output of these parallel layers is concatenated that is known as the inception module. The problem with the inception model is the number of filters starts building up rapidly when the inception module are stacked. This produces a large network of about 22 layers that contains a further set of a convolutional and pooling layer need a large dataset for the training. Further used global average pooling layer for the output of the model. ResNet introduced by He *et al.* [82].

The major contribution is the shortcut connections with the deep 152 layers architecture, in which the input kept in its original form instead of weighted and pass on to the deeper layer after skipping next. ResNet also consists deeper network with heavy computational tasks and large datasets. Where the major contribution of this study is to propose a neural network that efficiently used for smaller datasets with a small layered network for a fast and accurate prediction rate to save the patient's life even in an emergency case. Second the layered architecture of the CNN model designed to predict the brain hemorrhage known as BHCNet is specifically trained and tested with the CT scan images of the brain. The third BHCNet create with the minimum number of layers, small filters and lesser hyperparameter with efficient performance. BHCNet depends on 3 major blocks of the layers that are consist of 1 convolutional and 1 max-pooling layer in which 3rd block consists of the convolutional and global average pooling layer. The convolutional neural network extracts the useful feature with the filter size of $3 \times 3$ and max-pooling reduce the features for the net layers with $2 \times 2$ pooling size. Convolutional layers are implemented with ReLU activation function instead of simple tan or linear activation functions with $3 \times 3$ filter, as shown in Figure 5. ReLU handles the vanishing gradient problems with a dropout rate that handles the overfitting problem. On the whole, the proposed BHCNet is proposed after analyzing the previous application of the CNN model that consisted of the larger dataset for training and deeper networks. Where we try to achieve the highest and accurate prediction results by conquering the challenges like overfitting, time consumption, accuracy, computation, complexity, and model depth.

The main aim is to develop the methodology to diagnose the brain hemorrhage with effective fast prediction using small CT scan images dataset which represents the novelty of this study. Hence the experiments with the proposed methodology are carried out successfully to overcome all the challenges. The proposed CNN model outperforms the hybrid models and achieve a perfect diagnosis with faster speed and highest accuracy rate to save life of the patient. The performance evaluation and comparative experimental results shows the effectiveness of the proposed study. The novelty of this study is the structure of the methods and methodology that is proposed with small dataset for the crucial times using deep learning CNN model.

## VI. CONCLUSION

The identification of the brain hemorrhage is proposed in this study based on the deep learning models i.e. CNN, hybrid CNN + LSTM and CNN + GRU. The image augmentation is adopted to increase the number of the training dataset from 180 to 1000 images. The experiments are carried out with balanced and imbalanced dataset. The first experimental phase implemented with the balanced dataset in which brain hemorrhage and non-brain hemorrhage classes are in equal numbers. The 95% accuracy achieved with balanced dataset using CNN model that shows the loss of one life because

the CNN model concentrates on the false-negative results mean the patient has brain hemorrhage in actual but CNN model predicts it non-brain hemorrhage. To eliminate all false negative cases, the second phase of the experiments achieved by imbalancing the dataset. The CNN model outperforms the CNN + LSTM and CNN + GRU by achieving 100% accuracy, 95.54% precision, 100% sensitivity, 100% specificity and 95.0% F1-score without any false prediction that save the patient life. Thus, the proposed model can diagnose the brain hemorrhage accurately with fast-speed and can help to save precious lives by predicting with 100% accuracy rate. The image segmentation will be considered as a future work because the color separation through segmentation will elaborate the regions of internal bleeding in CT scan more specifically.

## REFERENCES

[1] X. Ma, Y. Cheng, R. Garcia, and J. Haorah, "Hemorrhage associated mechanisms of neuroinflammation in experimental traumatic brain injury," *J. Neuroimmune Pharmacol.*, vol. 15, no. 2, pp. 181–195, Jun. 2020.

[2] A. Barman, V. Lopez-Rivera, S. Lee, F. S. Vahidy, J. Z. Fan, S. I. Savitz, S. A. Sheth, and L. Giancardo, "Combining symmetric and standard deep convolutional representations for detecting brain hemorrhage," *Proc. SPIE*, vol. 11314, Mar. 2020, Art. no. 113140D.

[3] M. Fallenius, M. B. Skrifvars, M. Reinikainen, S. Bendel, S. Curtze, G. Sibolt, N. Martinez-Majander, and R. Raj, "Spontaneous intracerebral hemorrhage," *Stroke*, 2019.

[4] M. Otterskog, N. Petrovic, and P. O. Risman, "A multi-layered head phantom for microwave investigations of brain hemorrhages," in *Proc. IEEE Conf. Antenna Meas. Appl. (CAMA)*, Oct. 2016, pp. 1–3.

[5] Y. Hu and Y. Zheng, "A GLCM embedded CNN strategy for computer-aided diagnosis in intracerebral hemorrhage," 2019, *arXiv:1906.02040*. [Online]. Available: http://arxiv.org/abs/1906.02040

[6] V. Davis and S. Devane, "Diagnosis & classification of brain hemorrhage," in *Proc. Int. Conf. Adv. Comput., Commun. Control (ICAC)*, Dec. 2017, pp. 1–6.

[7] R. Mahajan and P. M. Mahajan, "Survey on diagnosis of brain haemorrhage by using artificial neural network," *Int. J. Sci. Res. Eng. Technol.*, vol. 5, no. 6, pp. 378–381, 2016.

[8] M. K. Abraham and W. T. W. Chang, "Subarachnoid hemorrhage," *Emergency Med. Clinics*, vol. 34, no. 4, pp. 901–916, 2016.

[9] F. Ahmadabadi, M. Mirzarahimi, A. Ahadi, and Z. Alizadeh, "Frequency, causes, and findings of brain CT scans of neonatal seizure at Ardabil city hospital, Ardabil, Iran," *Int. Surg. J.*, vol. 7, no. 8, pp. 2485–2489, 2020.

[10] C. S. Anderson *et al.*, "Rapid blood-pressure lowering in patients with acute intracerebral hemorrhage," *New England J. Med.*, vol. 368, pp. 2355–2365, Jun. 2013.

[11] P. Annavi, "Method and system for identification of cerebrovascular abnormalities," U.S. Patent 16 365 548, Jan. 16, 2020.

[12] F. M. Siddiqui, S. V. Bekker, and A. I. Qureshi, "Neuroimaging of hemorrhage and vascular defects," *Neurotherapeutics*, vol. 8, no. 1, pp. 28–38, Jan. 2011.

[13] J. R. Little, B. Dial, G. Bélanger, and S. Carpenter, "Brain hemorrhage from intracranial tumor," *Stroke*, vol. 10, no. 3, pp. 283–288, May 1979.

[14] C. A. Taylor, J. M. Bell, M. J. Breiding, and L. Xu, "Traumatic brain injury–related emergency department visits, hospitalizations, and deaths—United States, 2007 and 2013," *MMWR. Surveill. Summaries*, vol. 66, no. 9, pp. 1–16, Mar. 2017.

[15] U. Balasooriya and M. U. S. Perera, "Intelligent brain hemorrhage diagnosis using artificial neural networks," in *Proc. IEEE Bus., Eng. Ind. Appl. Colloq. (BEIAC)*, Apr. 2012, pp. 128–133.

[16] J. Kim, T. Thayabaranathan, G. A. Donnan, G. Howard, V. J. Howard, P. M. Rothwell, V. Feigin, B. Norrving, M. Owolabi, J. Pandian, and L. Liu, "Global stroke statistics 2019," *Int. J. Stroke*, vol. 15, no. 8, 2020, Art. no. 1747493020909545.

[17] J. Morita, K. Miwa, T. Kitasaka, K. Mori, Y. Suenaga, S. Iwano, M. Ikeda, and T. Ishigaki, "Interactions of perceptual and conceptual processing: Expertise in medical image diagnosis," *Int. J. Hum.-Comput. Stud.*, vol. 66, no. 5, pp. 370–390, May 2008.

[18] L. M. Nentwich and B. Grimmnitz, "Neurologic emergencies in the elderly," *Emergency Med. Clinics*, vol. 34, no. 3, pp. 575–599, 2016.

[19] J. A. Neutze, "Computed tomography," in *Radiology Fundamentals: Introduction to Imaging & Technology*. 2020.

[20] R. W. Chan, J. Y. Lau, W. W. Lam, and A. Z. Lau, "Magnetic resonance imaging," in *Encyclopedia of Biomedical Engineering*. 2018.

[21] M. D. Hssayeni, M. S. Croock, A. Al-Ani, H. F. Al-khafaji, Z. A. Yahya, and B. Ghoraani, "Intracranial hemorrhage segmentation using deep convolutional model," pp. 1–18, 2019, *arXiv:1910.08643*. [Online]. Available: https://arxiv.org/abs/1910.08643

[22] F. A. Stevie and C. L. Donley, "Introduction to X-ray photoelectron spectroscopy," *J. Vac. Sci. Technol. A, Vac., Surf., Films*, vol. 38, no. 6, 2020, Art. no. 063204.

[23] A. Zamani, N. P. Ryan, D. K. Wright, K. Caeyenberghs, and B. D. Semple, "The impact of traumatic injury to the immature human brain: A scoping review with insights from advanced structural neuroimaging," *J. Neurotrauma*, vol. 37, no. 5, pp. 724–738, Mar. 2020.

[24] S. Badillo, B. Banfai, F. Birzele, I. I. Davydov, L. Hutchinson, T. Kam-Thong, J. Siebourg-Polster, B. Steiert, and J. D. Zhang, "An introduction to machine learning," *Clin. Pharmacol. Therapeutics*, 2020.

[25] S. Vieira, W. H. L. Pinaya, R. Garcia-Dias, and A. Mechelli, "Deep neural networks," in *Machine Learning*. 2020.

[26] M. Shahroz, M. Mushtaq, M. Ahmad, S. Ullah, A. Mehmood, and G. Choi, "IoT-based smart shopping cart using radio frequency identification," *IEEE Access*, vol. 8, pp. 68426–68438, 2020.

[27] S. Sriram, R. Vinayakumar, M. Alazab, and K. Soman, "Network flow based IoT botnet attack detection using deep learning," in *Proc. IEEE INFOCOM, IEEE Conf. Comput. Commun. Workshops (INFOCOM WKSHPS)*, Jul. 2020, pp. 189–194.

[28] V. Ravi, M. Alazab, S. Srinivasan, A. Arunachalam, and K. P. Soman, "Adversarial defense: DGA-based botnets and DNS homographs detection through integrated deep learning," *IEEE Trans. Eng. Manag.*, early access, Mar. 12, 2021, doi: 10.1109/TEM.2021.3059664.

[29] H. Chen, S. Khan, B. Kou, S. Nazir, W. Liu, and A. Hussain, "A smart machine learning model for the detection of brain hemorrhage diagnosis based Internet of Things in smart cities," *Complexity*, vol. 2020, pp. 1–10, Sep. 2020.

[30] A. Shalini, P. V. Vijayashree, S. Vishal, and S. C. Chacko, "An automated system of brain hemorrhage detection using deep learning," *Int. J. Sci. Res. Eng. Trends*, vol. 6, no. 2, pp. 1–5, 2020.

[31] V. Batumalai, S. Burke, D. Roach, K. Lim, G. Dinsdale, M. Jameson, C. Ochoa, J. Veera, L. Holloway, and S. Vinod, "Impact of dosimetric differences between CT and MRI derived target volumes for external beam cervical cancer radiotherapy," *Brit. J. Radiol.*, vol. 93, no. 1114, Oct. 2020, Art. no. 20190564.

[32] D. C. Ugwuanyi, T. F. Sibeudu, C. P. Irole, M. P. Ogolodom, C. T. Nwagbara, A. M. Ibekwe, and A. N. Mbaba, "Evaluation of common findings in brain computerized tomography (CT) scan: A single center study," *AIMS Neurosci.*, vol. 7, no. 3, p. 311, 2020.

[33] E. F. M. Wijdicks, "The first CT scan of the brain: Entering the neurologic information age," *Neurocrit. Care*, vol. 28, no. 3, pp. 273–275, Jun. 2018.

[34] F. Xu, Z. Pan, and R. Xia, "E-commerce product review sentiment classification based on a Naïve Bayes continuous learning framework," *Inf. Process. Manage.*, vol. 57, no. 5, 2020, Art. no. 102221.

[35] B. Gao, Y. Yang, H. Gouk, and T. M. Hospedales, "Deep clusteringwith concrete $K$-means," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2020, pp. 4252–4256.

[36] I. R. I. Haque and J. Neubert, "Deep learning approaches to biomedical image segmentation," *Informat. Med. Unlocked*, vol. 18, Jan. 2020, Art. no. 100297.

[37] Y. C. Hsu, Z. Lv, J. Schlosser, P. Odom, and Z. Kira, "Multi-class classification without multi-class labels," in *Proc. 7th Int. Conf. Learn. Represent. (ICLR)*, 2019, pp. 1–16.

[38] A. Sherstinsky, "Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network," *Phys. D, Nonlinear Phenomena*, vol. 404, Mar. 2020, Art. no. 132306.

[39] 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, Oct. 2020.

[40] J. Wu, "Introduction to convolutional neural networks," Nat. Key Lab Novel Softw. Technol., Nanjing Univ., Nanjing, China, Tech. Rep., , 2017, p. 23, vol. 5.

[41] 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.

[42] A. Helwan, G. El-Fakhri, H. Sasani, and D. U. Ozsahin, "Deep networks in identifying CT brain hemorrhage," *J. Intell. Fuzzy Syst.*, vol. 35, no. 2, pp. 2215–2228, Aug. 2018.

[43] A. Sharma and S. Jain, "Brain region segmentation using low MSE based active contour model and convolutional neural network," *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 9, no. 2, pp. 1848–1853, Apr. 2020.

[44] J. Tan, Y. Gao, Z. Liang, W. Cao, M. J. Pomeroy, Y. Huo, L. Li, M. A. Barish, A. F. Abbasi, and P. J. Pickhardt, "3D-GLCM CNN: A 3-dimensional gray-level co-occurrence matrix-based CNN model for polyp classification via CT colonography," *IEEE Trans. Med. Imag.*, vol. 39, no. 6, pp. 2013–2024, Jun. 2020.

[45] Z. Bi and X. Wang, *Computer Aided Design and Manufacturing*. 2020.

[46] A. M. Bagirov, N. Karmitsa, and S. Taheri, "Introduction to clustering," in *Partitional Clustering via Nonsmooth Optimization*. Springer, 2020, pp. 3–13.

[47] G. Yang, Y. Zhang, J. Yang, G. Ji, Z. Dong, S. Wang, C. Feng, and Q. Wang, "Automated classification of brain images using wavelet-energy and biogeography-based optimization," *Multimedia Tools Appl.*, vol. 75, no. 23, pp. 15601–15617, Dec. 2016.

[48] M. Rastgarpour, J. Shanbehzadeh, and H. Soltanian-Zadeh, "A hybrid method based on fuzzy clustering and local region-based level set for segmentation of inhomogeneous medical images," *J. Med. Syst.*, vol. 38, no. 8, p. 68, 2014.

[49] S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a convolutional neural network," in *Proc. Int. Conf. Eng. Technol. (ICET)*, Aug. 2017, pp. 1–6.

[50] M. Yu, Q. Huang, H. Qin, C. Scheele, and C. Yang, "Deep learning for real-time social media text classification for situation awareness—Using Hurricanes Sandy, Harvey, and Irma as case studies," *Int. J. Digit. Earth*, vol. 12, no. 11, pp. 1230–1247, 2019.

[51] S. Sriram, R. Vinayakumar, V. Sowmya, M. Alazab, and K. P. Soman, "Multi-scale learning based malware variant detection using spatial pyramid pooling network," in *Proc. IEEE INFOCOM, IEEE Conf. Comput. Commun. Workshops (INFOCOM WKSHPS)*, Jul. 2020, pp. 740–745.

[52] B. Weinstein, S. Fine, and Y. Hel-Or, "Margin-based regularization and selective sampling in deep neural networks," 2020, *arXiv:2009.06011*. [Online]. Available: http://arxiv.org/abs/2009.06011

[53] Q. Dou, H. Chen, L. Yu, L. Zhao, J. Qin, D. Wang, V. C. Mok, L. Shi, and P. A. Heng, "Automatic detection of cerebral microbleeds from MR images via 3D convolutional neural networks," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1182–1195, May 2016.

[54] M. Shahroz, M. F. Mushtaq, A. Mehmood, S. Ullah, and G. S. Choi, "RUTUT: Roman Urdu to Urdu translator based on character substitution rules and unicode mapping," *IEEE Access*, vol. 8, pp. 189823–189841, 2020.

[55] N. Kakhandaki, S. B. Kulkarni, K. Ramesh, and U. P. Kulkarni, "Classification of brain hemorrhages in MRI using Naïve Bayes-probabilistic kernel approach," *J. Cases Inf. Technol.*, vol. 21, no. 3, pp. 51–65, 2019.

[56] Y. Liu, R. Caballero, and J. M. Monteiro, "RadNet 1.0: Exploring deep learning architectures for longwave radiative transfer," *Geosci. Model Develop.*, vol. 13, no. 9, pp. 4399–4412, Sep. 2020.

[57] M. Grewal, M. M. Srivastava, P. Kumar, and S. Varadarajan, "RADNet: Radiologist level accuracy using deep learning for hemorrhage detection in CT scans," in *Proc. IEEE 15th Int. Symp. Biomed. Imag.*, Apr. 2018, pp. 281–284.

[58] K. Jnawali, M. R. Arbabshirani, A. E. Ulloa, N. Rao, and A. A. Patel, "Automatic classification of radiological report for intracranial hemorrhage," in *Proc. IEEE 13th Int. Conf. Semantic Comput. (ICSC)*, Jan. 2019, pp. 187–190.

[59] T. Lemos and J. H. Kalivas, "Self-optimized one-class classification using sum of ranking differences combined with a receiver operator characteristic curve," *Anal. Chem.*, vol. 92, no. 7, pp. 5354–5361, Apr. 2020.

[60] R. Vinayakumar, M. Alazab, S. Srinivasan, Q.-V. Pham, S. K. Padannayil, and K. Simran, "A visualized botnet detection system based deep learning for the Internet of Things networks of smart cities," *IEEE Trans. Ind. Appl.*, vol. 56, no. 4, pp. 4436–4456, Jul. 2020.

[61] M. Togacar, Z. Comert, B. Ergen, and U. Budak, "Brain hemorrhage detection based on heat maps, autoencoder and CNN architecture," in *Proc. 1st Int. Informat. Softw. Eng. Conf. (UBMYK)*, Nov. 2019, pp. 1–5.

[62] M. Farhadi and A. H. Foruzan, "Data augmentation of CT images of liver tumors to reconstruct super-resolution slices based on a multi-frame approach," in *Proc. 27th Iranian Conf. Electr. Eng. (ICEE)*, Apr. 2019, pp. 1783–1786.

[63] A. Simon, R. Vinayakumar, V. Sowmya, K. P. Soman, and E. A. A. Gopalakrishnan, "A deep learning approach for patch-based disease diagnosis from microscopic images," in *Classification Techniques for Medical Image Analysis and Computer Aided Diagnosis*. Amsterdam, The Netherlands: Elsevier, 2019, pp. 109–127.

[64] R. Vinayakumar, M. Alazab, K. Soman, P. Poornachandran, A. Al-Nemrat, and S. Venkatraman, "Deep learning approach for intelligent intrusion detection system," *IEEE Access*, vol. 7, pp. 41525–41550, 2019.

[65] M. D. Pandya, P. D. Shah, and S. Jardosh, "Medical image diagnosis for disease detection: A deep learning approach," in *U-Healthcare Monitoring Systems*. Amsterdam, The Netherlands: Elsevier, 2019.

[66] Q. Ke, J. Liu, M. Bennamoun, S. An, F. Sohel, and F. Boussaid, *Computer Vision for Human-Machine Interaction*. Amsterdam, The Netherlands: Elsevier, 2018.

[67] M. J. J. P. van Grinsven, B. van Ginneken, C. B. Hoyng, T. Theelen, and C. I. Sanchez, "Fast convolutional neural network training using selective data sampling: Application to hemorrhage detection in color fundus images," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1273–1284, May 2016.

[68] M. Lin, Q. Chen, and S. Yan, "Network in network," in *Proc. 2nd Int. Conf. Learn. Represent. (ICLR)*, 2014, pp. 1–10.

[69] M. Hashemi, "Enlarging smaller images before inputting into convolutional neural network: Zero-padding vs. interpolation," *J. Big Data*, vol. 6, no. 1, p. 98, Dec. 2019.

[70] A. F. Agarap, "Deep learning using rectified linear units (ReLU)," no. 1, pp. 2–8, 2018, *arXiv:1803.08375*. [Online]. Available: http://arxiv.org/abs/1803.08375

[71] C. Nwankpa, W. Ijomah, A. Gachagan, and S. Marshall, "Activation functions: Comparison of trends in practice and research for deep learning," pp. 1–20, 2018, *arXiv:1811.03378*. [Online]. Available: http://arxiv.org/abs/1811.03378

[72] A. Raza, A. Mehmood, S. Ullah, M. Ahmad, G. S. Choi, and B.-W. On, "Heartbeat sound signal classification using deep learning," *Sensors*, vol. 19, no. 21, p. 4819, Nov. 2019.

[73] A. M. Dawud, K. Yurtkan, and H. Oztoprak, "Application of deep learning in neuroradiology: Brain haemorrhage classification using transfer learning," *Comput. Intell. Neurosci.*, vol. 2019, pp. 1–12, Jun. 2019.

[74] P. Chang, E. Kuoy, J. Grinband, B. D. Weinberg, M. Thompson, R. Homo, J. Chen, H. Abcede, M. Shafie, L. Sugrue, and C. G. Filippi, "Hybrid 3D/2D convolutional neural network for hemorrhage evaluation on head CT," *Amer. J. Neuroradiol.*, vol. 39, no. 9, pp. 1609–1616, 2018.

[75] D. T. Ginat, "Analysis of head CT scans flagged by deep learning software for acute intracranial hemorrhage," *Neuroradiology*, vol. 62, no. 3, pp. 335–340, Mar. 2020.

[76] G. Vrbancic, M. Zorman, and V. Podgorelec, "Transfer learning tuning utilizing grey wolf optimizer for identification of brain hemorrhage from head CT images," in *Proc. 6th Student Comput. Sci. Res. Conf. (StuCoSReC)*, 2019, pp. 61–66.

[77] S. P. Singh, L. Wang, S. Gupta, B. Gulyás, and P. Padmanabhan, "Shallow 3D CNN for detecting acute brain hemorrhage from medical imaging sensors," *IEEE Sensors J.*, vol. 21, no. 13, pp. 14290–14299, Jul. 2021.

[78] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

[79] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 25, 2012, pp. 1097–1105.

[80] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: http://arxiv.org/abs/1409.1556

[81] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 1–9.

[82] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.

**MUHAMMAD FAHEEM MUSHTAQ** received the B.S. degree in IT and the M.S. degree in CS from The Islamia University of Bahawalpur, Punjab, Pakistan, in 2011 and 2013, respectively, and the Ph.D. degree from the Department of Information Security, Faculty of Computer Science and Information Technology, University Tun Hussein Onn Malaysia (UTHM), Malaysia, in 2018.

He has made several contributions through research publications and book chapters toward information security. He received Microsoft certifications of internet security and acceleration (ISA) server, Microsoft certified professional (MCP), and Microsoft certified technology professional (MCTS), in 2010. He is currently working as the Head of the Department of Artificial Intelligence, The Islamia University of Bahawalpur. Previously, he was working as the/an Head/Assistant Professor of the Department of Information Technology, Khwaja Fareed University of Engineering and Information Technology, Rahim Yar Khan, Pakistan. He has been working as a Research Assistant during his Ph.D. degree, from March 2016 to August 2018. He has been appointed as the Vice President of the UTHM's Graduates Student Association from 2017 to 2018. His research interests include information security, artificial intelligence, and cognitive systems and applications.

**MOBEEN SHAHROZ** received the MCS and M.S. degrees in computer science from the Department of Computer Science, Khawaja Fareed University of Engineering and Information Technology (KFUEIT), Rahim Yar Khan, Pakistan, in 2018 and 2020, respectively. He is currently pursuing the Ph.D. degree in computer science with The Islamia University of Bahawalpur. His current research interests include the Internet of Things (IoT), artificial intelligence, data mining, natural language processing, machine learning, deep learning, and image classification.

**ALI M. ASEERE** received the Ph.D. degree in computer science from the University of Southampton, U.K., in 2012. He has published many research papers in refereed international journals/proceedings/books. He is currently working as an Assistant Professor at the Department of Computer Science, King Khalid University, and also the Dean of the College of Computer Science. He served as the session chair and organizing committee member of various conferences. His research interests include intelligent and multi-agent systems, agent-based models, agent mining, and image processing.

**HABIB SHAH** received the Ph.D. degree from the Faculty of Computer Science and Information Technology, University Tun Hussein Onn Malaysia, in 2013. He is currently an Assistant Professor with the Department of Computer Science, College of Computer Science, King Khalid University, Saudi Arabia. His research interests include artificial intelligence, learning algorithms, data mining techniques, time series analysis, and numerical optimization. He has successfully published more than 40 articles in various international SCI and Scopus journals and conference proceedings. He is a member of an editorial board, a guest editor, and acts as a reviewer for various journals and conferences as well. He has also served as a Program Committee Member and a co-organizer for numerous international conferences/workshops. He is currently working on three research projects of KKU and KSA.

**RIZWAN MAJEED** received the B.Sc. degree in IT from the University of Punjab, in 2008, and the M.Phil. degree in computer sciences from NCBA, Pakistan, in 2018. He is currently pursuing the Ph.D. degree with UTHM, Malaysia. He joined The Islamia University of Bahawalpur as the Director of IT. His current research interests include the IoT-based smart systems, artificial intelligence, data mining, machine learning, image processing, and deep learning.

**DANISH SHEHZAD** received the B.S. degree from COMSATS University, in 2010, the M.S. degree in computer sciences from Hazara University, Pakistan, in 2014, and the Ph.D. degree from Kadir Has University, Turkey, in 2019. He joined the FAST National University of Computer and Emerging Sciences (NUCES) as an Assistant Professor, in January 2019. Before joining NUCES, he has teaching and research experience of five years. He worked as a Researcher on the brain-inspired run-time system for a very large scale brain simulation (BiRTS) project funded by the Scientific and Technological Research Council of Turkey. He also worked as a Senior Researcher on a joint venture between Kadir Has University and Selcuk University, Turkey, for the automatic detection of various diseases through medical image processing. He has published his research work in well-reputed international journals and also serves as a reviewer in multiple international journals and conferences.

**ALI SAMAD** received the B.Sc. degree from the University of Punjab, in 2004, the master's degree in computer sciences from NCBA, Pakistan, in 2008, and the M.S. degree, in 2013. He is currently pursuing the Ph.D. degree with UTHM, Johor, Malaysia. He joined The Islamia University of Bahawalpur. His current research interests include the IoT-based smart systems, artificial intelligence, data mining, machine learning, image processing, and deep learning.

• • •