European Journal of Molecular & Clinical Medicine
ISSN 2515-8260 Volume 07, Issue 07, 2020

Brain Tumor And Intracranial Haemorrhage Feature Extraction And Classification Using Conventional And Deep Learning Methods

R. Aruna Kirithika[1], S. Sathiya[2], M. Balasubramanian[3], P. Sivaraj[4]

[1]*PhD scholar in Computer and Information Science,*
[2]*Assistant Professor in Computer Science Enginerring,*
[3]*Associate Professor in Computer Science Enginerring,*
[4]*Associate Professor in Manufacturing Engineering*
[1,2,3,4]Annamalai University, Tamilnadu, India.

E-mail : [1]ashassss79@gmail.com, [2]sathiya.sep05@gmail.com, [3]balu_june1@gmail.com, [5]cemajorsiva@gmail.com

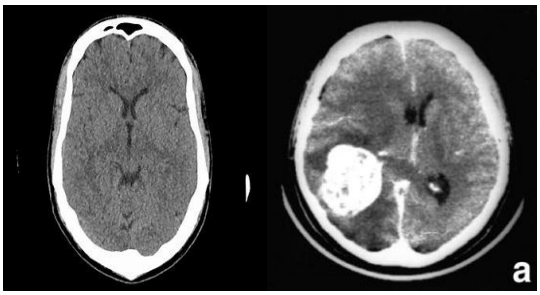
Abstract: Presently, brain tumor (BT) and Intracranial hemorrhage (ICH) detection and classification processes become essential to save human lives. Automated diagnosis model using deep learning (DL) models finds useful to attain improved diagnostic outcome. This paper presents an ensemble of handcrafted and deep features for BT and ICH diagnosis. The proposed model comprises of three important processes, such as preprocessing, feature extraction and classification. The preprocessing of the input image takes place in three ways namely skull stripping, bilateral filtering (BF) and contrast limited adaptive histogram equalization (CLAHE) based contrast enhancement. In addition, scale invariant feature transform (SIFT) and AlexNet models are used for feature extraction process. In order to classify the existence of BT and ICH, two classification models is carried out such as gaussian naïve bayes (GNB) and random forest (RF).For validating the effective diagnostic performance of the proposed model, a set of simulations were carried out to determine the different class labels. The simulation outcome indicated the effective performance with the maximum sensitivity of 92.41%, specificity of 100%, and accuracy of 94.26%.

Keywords: AlexNet, Brain tumor, Classification models, Feature extraction, Intracranial haemorrhage.

1. INTRODUCTION

In general, Brain Tumour (BT) is defined as a group of biological cells developed within the brain tissues. The anomalous cell development is increased gradually inside the skull which covers the brain. As a result, severe consequences are experienced by the patient where the mass growth inside the skull promotes to cultivate more abnormal tissues. BT is classified into 2 types namely, Benign/non-cancerous tumor and malignant/cancerous named as malignant neoplasm. These BTs are highly dangerous for the patient which tends to cause severe problems that reduce the lifetime of a human being. Human brain is an important internal organ which is embedded with massive number of cells. The unwanted cell development is evolved from the unrestrained cell segmentation, called as a tumor. Then, BT is one of the dreadful diseases that come under the class of cancer disease. In order to limit

the cell growth, earlier prediction of BT is more important to find the root causes of disease and treat accordingly to save the life of a patient. Nowadays, severe cancer disease is also treated by developing medical services, particularly in the earlier phase of disease. The probability of survival rate can be increased only when it is predicted in the earlier stage and acquire the treatment accordingly. In recent times, BT is caused for massive peoples globally. An interface has to be developed to examine the dangerous cells where the increasing mortal rate can be reduced in the earlier phase. Also, BT is caused for both male and female and at any age group. In last decades, numerous peoples were subjected for benign tumors.



(a) (b)

Fig. 1. (a) Normal brain image (b) Tumor image

Under the application of clinical imaging, different types of brain and Central nervous system (CNS) tumors [1] have been predicted. Even though there are enormous models used for BT classification, still massive number of constraints is involved which has to be resolved gradually. The benign and malignant tumor classification is considered to be binary classification which is insufficient for radiologists to take precautionary measures and treat the patient. In order to get a clear suggestion for radiotherapists, multi-class classification is essential to classify the tumour and relevant types. Also, the lack of advanced knowledge is one of the major challenging issues for developers to gain effective outcomes. In order to overcome these problems, a Convolutional Neural Network (CNN) approach has been applied on diverse data augmentation to achieve the consequences for multi-grade BT classification.

On the other hand, Intracranial Hemorrhage (ICH) is an alternate critical stage for human which exists globally. A hemorrhage occurs inside the brain parenchyma (intra-axial) or the cranial vault; however, it exists externally to brain parenchyma (extra-axial). These intra-axial and extra-axial hemorrhages are highly dreadful locations where productive medical service is essential. For instance, intra-axial hemorrhage occurs in massive number annually in US with a short time period. Additionally, the survivors of subarachnoid hemorrhage (extra-axial hemorrhage) endure fixed cognitive impairment. Hospital entries of ICH have enhanced gradually in last decades because of the increasing lifestyle and worse blood pressure control. Specifically, the primary analysis of ICH is a complex clinical service as many number of people die out of ICH and related disease within limited time intervals and earlier predictions enhance the health condition.

There are several ICH catogeries as mentioned in the Fig.2.EDH-epidural haemorrhage,SDH-subdural haemorrhage,SAH-subarachnoidal haemorrhage, ICH-intraparenchymal haemorrhage and IVH-intraventricular haemorrhage.

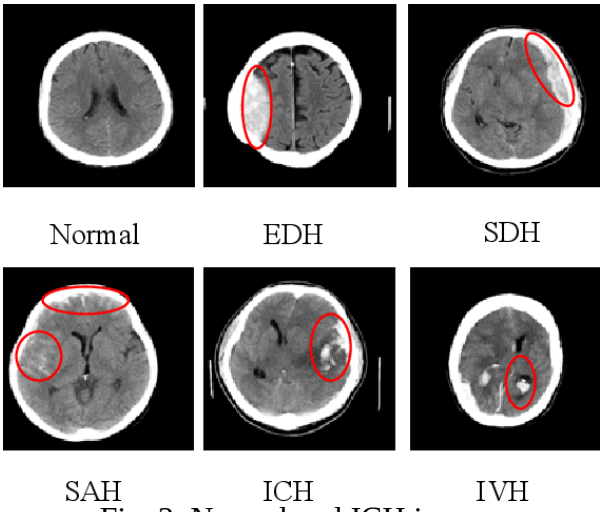


Fig. 2. Normal and ICH images

Computed tomography (CT) is one of the well-known screening mechanisms applied for acute ICH diagnosis, and duration of examination depends upon how rapidly a head CT is completed and interpreted by a physician. Moreover, interpretation duration of radiological works is related on preference of physicians and by considering the patient's health condition (inpatient vs. outpatient). Typically, several studies are examined with limited time whereas routine outpatient studies take maximum time which depends upon accessible radiology workforce. Hence, the predictions of ICH in routine works are developed significantly. ICH also exists in outpatient settings, albeit with a low frequency when compared with inpatient or casualty setting. For instance, aged outpatients on anticoagulation therapy are prone to ICH. Especially, earlier symptoms might be vague, frequent non-emergent, routine head CT.
Though several diagnosis models for ICH and BT are available in the literature there is still a requirement to achieve enhanced diagnostic outcomes. In this view, this paper presents a new ensemble of handcrafted with deep features model for BT and ICH diagnosis. The presented model comprises different subprocesses namely preprocessing, feature extraction, and classification. The presented model involves scale-invariant feature transform (SIFT) and AlexNet models are used for feature extraction process. Besides, two classification models are used to classify the existence of BT and ICH namely gaussian naïve bayes (GNB) and random forest (RF).

Related works

This section offers a brief survey of BT and ICH diagnosis models to identify the existence of disease with the allocation of different class labels. A set of ML and DL based diagnosis models for BT and ICH are reviewed in detail.

Prior BT classification models

In recent times, computer-aided clinical examination offers better functions and results which intend to apply Deep Learning (DL) paradigms. The DL principles were applied widely in clinical image analysis of breast cancer and lung cancer diagnosis. [2] deployed a DL method for human skin prediction which is a portion of dermatology. [3] employed a deep CNN to observe brain metastases. Moreover, the specific type of DL is named as Deep Transfer Learning (DTL) which are considered as effective studies on visual classification, object prediction, and image categorization issue. Moreover, TL enables the application of pre-trained CNN mechanism that is deployed for related works. [4] utilized a pre-trained InceptionV3 method for discriminating benign and malignant renal tumors on CT photographs. [5] projected a classifying breast cancer on histopathologic photographs.

Developers have applied a pre-trained VGG-16 framework and fine-tuned AlexNet to extract features are classified under the application of Support Vector Machine (SVM). [6] established a learning mechanism for computing lung tumor and pancreatic tumor classification. A learning scheme depends upon knowledge transfer with 3D CNN structure. In particular, TL has gained maximum attention from neuro-oncology works. Hence, the extraction of deep features from brain magnetic resonance imaging (MRI) images under the application of pre-trained systems. Therefore, literature depicted the ability of TL to be operated with tiny datasets.
[7] employed AlexNet and GoogLeNet for ranking the glioma from MRI images. By means of performance metrics, GoogLeNet ensured the supremacy of AlexNet under various operations. [8] established standard classification function with DTL on brain anomalous classification. Researchers have employed ResNet-34 and experiments with training of dense layers which is trained with data augmentation and fine-tune the TL scheme. The performance outcomes have represented DTL framework is employed in clinical image categorization, with reduced pre-processing. [9] utilized a pre-trained VGG-16 system for diagnosing Alzheimers disease from MRI. The TL has been utilized for Content-Based Image Retrieval (CBIR) for BTs. The performance validation on commonly accessible dataset and accomplished effective outcomes.

Prior works on ICH Diagnosis
Nowadays, Artificial Intelligence (AI) has exhibited challenging results in clinical imaging applications. Only few works have managed to predict the anomalies in head CT with ICH under the application of DL or Machine Learning (ML) methodologies. [10] illustrated the usage of DL for predicting test findings in head CT with the help of tiny dataset with acute ICH cases. [11] addressed maximum diagnostic value for SAH forecasting with the help of supervised ML method to numerous subjects with malicious SAH. The recent study by [12] utilized a hybrid CNN under the application of slice slabs on a dataset with massive training CT scans and testing CT scans from single institution for ICH prediction as well as quantification. Therefore, maximum dataset with minimum ICH affected cases and not all ICH subtypes have been examined in this literature. Alternatively, [13] applied DL for automated prediction of critical findings in head CT scans with ICH with maximum count of scans. A 2-phase approach has been utilized which means that 2D-CNN is employed for accomplishing slice-level confidence as well as RF is utilized for predicting subject-level possibility. It has to be pointed that, the models are relied on 2D or slice slabs, and subject-level detection is achieved by traversing by each slice with slice-level results with post-processing. Slice-level labels are essential for training. Huge efforts were taken by [14] to use 3DCNN-based framework for predicting ICH, where CNN system with 5 conv. layers and 2 Fully Connected (FC) layers have been applied and subject-level labels were employed as ground truths for training. The function of plain 3D CNN is extendable with maximum AUC, sensitivity, and specificity at selected operating point. It is still unknown whether straightforward models like 2D, hybrid, or simple 3D are applicable to produce scalable predictions.

The Proposed Model
The working process involved in the proposed method for BT and ICH diagnosis is demonstrated in Fig. 1. The figure states that the input image is initially preprocessed in three levels to enhance the image quality. Then, a set of handcrafted and deep features are extracted by the use of SIFT and AlexNet model. At last, the extracted feature vectors are fed as input to the GNB and RF to identify the different set of class labels present in the image. The elaborative explanation of the presented method is discussed below.

Algorithm:
Step 1: Acquiring Input (deferred images-both tumor and ICH)
Step 2: Pre-processing the acquired input
2.1 Skull stripping
2.2 Bilateral Filtering
2.3 Contrast Enhancement
Step 3: Feature Extraction
Step 4: Classification – two catogeries i) Tumor
ii) ICH.

The algorithm given above describes the steps that are followed in classifying the acquired brain image whether it belongs to the catogery of tumor or intracranial haemorrhage.

Pre-processing

Firstly, the input image is preprocessed in three levels skull stripping, noise removal, and contrast enhancement. Generally, it is needed to remove the skull region from the background region for better clarity. Afterward, bilateral filtering (BF) technique is applied on the image to discard any noise exists in it. Besides, contrast limited adaptive histogram equalization (CLAHE) technique is employed to increase the contrast level of the applied image.

Skull Stripping

Initially, skull stripping is performance on the input brain MRI image and it is essential to remove the skull from background region from MRI for

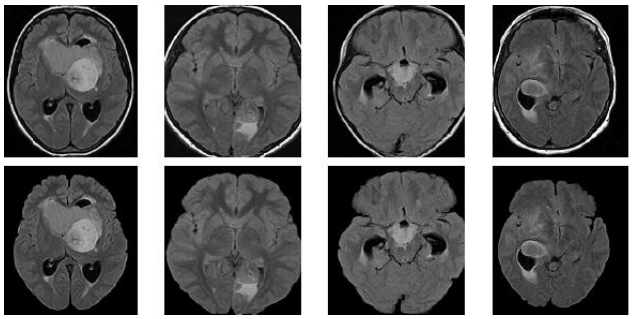


Fig. 3. Images of CT brain image –input image / pre-processed(skull stripped)

quantitative analysis. Usually, skull stripping

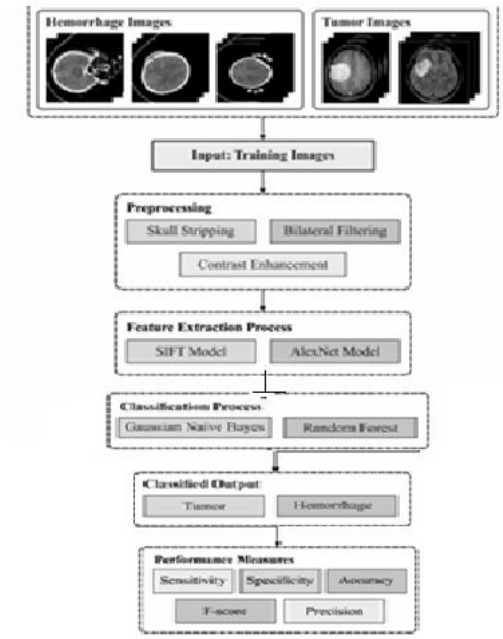
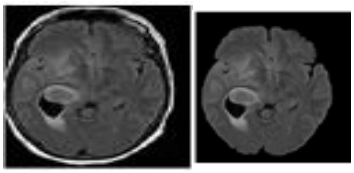


Fig. 4. Block diagram of proposed model

is processed with the application of image filter, which isolated the skull and remaining image sections by covering the pixels with identical intensity levels. For MRI images, the skull or bone section has higher threshold value (threshold > 200) when compared to tumor and alternate brain regions. Therefore, image filter has been employed to isolate brain regions according to the selected threshold value. Besides, using a solidity feature, skull is cropped from brain MRI.



(a) (b)

Fig. 5. (a) Acquired input image
(b)Skull stripped image

Bilateral Filtering

Once the skull stripping process is done, the noise removal process is carried out using BF technique. Tomasi and Manduchi [15] developed a bilateral filter which is considered as a non-iterative and non-linear filter to conserve the edges at the time of noise elimination. The neighboring pixel's geometric closeness and the likenesses of gray level have been assumed. In case of local neighborhood, the BF calculates the weighted sum of pixels. Each pixel has a neighboring weighted average used to replace the pixel. Then, the neighborhood weights are obtained using spatial and intensity distance of pixels. For a pixel neighborhood, the spatial distance has been found in domain (spatial) filter coefficients whereas the range filter weight is relevant to pixel radiometric distance.

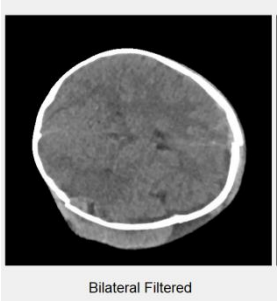


Fig.6.Bilateral filtering output

Therefore, the BF output is attained from various pixel locations as given below,

$$i'(x) = \frac{1}{\mathrm{N_c}} \sum_{y \in n(x)} e^{\left(\frac{-||y-x||^2}{2sd_d^2}\right)} e^{\left(\frac{-||i(y)-i(x)||^2}{2sd_r^2}\right)} i(y) \quad (1)$$

where spatial neighborhood of $i(x)$ is exhibited as $n(x)$, sd_d and sd_r are 2 variables employed to control the tradeoff between spatial and intensity domains weight. Hence, the normalization constant from given function is defined as

$$N_c = \sum_{y \in n(x)} e^{\left(\frac{-||y-x||^2}{2sd_d^2}\right)} e^{\left(\frac{-||i(y)-i(x)||^2}{2sd_r^2}\right)} \quad (2)$$

However, the edge conservation as well as noise elimination contributes in the BF technique, which is referred to be effective unlike classical filters [16]. But, the implementation of BF technique depends upon the d_d, sd_r, and $n(x)$ filter variables.

Enhanced Local contrast enhancement

Next to noise removal process, the contrast level of the resultant image is improved by the Enhanced Local Contrast Enhancement(CLAHE technique). Normally, histogram equalization (HE) is a special case of common histogram remapping models. It is highly referred by the developers due to its remarkable advantages like robustness and supreme effect which tends to improve the contrast of MRI images.

Histogram is defined as a function of gray level that represents a gray level of each pixel. Thus, a contrast ratio can be maximized using gray nonlinear transform to change the accumulation process whereas gray in minimum radius would be converted as complete field. A histogram is considered to be a discrete function which is depicted as follows:

$$p_r(r_k) = \frac{n_k}{n} \quad (3)$$

where n means the overall pixels of an MRI image and $n_k$refers pixel value of r_k gray level. Hypothesis of a gray transfer function is $s = T(r)$, where slope is mitigated to non-minus continuum monotone increasing function, and the input image $I(x, y)$ is changed into output image $I'(x, y)$. Consider $p_r(r)$ and $p_s(s)$implies the probability density function of random variables$r$ and s, rrefers a gray level of an input image and $s$defines the gray level of an output image [17]. Based on the HE, it cumulates density function, actual image histogram, and computed histogram regions are symmetric to given function

$$p_s(s) = p_r(r)\frac{dr}{ds}. \quad (4)$$

Consider that s belongs to $[0, L-1]$, then gray transfer function is depicted as:

$$s = T(r) = (L-1)\int_0^r p_r(w)dw, \quad (5)$$

where w denotes an integral dummy feature.

Based on the features of integral, the expression of Eq. (5) is formulated as.

$$\frac{ds}{dr} = \frac{dT(r)}{dr}(L-1)\frac{d}{dr}\left[\int_0^r p_r(w)dw\right] = (L-1)p_r(r). \quad (6)$$

Furthermore, apply the Eq. (6) into (4) and accomplish a novel function as given below:

$$p_s(s) = p_r(r)\left|\frac{dr}{ds}\right| = p_r(r)\left|\frac{1}{(L-1)p_r(r)}\right| \quad (7)$$

$$= \frac{1}{L-1}, 0 \le s \le L-1.$$

Based on (3) and a gray transfer function (5) transform, the discrete format has been applied:

$$s_k = T(r_k) = (L-1)\sum_{j=0}^{k} p_r(r_j) = \frac{L-1}{n}\sum_{j=0}^{k} n_j, \quad (8)$$

$$k = 0,1,2,\dots,L-1.$$

Generally (8) is a gray level remapping process. When the complete HE is compared, Adaptive HE (AHE) is beneficial for good local contrast enhancement. Then, AHE requires computation of local histogram and accomplishes distribution function for all pixels which is highly intensive. Followed by, AHE is sensible to noise. The AHE improves image contrast and suppress the noise. At certain point, the enhancement process intends in image dispersion so that visible analysis is affected. The feature and image contrast has to be improved and reduce magnified noise.

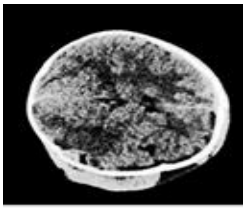


Fig. 7. Contrast Enhanced output

Therefore, reducing the contrast function to AHE for all blocks is required to produce transform function. The primary adjustment is essential for every block at the pyramid level image. Hence, a limited function has been applied for limiting a gray level probability density as well as to manage the additional histogram.

Feature Extraction

In this stage, the preprocessed images are fed as input to the feature extraction module, where a set of handcrafted features using SIFT and deep features using AlexNet are extracted.

Scale-invariant Feature Transform

David Lowe introduced Scale-Invariant Feature Transform, which is defined as a digital image predictor for mapping and examination of digital image. Some of the SIFT descriptors are composed of descriptor vectors that is applied in point matching from various perceptions of a scene and object detection in computer vision. The descriptors are generally, rapid in rotation, scaling, translation conversion in image, and robust for different brightness. Thus, the SIFT are applicable for image matching and prediction in the real-time platform.

The SIFT descriptor is composed of predicting interest point from an image that is a grey level image for a local gradient direction of image intensities are combined to show the extensive definition of an image structure of local neighborhood over the key-point that is applied in mapping equivalent keypoints among various image [18]. Next, SIFT descriptor has been utilized at huge grids which results in better function of object classification, texture

categorization, image organization as well as biometrics. There are 4 main phases in SIFT model.

Scaled-space Extrema Detection

This stage is the initial state of searching all scales as well as image position. Then, it applies a difference-of Gaussians function to find best points that are considered as invariant to scale and orientation. Laplacian of Gaussian (LoG) has been estimated for an image with different σ values that is considered as a blob detector to predict the blobs of various sizes along with a modified σ. Gaussian kernel with minimum σintends to generate maximum image whereas Gaussian kernel with higher σ, which fits correctly for edge corner. Therefore, it identifies a local maxima over distinct scale and space with a vector of (x, y, σ) values. Thus, J scale is a potential keypoint at (x,y) where it is depicted in Eq. (9).

$$L(x, y, \sigma) = G(x, y, \sigma) * I(\mathrm{x}, y\} \tag{9}$$

where Ldenotes the blurred image with σquantity of blur, $G = \frac{1}{2\pi\sigma^2} e^{-\frac{-(x^2+y^2)}{2\sigma^2}}$ means the Gaussian Blur operator,$I(x, y)$ defines the pixel at row x and column y of an image I, and$*$refers 2D convolution operator in x and y. When the quantity of blur image is σ, then the volume of blur in upcoming level might be $k \times \sigma, k$defines a constant, which is represented by $2^{\frac{1}{Numberofblurrediimages+1}}$. LoG is expensive when compared with Difference of Gaussians (DoG), hence SIFT applies DoG which is the extension of LoG. The DoG is processed from the variations of Gaussian blurring with 2 nearby values of σ, and assume the σ and kσ as given in Eq. (10)

$$D(x, y, \sigma) = L(x, y, k\sigma) - L(x, y, \sigma) \tag{10}$$

Where $L(x, y, k\sigma)$ and $L(x, y, \sigma)$represents the blurred image with blur quantity $k\sigma$ and σcorrespondingly. Afterward, the local extrema over scale and space have been explored in the image. For each pixel with spatial location x and y, I image I', the pixel with spatial position $(x - 1\ y - 1), (x - 1, y), (x - 1, y + 1), (x, y - 1), (x, y + 1), (x + 1, y - 1), (x + 1, y)$ and $(x + 1, y + 1)$ on recent image I'as well as subsequent scale images I'' and I'''mimics the neighboring pixels in conjunction with positions (x, y) in I'' and I'''. When a pixel value is higher, then it is named as maxima point whereas if the pixel value is lower, then it is termed as minima point. These maxima and minima points are assumed to be the candidate keys. An effective keypoint has been found when it is identified as local extrema that refer that keypoint is discovered.

Keypoint Localization

In all candidate positions, a brief method is fit to compute a location and scale. Keypoints are decided on the basis of measures which are reliable. Once the keypoint location is determined, it is refined for accomplishing effective outcomes. Taylor series expansion of scale-space was employed for attaining exact position of extrema. Followed by, a keypoint is eliminated when the intensity of extrema is lower than a previous threshold value. Moreover, the edges are rejected as DoG is high in response for edges under the application of Harris corner detector. The principle curvature can be computed by applying a 2x2 Hessian matrix (H). Additionally, candidate key-point $p(i, j)$ at coordinate (i, j), Hessian matrix is estimated as given in the following:

$$H = \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix} \tag{11}$$

Where h_{11},h_{12}, h_{21}, h_{22} are

$$h_{11} = p(i + 1, j) + p(i - 1, j) - 2 * p(i, j)$$
$$h_{12} = h_{21} = p(i + 1, j) + p(i, j - 1) - 2 * p(i, j)$$

$$h_{22} = (p(i+1, j+1) - p(i+1, j-1) - p(i-1, j+1) + p(i-1, j-1))/4$$

If $\frac{(h_{11}+h_{22})^2}{(h_{11}*h_{22})-(h_{22})^2} < \frac{(C_{edge}+1)^2}{C_{edge}}$, then maintain the key-point, else remove them. Here, Cedge is a ratio among maximum and non-zero eigenvalues in an image. From Harris corner, it is evident that one value is higher when compared with alternate values. Hence, keypoint would be eliminated when the ratio is greater than a threshold. The low contrast keypoint as well as edge key-point have been eliminated where keypoint is accomplished.

Orientation Assignment

According to the local image gradient directions, massive orientations have been allocated to a keypoint position. The image operation is converted where the allocated orientation, scale, and position for every feature that offers invariance to these conversions. For accomplishing invariance to image rotation, the orientation has been allocated to all key points. Based on the scale, a neighborhood point has been selected over the keypoint place. Followed by, the gradient magnitude and direction have been estimated. Usually, by gradient magnitude as well as Gaussian-weighted circular window. Next, a histogram with maximum peak has been measured and the orientation develops a key point with identical position and scale. Moreover, it contributes in stability matching.

Keypoint Descriptor

The decided scale, local image gradients have been estimated in a region for all key points. Followed by, it is transformed as a representation which is operated ineffective levels of local shape distortion and brightness. Afterwards, 16x16 neighborhoods over a key point have been selected and classified as 16 sub-blocks of 4x4 size block. For each sub-block, an 8 bin orientation histogram was developed that results in 128 bin value which implies a vector developing keypoint descriptor. Moreover, it is mainly applied to accomplish rapid illumination, rotation, noise, and so on.

Keypoint Matching

While identifying the nearest neighbors, keypoint among same images are mapped. Moreover, it is referred as latter closest-match might be differed from first portion because of the existence noise and alternate aspects. At this point, proportion of closest-distance for second-closest distance has been accomplished. It is eliminated when it is higher than 0.8. Hence, the phases are removed to a greater extent and minimum proportion is added or retained.

AlexNet

Usually, CNN is developed as a multi-layer interconnected NN, where the energetic minimum-, intermediate-, and high-level features have been extracted in a hierarchical manner. The CNN model is composed of 2 major layers namely, convolutional and pooling layers which is jointly named as convolutional base of the system. Some of the models are AlexNet and VGG, which are comprised of FC layers. Initially, conv. layer has an extracting function which filters the spatial characteristics from the images. The convolutional layers gain low-level features like edges and corners whereas the final convolutional layers gain high-level features like image architectures. As a result, CNN represents the efficiency to learn spatial hierarchical patterns. Moreover, conv. layers are described with the help of 2 elements namely, convolution patch size as well as depth of output feature map which means the filter count.

Specifically, a rectangular sliding window and permanent sized stride has been utilized for generating convoluted feature maps under the application of a dot product from the weights

of kernel and tiny region of input. A stride is a distance from 2 subsequent convolutional windows. Most probably, stride 1 is used in conv. layers as maximum stride values tend to make down-sampling in feature maps. Followed by, a feature map is defined as a novel new image produced by elegant convolution process which is visualized by obtained features. The weight-sharing features of CNN, count of attributes are limited when compared with FC layer, as the neurons in a specific feature map distribute similar attributes (weights and biases). A non-linearity function like Rectified Linear Unit (ReLU) is employed as element-wise nonlinear activation function for all components in a feature map. The ReLU function is highly beneficial to traditional activation functions employed in CNN like hyperbolic tangent or sigmoid functions, for the enclosing non-linearity to the system.

Next, ReLU stimulates training phase to classical functions with the help of Gradient Descent (GD). It is named as diminishing gradient problem where the functions of previous functions were extremely reduced in a saturating region and the updates for weights are diminished. Hence, pooling layers are applied after convolutional layer for mitigating the variance of features extracted with the help of maximizing or averaging operations. These processes are used in computing maximum and mean values, under the application of fixed-size sliding window as well as previous stride in feature maps in which it is conceptually same as conv. layer. Unlike to convolutional layers, a stride 2 has been employed in pooling layers for down-sampling the feature maps.As recommended above, few system have FC layers in prior to classifier layer which connects the results of various stacked convolutional and pooling layers into classifier layer. Consequently, final layer is named as classification layer that computes the posterior possibilities for every class.

Basically, AlexNet is a well-known model preferred by every researcher and several research communities due to its remarkable efficiency. The image classification is performed effectively when compared with traditional approaches. Before developing the DL mechanism, AlexNet is highly referred by every system with numerous counts of parameters and neurons. Initially, activation function was applied to enhance the performance. Activation function is employed in NN for providing non-linearity. Hence, the classical activation functions are logistic function, $\tanh$ function, $\arctan$ function, and so on. However, in deep models, the above-mentioned intends to implement gradient diminishing issues as gradient is considered as maximum value only if the input is minimum. These problems can be resolved by using a novel activation function in the form of ReLU. The actual definition of ReLU is expressed as:

$$ReLU(x) = \max(x, 0) \tag{12}$$

The gradient of ReLU is 1 when the input is higher than 0. It ensures that the deep networks with ReLU are the activation function which converges more rapidly than $\tanh$ unit. Therefore, the training speed has been increased. Then, dropout is utilized for eliminating the over-fitting issues. It is employed in FC layers. In case of dropout, the neurons undergo training for all iterations. Also, it promotes a neuron to work with alternate neighbors so that the joint adaptation among neurons is reduced and maximizes the generalization process [19].

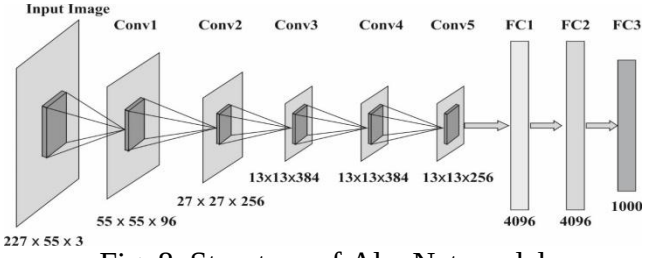


Fig. 8. Structure of AlexNet model

A network is classified into various sub-networks along with a dropout. Even though it has single sub-network, it is over-fitted for a greater limit, however, it shares the identical loss function. Next, the result of complete network is considered as the average of sub-networks. Finally, the dropout enhances efficiency. Fig. 8 shows the structure of AlexNet model.
Convolution and pooling have been applied for automated feature extraction and limitation. Convolution is applied in signal analysis. The image M in size of (m, n), the convolution is illustrated as,

$$C(m,n) = (M * w)(m,n) = \sum_{k}\sum_{l} M\,(m-k, n-l)w(k,l) \tag{13}$$

where w implies the convolution kernel in size of (I_k, l). It provides a solution for learning features from images and the variables to limit the model difficulty. Afterward, pooling is treated as feature reduction approach. It is assumed to be a collection of neighboring pixels in feature maps and produces a value. A feature map has 4×4, and max pooling offers a max value of 2×2 block, so that feature dimension might be limited.
Cross-channel normalization is evolved from local normalization scheme which enhances the generalization. Followed by, feature maps undergo normalization in prior to feed the upcoming layers. Moreover, it provides a sum from various neighboring maps simultaneously. It is identified in practical neurons. FC layers are applied in classification process. Hence, neurons in next FC layers are connected directly. The activation function in FC layers are named as softmax that is demonstrated as,

$$softmax\,(x)_i = \frac{\exp(x_i)}{\sum_{j=1}^{n} \exp\,(x_j)} \tag{14}$$

In this scenario, the final layer of the AlexNet is considered as the GNB and RF classifier layers to carry out the classification processes.

Classification
Finally, the extracted set of feature vectors are employed to the classification models to identify the existence of BT and ICH using GNB and RF classifiers.

Gaussian Naive Bayes (GNB)
The Naive Bayes (NB) classifier is a well-known Bayesian network with single root node which implies a class and n leaf nodes show the attributes. Consider that C is a class label with $k$feasible values, and $X_1 \ldots X_n$is a collection of parameters involved in the finite domain $D(X)$ where $i = 1..n$. A classifier is provided by the integration of Bayesian probabilistic model along with Maximum A Posteriori (MAP) rule, also named as discriminant function. Therefore, NB classification is illustrated as given in the following:

$$NBayes(a) = arg\max_{c \in C} P(c) \prod_{i=1}^{n} P\,(x_i|c) \tag{15}$$

where $a = \{X_1 = x_1, \ldots X_n = x_n\}$means a complete designation of features, x_i refers a short for $X_i = x_i$ and $c$implies short for $C = c$. The function is considered with conditional independence between attributes.
In order to deal continuous variables, domain of parameters is portioned; however, it results in data loss. Effective technology has been presented in [20] called as Fuzzy Bayesian classification scheme which is a hybrid algorithm where attributes undergo fuzzification in prior to computing classification. Here, degrees of truth have been considered as possibility of $P(x_i|a) = \mu_{x_i}$ and $P(c|a) = \mu_c$. Though the degrees of truth exhibit membership values of classes rather in probabilities, it is an extended version of NB classifier by the Bayes' rule and considers the independence between these features:

$$P(c|a) = \sum_{x_1 \in X_1, .. x_n \in X_n} P\,(c|x_1 .. x_n) P(x_1|a) .. P(x_n|a) \quad (16)$$

$$P(c|a) = \sum_{x_1 \in X_1, .. x_n \in X_n} \frac{P(x_1|c) .. P(x_n|c) P(c)}{P(x_1) .. P(x_n)} \mu_{x_1} .. \mu_{x_n} \quad (17)$$

The Fuzzy NB classification method is provided below:

$$FNBayes(a) = arg\max_{c \in C} P(c).$$

$$\sum_{x_{1j} \in X_1} \frac{P(x_{1j}|c)}{P(x_{1j})} \mu_{x_{1j}} .. \sum_{x_{nj} \in X_n} \frac{P(x_{nj}|c)}{P(x_{nj})} \mu_{x_{nj}} \quad (18)$$

where $j = 1..D(X_i)$ and $\mu_{x_{ij}} \in [0,1]$ refers a Membership Function (MF) of attribute value $x_{ij} \in X_i$ in a novel instance a. The degrees of truth should be generalized where $\sum_{x_{ij} \in X_i} \mu_{x_{ij}} = 1$ for all parameters $i = 1..n$. The possibilities essential by fuzzy model is determined as same as traditional NB classifier (15)

$$P(C = c) = \frac{\left(\sum_{e \in L} \mu_c^e\right) + 1}{|L| + |D(C)|} \quad (19)$$

$$P(X_i = x_i) = \frac{\left(\sum_{e \in L} \mu_{x_i}^e\right) + 1}{|L| + |D(X_i)|} \quad (20)$$

$$P(X_i = x_i | C = c) = \frac{\left(\sum_{e \in L} \mu_{x_i}^e \mu_c^e\right) + 1}{\left(\sum_{e \in L} \mu_c^e\right) + |D(X_i)|} \quad (21)$$

where Laplace-correction is used for smoothing estimations to eliminate the extreme values accomplished with tiny training sets. In this approach, Ldenotes the training samples e, where $e = \{X_1 = x_1, X_n = x_n, C = c\}$, $|L|$refers the count of instances $e \in L$, $\mu_c^e \in [0,1]$represents the degree of truth of $c \in C$ in sample $e \in L$, and $\mu_{x_i}^e \in [0,1]$means the membership of feature $x_i \in X_i$ in this instance. Likewise, degrees of truth should be generalized where $\sum_{c \in C} \mu_c^e = 1$ and $\sum_{x_i \in X_i} \mu_{x_i}^e = 1$.

A common model to manage continuous parameters in NB classifier to apply the Gaussian distributions to represent show the likelihoods of features acquired from the classes [21]. Hence, every attribute are illustrated by a Gaussian probability density function (PDF) as given below.

$$X_i \sim N(\mu, \sigma^2) \quad (22)$$

The Gaussian PDF is a bell-like structure which is demonstrated by the given function:

$$N(\mu, \sigma^2)(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (23)$$

where μ implies the mean and σ^2denotes the variance. In NB, the variables required are order of (nk), where n shows the count of features and $k$signifies count of classes. The major objective essential is a normal distribution $P(Xi|C) \sim N(\mu, \sigma^2)$ for continuous attributes. Therefore, parameters of normal distributions are accomplished by,

$$\mu_{X_i|C=c} = \frac{1}{N_c} \sum_{i=1}^{N_c} x_i \quad (24)$$

$$\sigma^2_{X_i|C=c} = \frac{1}{N_c} \sum_{i=1}^{N_c} \mathrm{x_i^2} - \mu^2 \quad (25)$$

where $N_c$means the count of instances in which $C = c$ and Nrefers the count of overall instances applied in training. The measurement of $P(C = c)$ for classes using relative frequencies as depicted in the below:

$$P(C = c) = \frac{N_c}{N} \quad (26)$$

Random Forest

Usually, RF is defined as a combination model where the predicted outcomes are considered as discrete value which is so-called as RF classification, and in case of continuous value, it is assumed as RF regression. The empirical works assured that RF model contains maximum prediction accuracy with optimal tolerance for noisy value. Next, RF classifier is operated in 2 stages. Initially, RF scheme filters subsamples from actual instances under the application of bootstrap re-sampling mechanism and develops Decision Trees (DTs) for every sample. Secondly, DT is classified and executed a simple vote with higher votes of classification as final outcomes. Fig. 3 illustrates the flowchart of RF classifier.

Also, it is operated on 3 phases:

(1) Choose a training set. Apply the bootstrap random sampling model for retrieving K training sets from actual dataset (M properties), with a size of training set is equal to actual training set.

(2) Develop an RF model. Deploy a classification-regression tree for bootstrap training sets in generating K DTs to make a "forest"; however, these trees non-pruned. Considering the development of a tree, this method does not select optimal features as interior nodes for branches; however, the branching operation is described as random selection of m⩽M features.

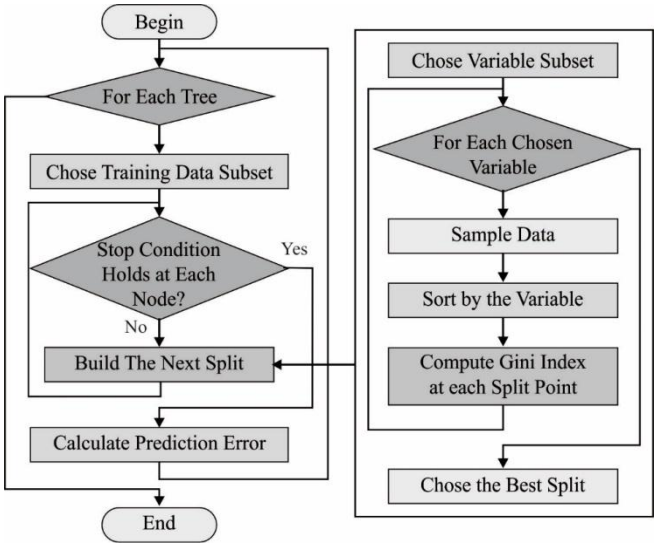


Fig. 9. Flowchart of RF classifier

(3) Develop simple voting. As the training process of DT is autonomous, training of the RF forests is computed in parallel fashion that enhances the efficiency. RF is developed by the combination of K DTs. While the input samples are classified, the final outcomes are related on simple voting of resultant DT. Moreover, it computes the instances by developing a sequence of autonomous and shared DTs for accomplishing the final class of samples.

2. EXPERIMENTAL VALIDATION

The performance of the proposed model is validated using a PC with i5-8600k processor, GeForce 1050Ti 4GB, 16GB RAM, 250GB SSD, and 1TB HDD. The simulation tool used is Python - 3.6.5 with different python packages namely TensorFlow (GPU-CUDA Enabled), keras, numpy, pickle, matplotlib, sklearn, pillow, and opencv-python. The dataset involved, measures, and the results are discussed in the subsequent sections.

Dataset used

For experimentation, two benchmark datasets namely brain MRI images [22] and ICH dataset [23] are used. The former dataset has a set of 147 images under tumor class and 341 images under hemorrhage class.

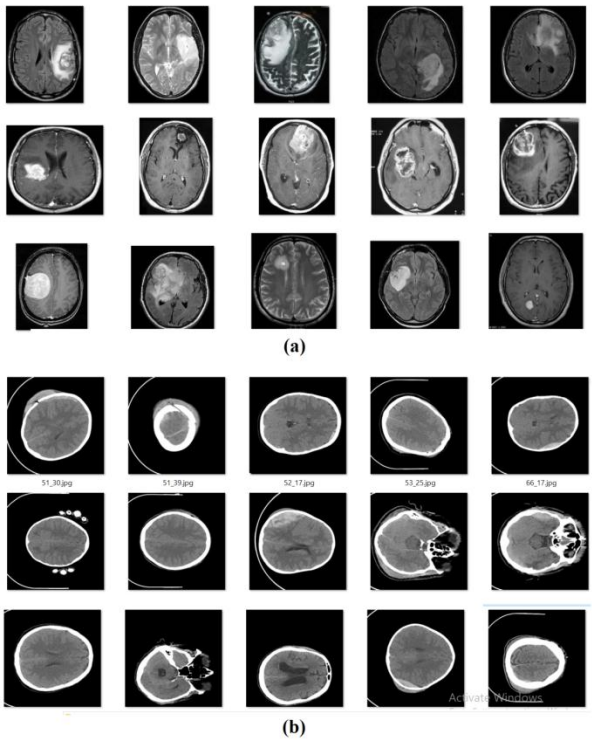


Fig. 10. Sample Images a) Tumor b) Hemorrhages

The tumor image size varies between 192*192 and 630*630. Besides, the hemorrhage image size is 512*512 pixels. The second dataset has CT scans of 75 subjects in NIfTI format. Some sample images from two datasets are illustrated in Fig. 4.

Performance Measures

The measures used to investigate the classifier results analysis of the proposed model are defined as follows.

Sensitivity: It determines the proportion of positive samples correctly classified.

$$\text{Sensitivity} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \quad (27)$$

Specificity: It evaluates the proportion of negative samples correctly classified.

$$\text{Specificity} = \frac{\text{True Negative}}{\text{True Negative} + \text{False Positive}} \quad (28)$$

Accuracy: It measures the proportion of correctly classified samples (positives and negatives) beside the entire samples (count of samples that have been classified).

$$\text{Accuracy} = \frac{\text{True Positive} + \text{True Negative}}{\text{True Positive} + \text{True Negative} + \text{False Positive} + \text{False Negative}} \quad (29)$$

Precision: It computes the count of true positives divided by the count of true positives plus the count of false positives

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \quad (30)$$

3. RESULTS AND DISCUSSION

Fig. 5 visualizes the qualitative results attained by the proposed model on the applied BT and hemorrhage images. Fig. 5a shows the outcome of the BT classification process where the first and second rows indicate the input and pre-processed tumor images. Similarly, Fig. 11b displays the sample results of the original and hemorrhage images. Table1. Displays the complete process as column (a) and (b)shows the classification process of a tumor image where as column(c) and (d) shows the classification of intracranial haemorrhage from the acquired input.

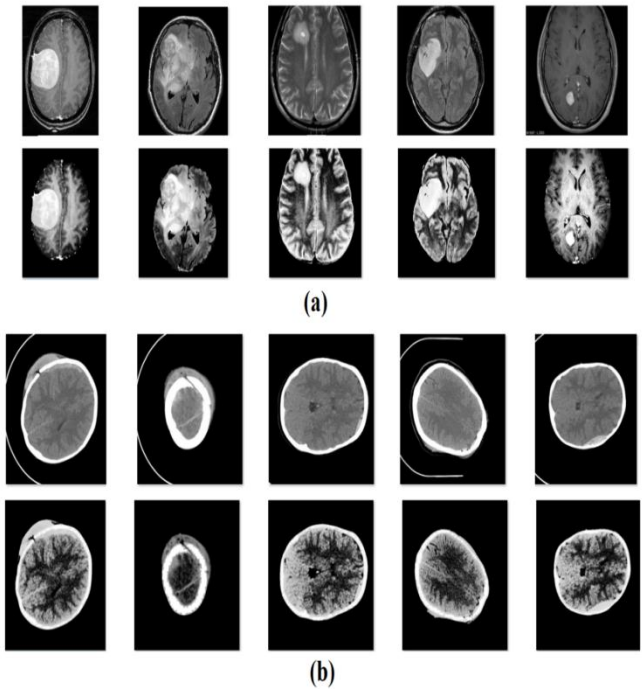


Fig. 11. a) First Row Tumor Original Images / Second Row Tumor Preprocessed Images b) First Row Hemorrhage Original Images b) Second Row Hemorrhage Preprocessed Images

Fig. 6 demonstrates the confusion matrices generated by the different proposed models on the classification of BT and ICH. Fig. 6a depicts that the SIFT-GNB model has properly classified a set of 303 images under Tumor class and 96 images under haemorrhage class. In addition, the SIFT-RF model has resulted to effective classification with the maximum of 338 images under Tumor class and 108 images under haemorrhage class. Moreover, the ANT-GNB model has proficiently classified a total of 341 images under tumor class and 112 images under haemorrhage class. Furthermore, the ANT-RF model has appropriately classified a total of 341 images under Tumor class and 119 images under haemorrhage class.

Confusion Matrix				
Feature extraction methods	**Haemorrhage/Tumor**			
	TP	**TN**	**FP**	**FN**
AlexNet-GNB	447	34	7	0
AlexNet-RF	460	1	0	27
SIFT-GNB	399	89	0	0
SIFT-RF	425	59	2	2

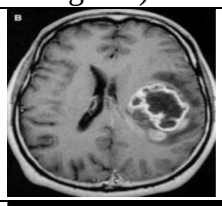
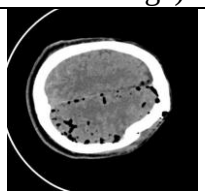
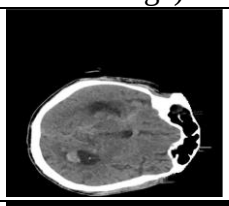
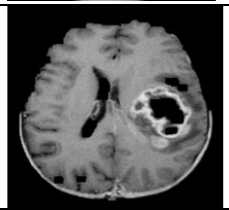
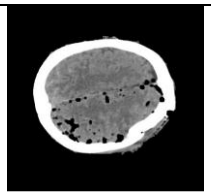
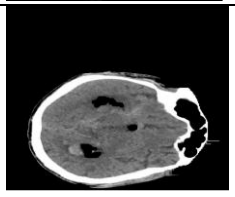
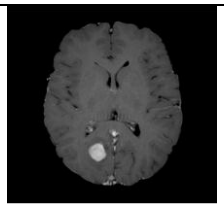
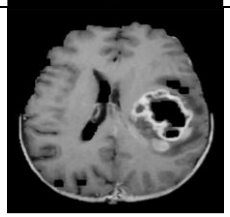
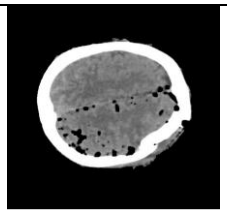
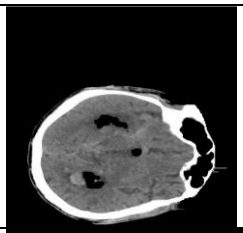
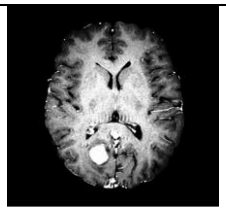
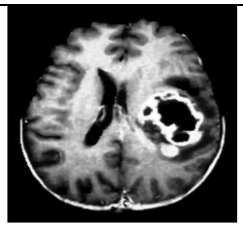
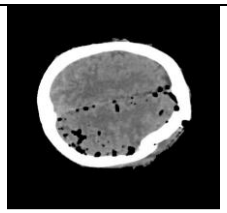
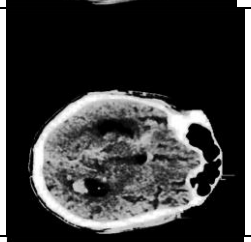
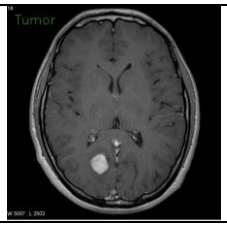
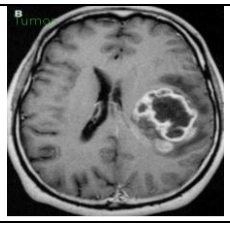
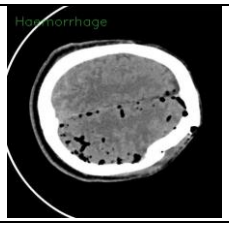
Fig. 12. Confusion Matrix for SIFT-GNB, SIFT-RF, ANT-GNB,ANT-RF

Table 2 and Figs. 7-8 displays the classification outcome of the proposed models with respect to distinct measures. From the obtained values, it is evident that the SIFT-GNB model has obtained a minimum sensitivity of 85.59%, specificity of 71.64%, accuracy of 81.76%, precision of 88.85%, and F-score of 87.19%. At the same time, the SIFT-RF model has resulted to slightly better performance over the SIFT-GNB model with the sensitivity of 89.66%, specificity of 97.29%, accuracy of 91.39%, precision of 99.12%, and F-score of 94.15%.

Fig. 12. Shows the confusion matrix for the predicted class as against the actual class, revealing the TP,TN,FP and FN values obtained using Alexnet and GNB in (a),values obtained using alexnet and RF in (b), values obtained using SIFT and GNB in (C) and values obtained using SIFT and RF in (d).

Table 1: Experimental results (a) Proess carried out (b) Image1-tumor(Benign),(c)Image3-ICH(epidural haemorrhage)(d)IVH-intra-venticular haemorrhage)

(a) (b) (c) (d) €

Process	Image1(Tumor-Benign)	Image2(Tumor-Malignant)	Image3(Intracranial haemorrhage)	Image4(Intracranial haemorrhage)
I/p image				
Skull stripped				
Bilateral Filtering				
CLAHE				
SIFT-GNB				

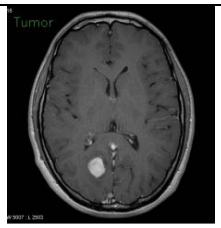
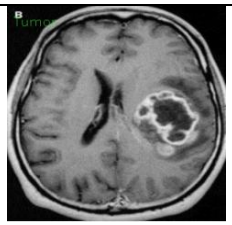
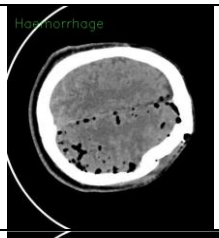
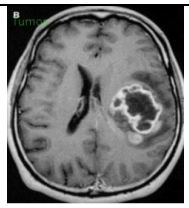
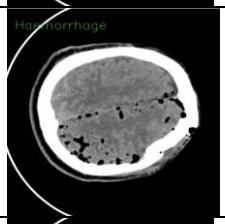
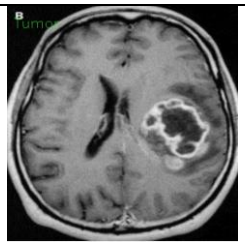
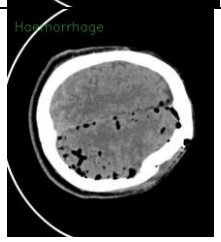
SIFT-RF				
DLIM-GNB				
DLIM-RF				

Table 2 Result Analysis of Proposed Methods interms of Sensitivity, Specificity, Accuracy, Precision, and F-score

Methods	**Sensitivity**	**Specificity**	**Accuracy**	**Precision**	**F-score**
DLAN-RF	92.41	100	94.26	100	96.05
DLAN-GNB	90.69	100	92.83	100	95.12
SIFT-RF	89.66	97.29	91.39	99.12	94.15
SIFT-GNB	85.59	71.64	81.76	88.85	87.19

Table 3 Result Analysis of Existing with Proposed Methods interms of Sensitivity, Specificity, Accuracy, Precision, and F-score

Methods	**Sensitivity**	**Specificity**	**Accuracy**	**Precision**	**F-score**
DLAN-RF	92.41	100	94.26	100	96.05
CNN-VGG16	81.25	88.46	89.66	84.48	85.25
CART	88.00	80.00	84.00	-	-
RF	90.00	80.00	88.00	-	-
k-NN	80.00	80.00	80.00	-	-
Linear SVM	91.20	80.00	88.00	-	-
WEM-DCNN	83.33	97.48	88.35	89.90	-
CNN	87.06	88.18	87.56	87.98	-
SVM	76.38	79.41	77.32	77.53	-

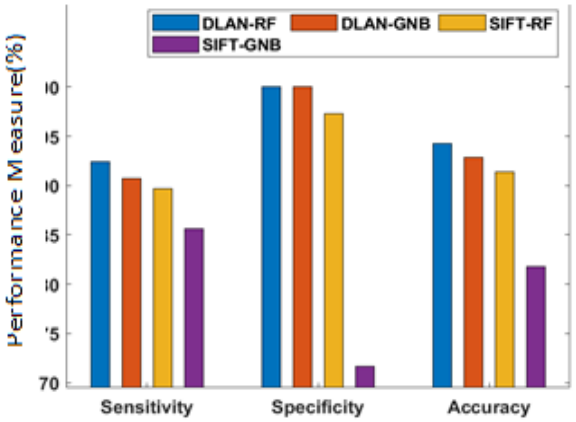


Fig. 13. Result analysis of DLAN-RF model interms of sensitivity, specificity, and accuracy

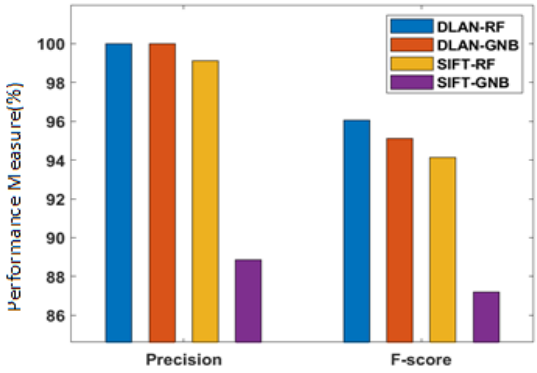


Fig. 14. Result analysis of DLAN-RF model interms of precision and F-score

Followed by, competitive performance is showcased by the DLAN-GNB model with the sensitivity of 90.69%, specificity of 100%, accuracy of 92.83%, precision of 100%, and F-score of 95.12%. However, the DLAN-RF model has outperformed the other three proposed models with the maximum sensitivity of 85.59%, specificity of 71.64%, accuracy of 81.76%, precision of 88.85%, and F-score of 87.19%.

Table 3 and Figs. 9-10 showcase the comparative results analysis of the DLAN-RF model with existing models [24-26] interms of different measures.

Fig. 15 investigates the classifier results analysis of the DLAN-RF model interms of sensitivity, specificity, and accuracy. The experimental results indicated that the SVM model achieves worse performance by obtaining a least sensitivity of 76.38%, specificity of 79.41%, and accuracy of 77.32%. In addition, the KNN model has achieved a slightly higher equivalent sensitivity, specificity, and accuracy of 80%. Along with that, the CART model has resulted to an even higher sensitivity of 88%, specificity of 80%, and accuracy of 84%. At the same time, the CNN model has tried to achieve moderate outcome with the sensitivity of 87.06%, specificity of 88.18%, and accuracy of 87.56%. Simultaneously, the RF model has achieved slightly manageable outcome with the sensitivity of 90%, specificity of 80%, and accuracy of 88%. Eventually, the linear SVM has exhibited somewhat satisfactory results with the sensitivity of 91.2%, specificity of 80%, and accuracy of 88%. Concurrently, the WEM-DCNN model has achieved reasonable outcome with the sensitivity of 83.33%, specificity of 97.48%, and accuracy of 88.35%. Though the CNN-VGG16 model has exhibited competitive outcome with the sensitivity of 81.25%, specificity of 88.46%, and accuracy of 89.66%, it failed to outperform the proposed DLAN-RF model which has obtained a maximum sensitivity of 92.41%, specificity of 100%, and accuracy of 94.26%.

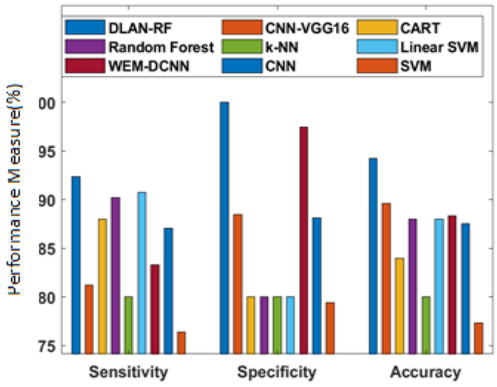


Fig. 15. Comparative analysis of DLAN-RF model

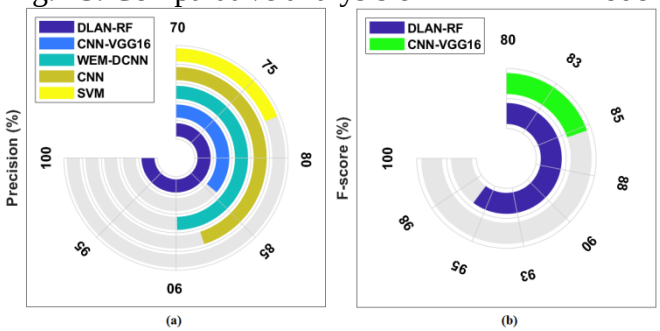


Fig. 16. Comparative analysis of DLAN-RF model interms of Precision and F-score

Fig. 16 examines the classifier outcomes analysis of the DLAN-RF (Deep learning AlexNet-Random Forest) method with respect to precision and F-score. The experimental outcomes exhibited that the SVM model attains worse performance by achieving a worst precision of 77.53%. Additionally, the CNN-VGG 16 model has reached a somewhat higher precision of 84.48% and F-score of 85.25%. Simultaneously, the CNN method has tried to attain moderate outcome with the precision of 87.98%. Concurrently, the WEM-DCNN model has attained slightly manageable outcome with the precision of 89.9%. However, the presented DLAN-RF model has reached a highest precision of 100% and F-score of 96.05%.
From the above mentioned tables and figures, it is obvious that the DLAN-RF model is found to be superior to other proposed and existing methods on the diagnosis of BT and ICH. The proposed DLAN-RF model has obtained a maximum sensitivity of 92.41%, specificity of 100%, and accuracy of 94.26. These values portrayed that it can be applied as a proper tool for medical diagnosis.

4. CONCLUSION

This paper has developed a novel DL based BT and ICH diagnosis model. The input image is initially preprocessed in three levels to enhance the image quality. Then, a set of handcrafted and deep features are extracted by the use of SIFT and AlexNet model. The integration of the handcrafted and deep features takes place to enhance the classifier results. At last, the extracted feature vectors are fed as input to the GNB and RF to identify the different set of class labels present in the image. In order to assess the classifier results analysis of the proposed model, an extensive experimental analysis is carried out to ensure its supremacy. The experimental results verified the effective diagnostic performance of the proposed model with the maximum sensitivity of 92.41%, specificity of 100%, and accuracy of 94.26%. In future, the diagnostic performance can be improved by the use of advanced DL models instead of AlexNet model.

5. REFERENCES

[1] McGuire, S. (2016). Health Organization, International Agency for Research on Cancer, World Cancer Report 2014. Geneva, Switzerland. Advances in Nutrition, 7, 418–419.
[2] H. Zuo, H. Fan, E. Blasch, H. Ling, Combining convolutional and recurrent neural networks for human skin detection, IEEE Signal Proc. Let. 24 (3) (2017) 289-293.
[3] O. Charron, A. Lallement, D. Jarnet, V. Noblet, J.B. Clavier, P. Meyer, Automatic detection and segmentation of brain metastases on multimodal MR images with a deep convolutional neural network, Comput. Biol. Med. 95 (2018) 43-54.
[4] L. Zhou, Z. Zhang, Y.C. Chen, Z.Y. Zhao, X.D. Yin, H.B. Jiang, A deep learning-based radiomics model for differentiating benign and malignant renal tumors, Transl. Oncol. 12 (2) (2019) 292-300.
[5] Latchoumi, T. P., Sunitha, R. Multi-agent systems in distributed datawarehousing. In 2010 International Conference on Computer and Communication Technology (ICCCT), (2010) 442-447, IEEE.
[6] S. Hussein, P. Kandel, C.W. Bolan, M.B. Wallace, U. Bagci, Lung and pancreatic tumor characterization in the deep learning era: novel supervised and unsupervised learning approaches, IEEE Trans. Med. Imaging (2019). https://doi.org/10.1109/TMI.2019.2894349
[7] Y. Yang, L.F. Yan, X. Zhang, Y. Han, H.Y. Nan, Y.C. Hu, X.W. Ge, Glioma grading on conventional MR images: a deep learning study with transfer learning, Frontiers in Neuroscience 12 (2018).
[8] Latchoumi, T. P., Parthiban, L. Abnormality detection using weighed particle swarm optimization and smooth support vector machine. (2017).
[9] R. Jain, N. Jain, A. Aggarwal, D.J. Hemanth, Convolutional neural network based Alzheimer's disease classification from magnetic resonance brain images, Cogn. Syst. Res. (2019). https://doi.org/10.1016/j.cogsys.2018.12.015
[10] Prevedello LM, Erdal BS, Ryu JL et al (2017) Automated critical test findings identification and online notification system using artificial intelligence in imaging. Radiology 285:923–931
[11] Ranjeeth, S., Latchoumi, T. P., Paul, P. V. A Survey on Predictive Models of Learning Analytics. Procedia Computer Science, (2020) 167, 37-46.
[12] Chang P, Kuoy E, Grinband J et al (2018) Hybrid 3D/2D convolutional neural network for hemorrhage evaluation on head CT. AJNR Am J Neuroradiol 39(9):1609–1616
[13] Chilamkurthy S, Ghosh R, Tanamala S et al (2018) Development and validation of deep learning algorithms for detection of critical findings in head CT scans. arXiv preprint arXiv:1803.05854
[14] Arbabshirani MR, Fornwalt BK, Mongelluzzo GJ et al (2018) Advanced machine learning in action: identification of intracranial hemorrhage on computed tomography scans of the head with clinical workflow integration. npj Digit Med 1:9
[15] Latchoumi, T. P., Balamurugan, K., Dinesh, K., Ezhilarasi, T. P. Particle swarm optimization approach for waterjet cavitation peening. Measurement, (2019) 141, 184-189.
[16] Anoop, V. and Bipin, P.R., 2019. Medical Image Enhancement by a Bilateral Filter Using Optimization Technique. Journal of Medical Systems, 43(8), p.240.
[17] Wu, S., Zhu, Q., Yang, Y. and Xie, Y., 2013, August. Feature and contrast enhancement of mammographic image based on multiscale analysis and morphology. In 2013 IEEE International Conference on Information and Automation (ICIA) (pp. 521-526). IEEE.

[18] Rajkumar, R. and Singh, K.M., 2015, September. Digital image forgery detection using SIFT feature. In 2015 International Symposium on Advanced Computing and Communication (ISACC) (pp. 186-191). IEEE.
[19] Lu, S., Lu, Z. and Zhang, Y.D., 2019. Pathological brain detection based on AlexNet and transfer learning. Journal of computational science, 30, pp.41-47.
[20] Loganathan, J., Latchoumi, T. P., Janakiraman, S., parthiban, L. A novel multi-criteria channel decision in co-operative cognitive radio network using E-TOPSIS. In Proceedings of the International Conference on Informatics and Analytics (2016) 1-6.
[21] Bustamante, C., Garrido, L. and Soto, R., 2006, November. Comparing fuzzy naive bayes and gaussian naive bayes for decision making in robocup 3d. In Mexican International Conference on Artificial Intelligence (pp. 237-247). Springer, Berlin, Heidelberg.
[22] https://www.kaggle.com/navoneel/brain-mri-images-for-brain-tumor-detection
[23] https://physionet.org/content/ct-ich/1.3.1/
[24] Latchoumi, T. P., Ezhilarasi, T. P., & Balamurugan, K. (2019). Bio-inspired weighed quantum particle swarm optimization and smooth support vector machine ensembles for identification of abnormalities in medical data. SN Applied Sciences, 1(10), 1137.
[25] Gupta, T., Gandhi, T.K., Gupta, R.K. and Panigrahi, B.K., 2017. Classification of patients with tumor using MR FLAIR images. Pattern Recognition Letters.
[26] Karki, M., Cho, J., Lee, E., Hahm, M.H., Yoon, S.Y., Kim, M., Ahn, J.Y., Son, J., Park, S.H., Kim, K.H. and Park, S., 2020. CT window trainable neural network for improving intracranial hemorrhage detection by combining multiple settings. Artificial Intelligence in Medicine, p.101850.
0