International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 23s, 2025
https://theaspd.com/index.php

Network Pruning Assisted Dense Convolution Knowledge Distillation Transformer for Brain Cancer Classification with Effective Tumor Identification Approach

Sangeeta Jawar[1], Shikha Tiwari[2], Ramgopal Kashyap[3]
[1]Research Scholar Amity University Chhattisgarh, Raipur, India
[2]Assistant Professor, Amity University Chhattisgarh, Raipur, India
[3]Professor, [3]Department of Information Technology Guru Ghasidas Vishwavidyalaya, Bilaspur, India
[1]sangeetajawar8@gmail.com, [2]stiwari@rpr.amity.edu and [3]ram1kashyap@gmail.com

ABSTRACT
The survival rate of patient's has been reduced if brain tumor types are misdiagnosed, which can stop an efficient response to medical intervention. In traditional method, brain tumor have been differentiated by inspecting the MRI images of the patient's brain. However in existing with large amount of data and different type of tumor identification consume more time and prone to human errors. The introduction of attention-based mechanism in medical imaging have been shown an accurate diagnostic, especially in identification and classification of brain tumor. In this study, an novel of hybrid knowledge distillation transformer model with improved U-net model is used for brain tumor detection. This paper is divided into four phases: pre-processing, identification, feature extraction and classification. Initially input images were pre-processed by using Trimmed pixel density based median filter (Pix-TrMed). From pre-processed image identify the tumor, which have been performed by Grouped dense residual convolution based U-Net model (Grp-DRcU-Net) model. After tumor identification, extract the important feature from the identified tumor outcomes, which is performed by improved ResNet (Im-ResN) model. Based on these extraction, classify the various types of tumors such as necrotic tumor core (NCR), peritumoral edema (ED) and enhancing tumor (ET) . The classification of brain tumor is performed by Network pruning assisted tripartite attention based knowledge distillation in dense convolutional transformers (Netr-HDCViT). The experimental result of this study is conduct on two datasets, such as BRATS 2020 and 2021, the analysis done on this datasets has achieve better performances in term of accuracy at values of 98.83% and 98.56% respectively
Keywords: *Brain tumor, Trimmed pixel, Dense Residual, Tripartite Attention, Knowledge distillation, Pruning*

1. INTRODUCTION

The mortality rate of the brain cancer has been reduced by detecting brain tumors in the earlier stages using various imaging techniques like positron emission tomography (PET), computed tomography (CT). Among these images, magnetic resonance images (MRI) are used in the medical field for identification of brain tumor these imaging techniques. In recent years, advanced deep learning (DL) and machine learning (ML) algorithms are used for automatic detection of brain tumor, identification and classification [1, 2]. Brain tumors are usually divided into primary and secondary tumors. Primary tumors are originated from central nervous system and secondary tumors are originated from lungs, mammary glans and digestive tract [3]. For brain tumor segmentation, various algorithms are introduced such as U-Net, threshold based segmentation approach and so on. Classification of brain tumor has been performed by various algorithms like k-nearest neighboring (KNN), support vector machine (SVM), decision tree (DT), principal component analysis (PCA) and so on are used to enhance the performances [4]. For early detection and accurate diagnosis of brain tumor treatment, computer aided diagnosis (CAD) model has implemented medical image analysis precisely and rapidly with reproducible results [5].

ML algorithms are rapidly increasing and helpful tool for various prognosis and diagnosis of diseases, tissue segmentation and image classification. Convolutional neural networks (CNN) and autoencoder is mostly used for successful image processing technique [6]. For analysis, classification and detection process in healthcare industry, DL models such as ResNet, AlexNet, GoogleNet, VGG and so on were used for accurate prediction [7]. In order to extract high relevant features and making accurate predictions of brain tumor from input brain images using deep neural networks (DNN). This model has use millions of parameters, hierarchical model and learn from large database to solve classification problems [8]. MRI images are taken from three different directions such as sagittal, axial and coronal, which is mostly used in now-a-days for tumor detection in brain images. To overcome the issues of computational cost, very less model robustness by ResNet model [9, 10].

A novel DL ensemble model has been used to solve the issues like, important information were lost by traditional algorithm at the time of classification. Shallow CNN and VGG 16 models are designed on kera's frameworks, which are used to extract high-level information about tumors and helps to gathers deep features [11]. By using you only look once (YOLO) v7 algorithm are used for automatic brain tumor detection, which helps to reduce false detection and reduce the loss of human lives [12]. In various research, detection has been performed by various stages such as pre-processing, feature extraction, classification and so on. In pre-processing stage, the original images are blend into sharp images to enhance effects and these images are converted into gray scale images with resolution of 255 x 255 pixels [13]. A deep transfer learning CNN models were used for deep feature extraction from the input images and 3D incremental deep CNN model were widely used for automatic segmentation of brain tumor images [14, 15]. Various optimization algorithms like gray wolf, lizard, ant colony and so on were used for tuning purposes to reduce the loss function and model complexity.

1.1 Motivation and Problem statement

One of the deadliest and most prevalent types of disease in both young people and adults are brain tumors. Many deep learning models are developed to detect the brain tumors in MRI images but, a very few techniques are employed the segment the tumor regions. Also these approaches include some drawbacks in terms of reduced performance, misidentification, highly complex for training the data effectively, and limited features to train the model. In addition to that, some other existing limitations like high time consumption, less accuracy, less model robustness unable to extract long range features, global features and so on. By motivating this issues, introduce the novel method of hybrid knowledge distillation transformer model for brain tumor detection with effective identification approaches. In this research, novel model of tumor identification is Grp-DRcU-Net, here dense block has make the training easily and reduce the vanish gradient problem. However residual connection present in this novel model has make the feature mapping as easily and feature can be reusable. In traditional identification of tumor by using neural networks, which had require considerable of time for training. Meanwhile, in this study use advanced technique of Im-ResN model to extract the feature from the image. This model has require only minimal parameter, which has achieve good performance, in addition model has improve the quality of feature extraction and reduce the computational expensive. However batch size, learning rate and layers present in this extraction model has reduce the fitting issues. In conventional feature can be extracted by using CNN approach, but this baseline method reduce the evaluation of accuracy rate. The effectives of brain tumor classification have been proceed by advanced novel technique of Netr-HDCViT model. Here knowledge distillation has reduce the computational load and improve the accuracy over classification model. Additionally dense convolutional layer capture the long range dependencies, similarly the classification model is assist with tripartite attention, which is used within transformer block. However, these advanced attention mechanism, capture the local and global context effectively. In addition to that integrate the network pruning, this can make the classifier model as more efficiency and reduce the size of model as well as improve the performance. The major contribution of this research work is discussed in below,

- To identification tumor portion from the pre-processed brain MRI images using Grouped dense residual convolution based U-Net model (Grp-DRcU-Net).
- To extract essential features from identified images using improved ResNet (Im-ResN) model.
- To classify various types of brain tumor based on gathered features using Network pruning assisted tripartite attention based knowledge distillation in dense convolutional transformers (Netr-HDCViT), which helps to enhance model accuracy and robustness
- To reduce loss function network size of the classifier model using network pruning mechanism in the training phase of the classifier model.

The rest of the paper is organized as following: in section 2 explain the existing survey related to brain tumor classification, section 3 describes the proposed methodology of this research work, section 4 discuss the performance of the proposed work and section 5 conclude the research work with feature scope.

2. RELATED WORKS

Hossain et al. [16] suggested a transfer learning based multiclass classification model called IVX16 were used for accurate detection of brain tumor from MRI images. Here, the dataset contains totally 3264 MRI brain images for accurate classification, which were used for performances analysis like accuracy, precision, recall and f1-score. This suggested model was compared with various existing models like VGG

16, VGG 19, Inception, ResNet V2, ResNet 50 and Inception V3. To check the model validity by implementing local interpretable model agnostic explanation (LIME) analysis for this approach. This approach were deal with solving overfitting issues and reduce the complexity issues by dealing with complex pattern in images. This model didn't able to extract long range dependencies features from the input images was one of the major limitations in this model.
Aloraini et al. [17] suggested a hybrid transformer enhanced CNN (TECNN) model for brain tumor classification with effective attention mechanism. Here, local features were extracted by CNN model and global features were gathered by attention based transformer model. This approach has used two different datasets like BRATS 2018 and Figshare dataset, which contains various classes like glioma, high grade glioma, low grade glioma and meningioma. Very high weights model were used in this approach was one of the major limitations, which reduce the model superiority, efficiency and accuracy.
Liu et al. [18] suggested a hybrid CNN transformer network with semantic awareness for accurate brain tumor segmentation task. By integrating global and local features at the encoding stage by semantic mutual attention (SMA) for better effective performances. Here, the SMA was hybridized with Swin transformer model with benefits of depth-spatial separable convolution (DSConv). The performances such intersection of union (IoU) dicer score and so on were evaluated and for this approach, which was compared with various existing models like 3D U-Net, attention based U-Net model and so on.
Reddy, C et al. [19] suggested a fine tuned vision transformer model for brain tumor classification, which was compared with existing models like MobileNet v2, Efficient Net B0 and ResNet 50. Here, the dataset was collected from open source, which contains 7023 MRI images, which contains various classes like glioma, pituitary, meningioma and no tumor. Initially, image flipping, resizing, rotation, and normalization was performed in the pre-processing stages. At second stage, which was passed into pre-trained model for feature extraction and classifications.
Pacal et al. [20] suggested a hybrid shifted windows multi-head self-attention module (HSW-MSA with rescaled model used to enhance classification accuracy, simplified training complexity and reduced memory usage. Here, residual multi-layer perceptron (MLP) was used instead of traditional MLP in Swin transformer model to enhance model performances. The performances such as accuracy, precision, recall and f1-score were evaluated and compared with recent existing models to show the efficiency of this approach.
The segmentation of brain tumor was an essential for image analysis, fully convolutional neural networks (FCNN) model shown a standard 3D medical image segmentation. However due to limited kernel size in FCNN had degrade the performance of tumor segmentation. To overcome this issues, Hatamizadeh et al. [21] suggested the Swin UNET transformer model (SWIM UNETR) to segment the brain tumor. Here Swin transformer encoder extract features at five different resolution, also utilizing self-attention mechanism, which was connected to FCCN based decoder. However the performance of suggested segmentation model was illustrated by Brats 2021 dataset, moreover, the suggested model had demonstrated long range information and its limitation was segment a tumors with variable size.
In medical field, MRI brain tumor segmentation was essential for prediction and diagnosis. The difficulties in brain tumor segments was mainly because of shape, structure, frequency and their intensities, so researcher Aggarwal et al. [22] recently suggested the improved residual network (ResNet) for tumor segmentation. Here suggested model improve details present in connection links and improved the shortcut projection. However the suggested model improve the accuracy and precision, meanwhile speed up the learning process. The performance of suggested model analysis was evaluated by using BRATS 2020 dataset, and the model had occur improved more than 10% of accuracy, recall and F1-score over other existing methods like FCNN, CNN and so on. The limitation exhibits in this ResNet model was require more time to process and some other gradient diffusion become difficult to train.
In tumour diagnosis, segmentation was an important task, for treatment research in this area had detecting the location of tumor in early stage. Aboussaleh et al. [23] suggested the improved U-net model called inception U-net. This suggested model employed the inception block, and skip connection of U-Net had used the bi-directional feature pyramid neural (Bi-FPN) network for segmentation. However the suggested model was compared with several imaging segmentation and their performance was evaluated by publically available BraTS dataset. Similarly the suggested model employee promising result in term of accuracy at values of 87.8% respectively. Further limitation of this suggested model was increase the computational complexity.
The main requirement of brain tumor extraction was, segmentation and annotation of tumor boundaries

correctly. Gunasekara et al. [24] suggested the DL based chan-vese algorithm to detect the tumor boundaries for segmentation process. Initially introduced the convolutional neural network (CNN) classifier model, then implemented the region based convolutional neural network (R-CNN) model, which was performed on classified images and localize the tumor region. Finally the suggested model, concentrated to tumor boundary for segmentation process. Based on segmentation process, evaluate the overall performance in term of, PSNR and MAE, at values of 77.076 and 52.946 respectively. Compare to classical segmentation, bounder based segmentation is quite difficult and utilize more cost of effectiveness. Table 1 represent the comparative analysis of existing model

Author	Models	Merits	Demerits
Hossain et al. [16]	Multiclass classification model called IVX16	• Improve accuracy • Reduce saleable • More efficient	• Merging several model add more complexity
Aloraini et al. [17]	Hybrid transformer enhanced CNN (TECNN)	• Detecting at good • Robust • Training process were happen at end to end	• Sequential data possess limited effectiveness • Difficult to interpretable
Liu et al. [18]	Hybrid CNN transformer network	• More efficient • Improve the extraction of low level feature	• Prone to overfitting • Under extreme condition require limited performance
Reddy C et al. [19]	Vision transformer model	• More interpretable • Trained on large dataset • Efficient for segmentation	• Require large dataset to train effectively • Training become computationally expensive
Pacal et al. [20]	Hybrid shifted windows multi-head self-attention module (HSW-MSA	• Greater efficiency • Non overlapping of local window	• Limits in window size • Capturing of spatial and temporal become difficult.
Hatamizadeh et al. [21]	Swin UNET transformer model (SWIM UNETR)	• Achieve high predication in term of segmentation localization	• Global contextual and long term dependices become inefficient
Aggarwal et al. [22]	Improved residual network (ResNet)	• Reduce model complexity • Deeper network has train efficiently	• Heavy load of computation • Require more memory
Aboussaleh et al. [23]	Improved U-net model	• Better performance • Handle complex data	• Generalization issues • data become imbalance
Gunasekara et al. [24]	Chan-vese algorithm	• Flexible and powerful • Easy to implement	• Consume more time • Initial stage become more sensitive

In existing survey, various researcher had analysis the brain tumor segmentation by using various techniques. In existing, various limitations had noted such as model did not able to extract long range dependencies feature from the input image. However in existing model had reduce the efficiency, superiority and accuracy, meanwhile in existing segment the image with variable size. Moreover existing model had taken more time to process and gradient diffusion process were difficult to train. Furthermore limitations of this model was increase the computational complexity and so on. These all the overall limitations, noted from existing survey.

3. PROPOSED METHODOLOGY

In this research work a novel hybrid knowledge distillation transformer model with improved U-Net model is introduced for brain tumor detection. The work flow of this research work is represent in Figure 1

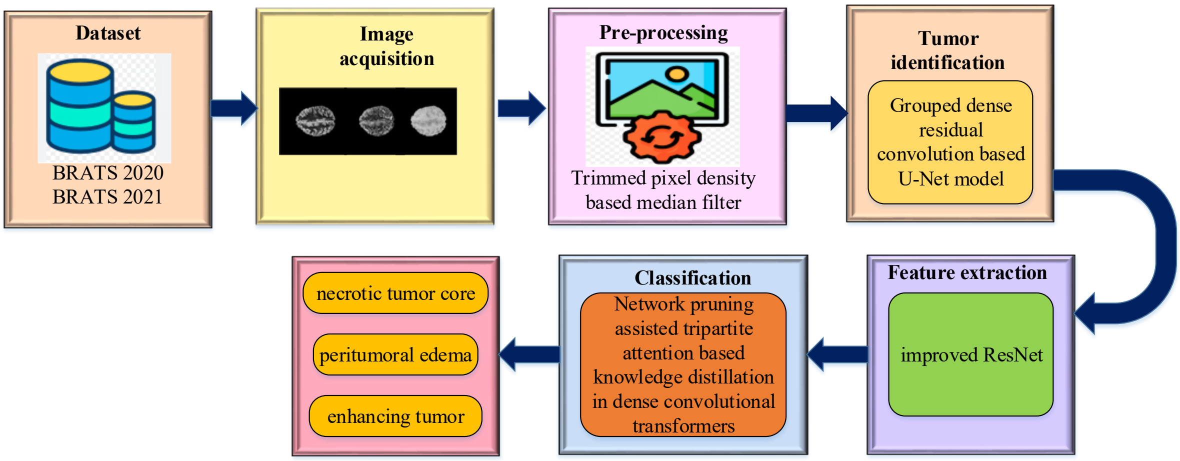


Figure 1: Overall architecture of proposed model

These architecture describes the work flow of research work, Initially, the input images are collected from open source datasets namely BRATS 2020 and 2021. Here, the input images are pre-processed by Pix-TrMed filter. From that pre-processed images, tumor identification has been performed by Grp-DRcU-Net .The essential features are extracted from the identified tumor outcomes using (Im-ResN model. Based on these selected features, various types of classes such necrotic tumor core (NCR), peritumoral edema (ED), enhancing tumor (ET) and normal class has been identified by Netr-HDCViT. These are the steps to follows for brain tumor classification and detection.

3.1 Pre-processing the image by using Trimmed pixel density based median filter

In this study for brain tumor classification, initially pre-processed the input image, which is collected from open source dataset. The images were present with some noises like speckle noise, salt and pepper noise, Gaussian noise and so on. These noises were removed by using bilateral filter, Gaussian filter and some other filters, but these filters had smoothen the input image properly. However, the images were shown the distortion on their edges and increase the intensity of value of pixel, which may loss the image sharpness. Similarly task may appear at edges can loss the important features, to tackle this issues, introduce Pix-TrMed filter [25]. Here Pix-TrMed filter woks at two stages, in first stage, filter tests the pixels and determine, whether it is decompose by salt and pepper noise or not. Next corrupted pixels were identified and this filter has check, if corrupted pixels is noisy or not. The maximum number of test pixel is 225 and it is present with window size of 3×3, so the filter has treat the 255 pixel at non-noisy one. For better understandable, the detail process of denoising is given in below algorithm. The steps to involves in Pix-TrMed are follow as

Step 1: Consider input noisy image $M := (M(i;j))$ and its pixel is $M(i.j)$

Step2: Restored the input image as $q(i,j)$

Step 3: Check and read the all i and j in input noisy image

Step 4: Select test pixel $M(i,j)$, if pixel is equal to zero or 255, it is said to be corrupted pixels and select the window size 3×3.

Step 4.1: If all the nine samples are in window size 3×3 of 0 and 225 only, then step 4.1 follow two cases are possible

Else Step 4.2.

Case 1: If atleast six samples in window mask are 0, then test pixel $M(i,j)$ is conserved as non-corrupted

pixels and current pixels is unaffected that is 0.
Case 2 : If atleast six samples in window mask are 255, then test pixel $M(i, j)$ is conserved as non-corrupted pixels and current pixels is unaffected that is 256.
Step 4.2: If suppose nine samples are selected in window mask are not 0 and 255, then following two cases are possible.
Case 1: If atleast one pixel in selected window size 3×3, has satisfy the condition of $0 < M(i, j) < 10$ or $245 < M(i, j) < 255$ then following steps are possible, else case 2
i. In selected window size maximum number of recurring pixels are identified
ii. The median values has calculated from those maximum repetitive pixel values
iii. Next obtained median values has replace the present processing pixel $M(i, j)$
Case 2: Remove every pixels, which having pixel values at 0 and 255. As per widow mask, next compute the median value for remaining element. Finally replace the current processing pixel $N(i, j)$ with computed median value.
Step 5: If $0 < N(i, j) < 255$, then test pixel is conserved as uncorrupted pixel and value of test pixel is kept as unaffected.
Finally, replaced or modified pixel has make as denoised pixel and this algorithm is suitable for entire pixel range of image, and its outcome become a denoised image.

3.2 Tumor identification by using Grouped dense residual convolution based U-Net model

After-preprocessing, the CNN model has identified the tumor, but this model had handle limited number of task. Based on overall analysis performance, the model had minimize its accuracy and increase the complexity [26]. In this study, identification model of Grp-DRcU-Net has effectively identified the tumor. Here initially used the grouped dense convolutional model, which has improve the structure of U-Net model with convolutional block attention module. However, these structure is added to the decoder module, which is focus towards the tumor feature in input image and extract the relevant feature. Meanwhile, upsampling is performed on decoder by the use of bilinear interpolation, which can improve the accuracy of tumor identification model. The Grp-DRcU-Net model contain four layer of encoder and decoder, in encoder grouped residual convolutional model is used for image downsampling with 64 filter. In between encoder and decoder, the skip connection used to introduce low-level information at appropriate scale into the decoder. In convolutional block attention mechanism module (CBAM) has employ low-level input data from the skip connection for additional feature learning, the decoder has combined with high-level feature data. Moreover, the image upsampled can be done by using bilinear interpolation technique, which has restored its original size. Figure 2 represent the tumor identification architecture

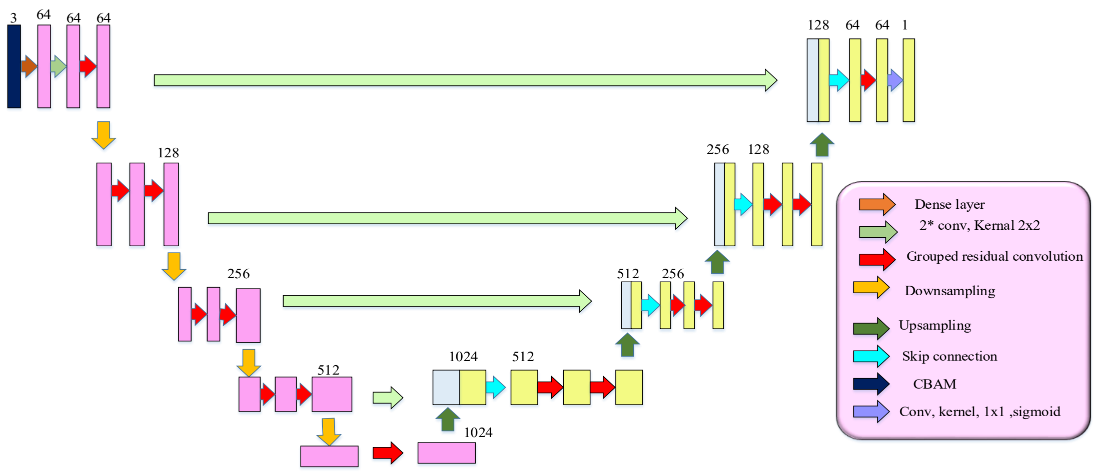


Figure 2: Architecture of grouped dense residual convolution based U-Net model
In CNN, the dense network layer has mapped the more number of feature and extracted the detailed feature. Basically, the high level features are at more representative and low level feature contain noise as well as raw information. While applying convolution operation at low level features, it helps to filters

more useful feature. Similarly to obtained effective feature information, implement the grouped convolutional. The grouped convolution consist of filter bank that convolves and separate the input feature at various depth. In between the filters has use the sparse relationship, which helps the grouped convolution to took redundant information and better features from respective filter part. Therefore, the model has enhance the feature extraction. Further, in grouped filtering, each group of filter had perform convolution calculation of image in their own group, and reduce its parameter and computation. Based on inspiration of ResNet model, integrate a residual connection to dense layer and indirectly learn the residual relationship between input and output layer. These improvement can enhanced the convergence rate of model and achieve higher identification accuracy.

3.3 Feature extraction by using improved ResNet model

In this section, extract the important feature from the identified tumor, an wide range of feature extraction technique were exist such as ResNet, CapsNet and CNN. These extraction technique has make the training process as complicated. Similarly, the use of advanced feature extraction model of Im-ResN has make the training process efficient [27]. The Im-ResN model has placed in first position to extract the feature, initially ResNet block is built with input parameter y and output target $G(y)$. This residual block has utilize the shortcut connection to learn input residual $H(y)=G(y)-x$ and make them us target output $[H(y)+y]$. However, due to many convolutional layer present in residual block has reduce the accuracy rate and degrade the performance. The two or more layer present in convolutional layer can be skipped with the help of skip connection and perform the identity mapping as directly. Due to learning some functions without reference (Y), residual has make as to create reference (Y) for each input layer's and learns to generate a residual function. The residual function can depend on number of network layer and make easier to optimize. The mapping function of residual is formulated in below equation

$$H = U_2\,\sigma(U_1 y) \qquad (1)$$

Here activation function of ReLU is denoted as σ, then y represent the output, which is obtained through second ReLU and shortcut connection. Figure 3: represent architecture of feature extraction.

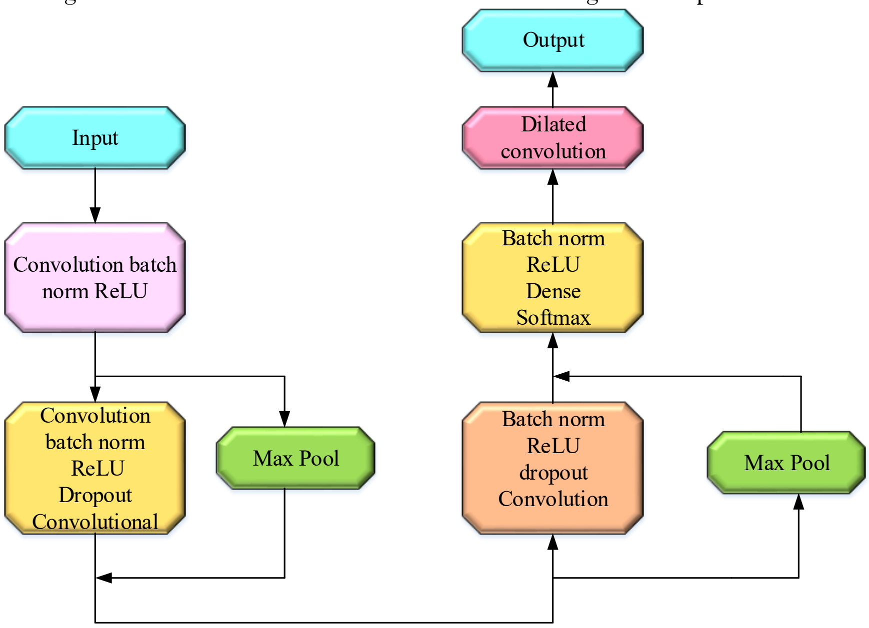


Figure 3: Architecture of Improved ResNet model

The ResNet 18 model is introduced here for feature extraction. Initially from input image extract the one-dimensional convolutional kernel with length of 32 similarly ResNet 18 model is used for two dimensional of feature extraction with convolutional of small kernel size of 3×3. Basically, the image resolution of direct input network is relatively low, but feature extraction has require large convolutional kernel to solve the problem. An ResNet 18 model is improved by adding a dilated convolutional layer. Generally Im-ResN model contain, four layers such as convolutional layer, classic ResNet-18 layer, fully connected layer (FC), and dilated convolutional layer. The initial stage of convolutional layer is act as basic feature extraction on input image. The second stage has use the classic ResNet-18 model which can act as best for tumor classification. In addition to that input layer are convoluted as double and ReLU, modified linear unit is added between two convolutions. However, it occur overfitting problem. On the other hand, before convolution of data are inputted into poling layer, which divides the input image into feature region and assign maximum value in that region, therefore region has reduced the number of parameter. The purpose of second stage is to avoid the optimization and make the models fast converge. In order to enhance the performance, use the Im-ResN model as the third part, here batch norm is added before the classical ResNet-18 structure to activate the training of network and maintain the model as stability. Similarly the features were sent to fourth stage of FC layer, here output features are mapped to FC layer at one-dimension of vector, and this vector is regressed by softmax function. The function of softmax is formulated in below equation

$$x = inputfeature \qquad (2)$$

$$P_i = soft\max\left(\lg_j\right) = \frac{e^{\lg_i}}{\sum_{i=0}^{n^{-1}} e^{\lg_i}} \qquad (3)$$

Here input feature is correspond to classification of tumor type image, x represent the thermal code and P_i denote the probability of tumor sample belongs to i^{th} value. However, the ith value of output vector logits is represent as $\lg_i$, which is used for probability of extraction result. Compare to existing model, the proposed of Im-ResN model make the extraction of image at shorter computational time as well as reduce the time of image training process.

3.4 Tumor classification Network pruning assisted tripartite attention based knowledge distillation in dense convolutional transformers (Netr-HDCViT)

The essential features has been extracted, and classify the tumor into various classes such as necrotic tumor core (NCR), peritumoral edema (ED), enhancing tumor (ET) and normal class has been classified by DL approaches such as CNN, RNN, DNN and many more. These technique cannot focus on main area of tumor also it cannot consider its relationship with the nearby tissues, because it show a crucial sign of tumor type. In this study, tumor have been classified by hybrid architecture of Netr-HDCViT model, here network pruning is used to reduce the size of parameter. However vision transformer assist in classifier model has reduce the complexity [28]. In DL, Knowledge distillation is one of the technique, here knowledge from complex model is referred as teacher model and transferred to simpler small model is known as student model. The function of knowledge distillation is to compress the knowledge of larger model into smaller one with loss of any performance. Before distilled knowledge, initially trained the teacher model, it is basically a large model, which is trained on respective dataset and achieve higher accuracy. The main goal behind the knowledge distillation is probability distribution of teacher model prediction and soft targets for student model. The output of teacher models are soften by applying a temperature to softmax function. The result of higher temperature in softer probabilities provide more information about different classes. Typically student model training has involves combination of two losses. The loss of distillation measures the difference between the softened output of teacher and student model. The original loss measures the variations between student prediction and original hard labels. After training, student model is the last one to be utilized in real-world application and it should assimilated a large portion of the teacher knowledge. Further resulting a model as smaller and quicker, but maintaining a comparatively high accuracy. Figure 4: represent the classification architecture

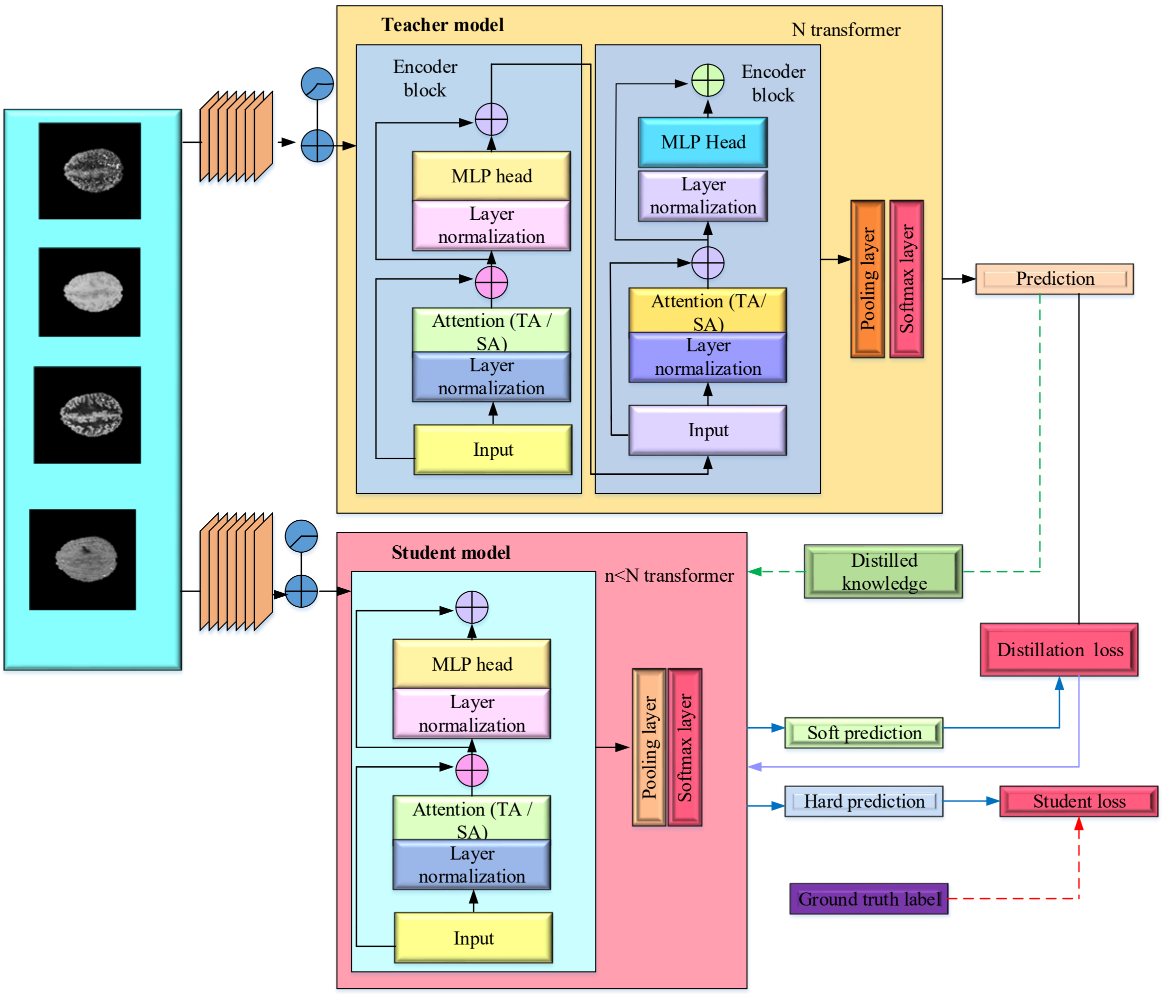


Figure 4: Network pruning assisted tripartite attention based knowledge distillation in dense convolutional transformers

In this research, knowledge distillation has employed transformer based attention for tumor classification. The process involved to be splitting the image into patches, then linearly embedded to them, and processing embedded patches. In teacher and student transformer, the images are split into patches and each patches are embedded into D-dimensional vector.

$$y_i = Embed\left(P_i\right) \qquad (4)$$

Here P_i represent the i patch and its embedding is denoted as y_i, the tumor classification based on distillation is computed as follows. In classifier, the standard softmax function is formulated in below equation.

$$soft\max\left(v\right)_i = \frac{e^{v_i}}{\sum_{j=1}^{c} e^{v_j}} \qquad (5)$$

Here v represent, logits vector produced on model and number of classes is denoted as c. In distilling knowledge, the modified temperature T is formulated in below equation.

$$Soft\max{}_T\left(v\right)_i = \frac{e^{v_i/T}}{\sum_{i=1}^{c} e^{v_i/T}} \qquad (6)$$

The standard softmax function $T = 1$ increases, the output of softmax become softer and uniform. Combining, the distillation loss and original loss has results the overall student loss utilized in knowledge distillation.

$$L = \beta L_{original} + \left(1-\alpha\right) L_{distillation} \qquad (7)$$

Here β specifies the hyperparameter, which weights the two losses, the original loss between, student prediction and true label is formulated by using standard cross-entropy.

$$L_{original} = \sum_{i=1}^{c} x_i \log\left(soft\max(s)_i\right) \qquad (8)$$

Here x represent true label of one-hot encoded, in student model, logits produced is mentioned as s. Based on computing formula, the $L_{distiilation}$ is the cross entropy between teacher and students softened prediction.

$$L_{distillation} = -\sum_{i=1}^{c} soft\max_T(t)_i \times \log\left(soft\max_T(s)_i\right) \qquad (9)$$

Here, logit produced by teacher model is denoted as t and student model is denoted as s. The temperature T has softened the both model. At training, teacher model is trained as isolation, then trained the student model and calculated the $L_{original}$ and $L_{distillation}$ then combined them by using L. In knowledge distillation, utilized four transformer block for teacher model, then each transformer block is well-found with two block of tripartite attention. However for student model utilized two block of transformer and each block is undergoing two cycle of tripartite attention. The tripartite attention is explained in below section.

3.4.1 Tripartite attention

The transformer block has adapted the tripartite attention (TA) for multiclass brain tumor classification. For detail process of tripartite attention, first understand the concept of self- attention mechanism. In transformer, self-attention mechanism has allow each position in the input sequence to attend every position in the preceding layer of sequence. This process concurrently done for every position is known as self-attention. Meanwhile, for multi-head attention, the process has done multi-heads to attend information from different subspace representation at different position. The patch of input image is linearly projected into three vectors such as queries (A), key (B) and values (C), similarly, this will happened for each head h using different linear projection.

$$A_h = YW_h^A, B_h = YW_h^B, C_h = YW_h^C \qquad (10)$$

Here Y is the input of self-attention layer, and the weight matrices of queries, keys and values are represented as W_h^A, W_h^B and W_h^C for head h. However, in each head, attention score is evaluated by using softmax of the dot product. The detailed operation is formulated in below equation.

$$Attention(A,B,C) = soft\max\left(\frac{AB^T}{\sqrt{d_b}}\right)C \qquad (11)$$

Here key vector dimension is denoted as d_b and each row of matrix has applied the softmax to create a probability distribution. However concatenated, the each head's output and once again linearly projected to obtained final value of multi-head attention output.

$$Multiheadattention(A,B,C) = concat(head_1,....,, head_h) \qquad (12)$$

Here each $head_h = Attention(A_h, B_h, C_h)$, often in transformer, each layer in attention has followed the residual connection and layer normalization. The tripartite attention follows three different attentions such as global attention, neighborhood attention and cross attention. In tripartite attention map has been computed by cross attention and token has been allowed from neighbourhood and global attention map sequences. The query sequence, A were embedding from neighborhood attention and context sequence (key and value) S were embedding from global attention. The sequence $A = AW^A, B = SW^B, C = SW^C$ are created by using weighted matrices sequence of W^A, W^B, and W^c, respectively.

$$Tripartite\ attent\left(A', B', C'\right) = soft\max(R)C' \qquad (13)$$

Here $S = \frac{AB^T}{\sqrt{d_b}}$ and B^T represent the matrix transpose of K. However, this way can tripartite attention model generate the output, and considered the relevant information from two different sequences. Moreover one sequence has operates the low contextual and other can operate as global contextual level, which can be leads for significant feature extraction of different brain tumors and classify its type accurately.

3.4.2 Transformer pruning

The architecture of transformer has noticed the decreasing of MSHA and MLP, so introduced the pruning technique, which can prune the dimension of linear projection by their related significant scores. In feature, $Y \in R^{n \times d}$, here denote the number of feature need to be pruned and dimension of the feature is represent as d. The main goal of pruning is to be preserve the important feature and remove the unwanted one from the linear projection. However optimal important score of $b^* \in \{0,1\}^d$, has generated the important features with their respective component of scores as 0 and 1. Based on significant scores, illustrated the pruned features

$$Y^* = Y\,diag(b^*) \qquad (14)$$

However, in b^* is difficult to optimize in neural network, so relax the b^* to real value as $\hat{b} \in R^d$. The soft pruned feature are obtained as

$$\hat{Y} = Y\,diag(\hat{b}) \qquad (15)$$

After training process, the transformer has obtained important score near to zero, then rank all the values with some importance score in transformer and obtained a threshold τ according to pre-defined purring rate. Based on threshold τ attained a discrete b^* by adjusting the values of threshold as below zero and higher values of one.

$$b^* = \hat{b} \geq \tau \qquad (17)$$

After pruning, the total pruned transformer has obtained with higher accuracy rate, the procedure of pruning is formulated in below equation

$$Y^* = \mathrm{Pr}une(Y). \qquad (18)$$

However, the pruning operation is applied to all MSHA and MLP block, the process of pruning is formulated in below equation.

$$A, B, C = FS_a'(prune(Y)), FS_b'(purne(Y)), FS_c'(prune(Y)) \qquad (19)$$

$$X = Y + FS_{out}'(prune(Attention(A, B, C))) \qquad (20)$$

$$Z = X + FS_2'(prune(FS_1'(prune(X)))) \qquad (21)$$

Here $FS_a', FS_b', FS_c', FS_{out}', FS_1', FS_2'$ are denoted as pruned linear projection which is corresponding to b^* and pruned feature. However these pruned network has reduce the complexity and loss function of network size also further improve the performance of training classifier.

4.1 RESULT AND DISCUSSION

In this section, the performance of the proposed model is evaluated by using two datasets namely such as BRATS 2020 and BRATS 2021. Meanwhile, the following section shows the detailed graphical representation of proposed as well as existing performance.

4.2 Dataset description

4.2.1 BRATS 2020 [30]

The BRATs 2020 dataset employed**,** MRI scans from multi-institutional and its primarily focus on segmentation of brain tumor. BRATS 2020 dataset mainly focus on prediction of patient overall survival rate based on classification of true tumor. The evaluation of BRATS 2020 is utilized on tumor segmentation. It contain three class of NCR, ED, and ET and sample of images to be pre-processed is 3100.

4.2.2 BRATS 2021[31]

The BRATS 2021 dataset utilized MRI scans and mainly focus on sub region segmentation of brain tumor, and its evaluation portion is depend on radiogenic classification of tumor. In task of segmentation, the class NCR is labeled as one, and ED and ET are labeled as two and three. However in this dataset contain totally 8,000 MRI scans from 2,000 patients and sample of images to be pre-processed is 1550.

4.3 Performance of BRATS 2020 dataset

In order to detect the performance of tumor classification, the proposed method is compared with several existing model such as AlexNet, MobileNet, Google Net, ResNet, CNN

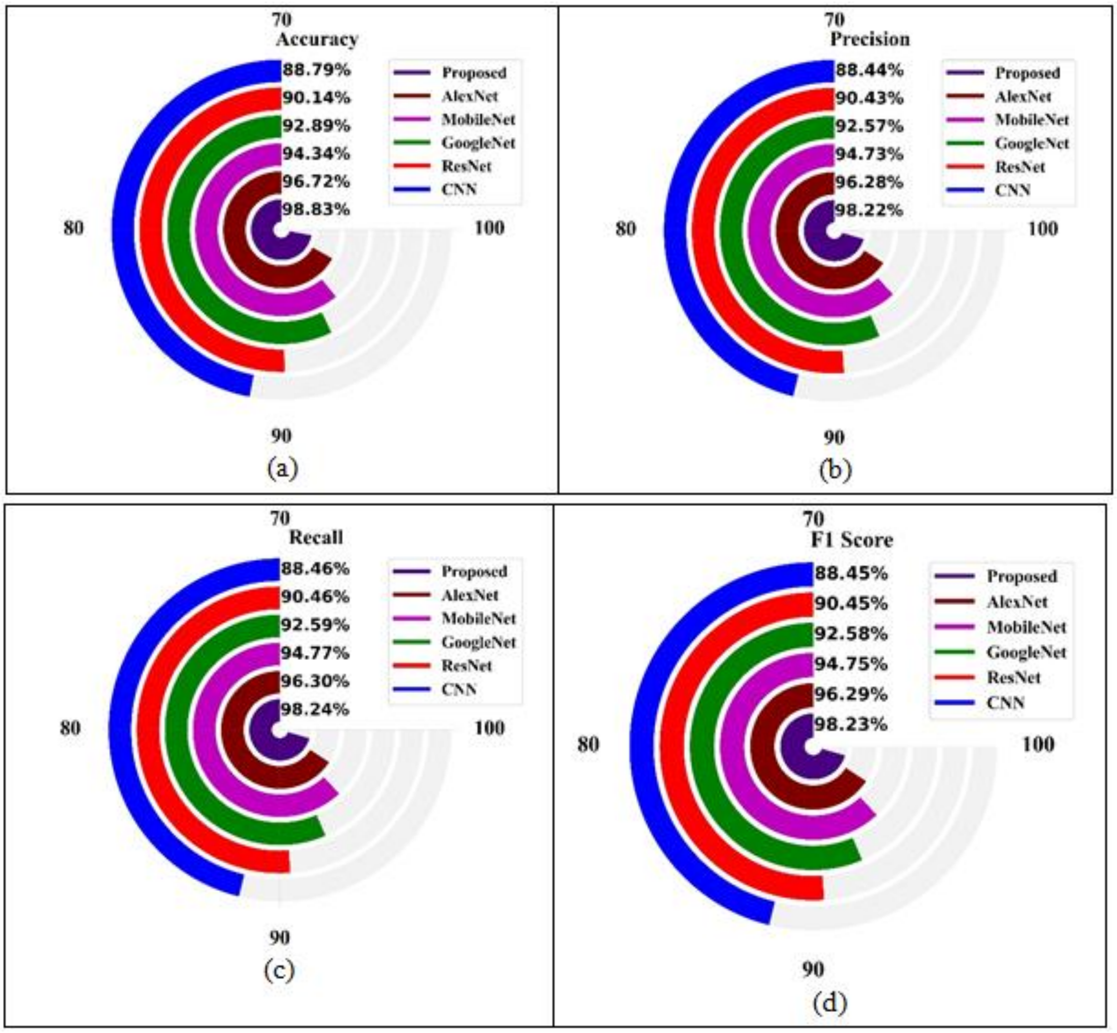
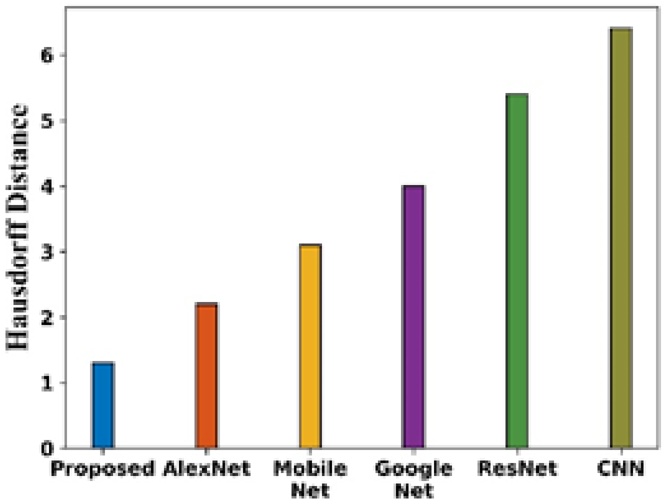
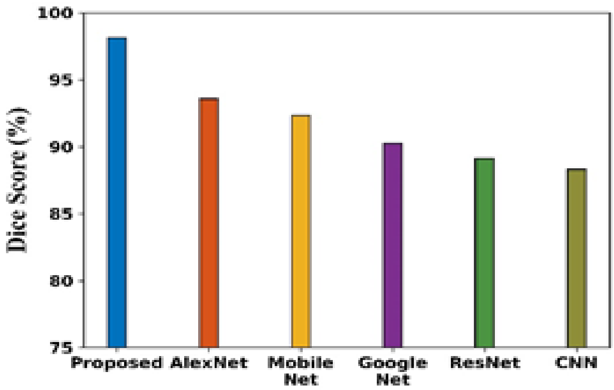


Figure 5 (a)-(b): Performance analysis of Accuracy, Precision, Recall and F1-score

Comparison of existing and proposed model of accuracy, precision, recall and F1-score is illustrated in Figure 5 (a)-(b). In proposed model, accuracy occur at values of 98.83 respectively, but existing model of Bi-LSTM and CNN had accuracy at low values of 90.14% and 88.79 % respectively. However, these existing model had shown the scalability issues and less efficiency of computational. The performance based on Precision and recall had achieved high values at 98.22 % and 98.24% respectively. In existing model of Ghost-Net and AlexNet very low performance, also these model had trainable only on small dataset. Meanwhile, based on comparative analysis, proposed has achieved high F1-score value at 98.23% respectively. The existing model of Dense Net had very low F1-score at values of 92.59%, however this existing model had perform at moderate size.



(a)



(b)

Figure 6 (a)-(b): Comparative analysis of Hausdroff distance and Dice score

Figure 6 (a)-(b) shown the performance analysis of existing and proposed model of Hausdroff distance and dice score. Here proposed model achieve high performance at values of 1.3031 and 2.2031 respectively. In existing model of AlexNet and CNN had very low performance dice score at values of 93.58 % and 88.30 % respectively. Similarly in existing of Ghost-net had evaluate the performance at very low range of 92.32% and 3.103 respectively. However in existing, those models were more sensitive to noise

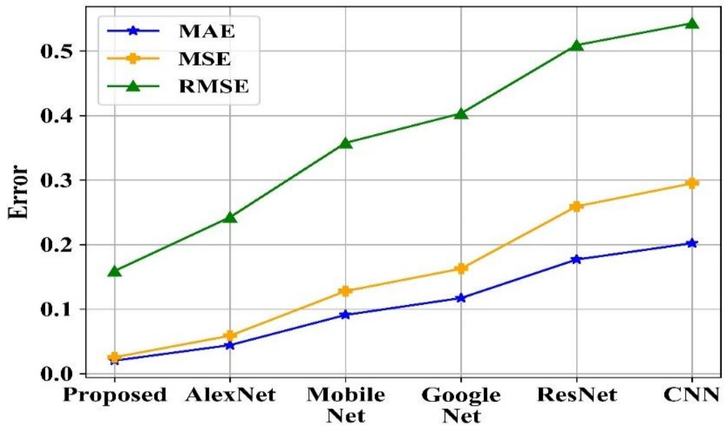


Figure 7: Evaluating error metrics performance of MAE, MSE and RMSE

The existing and proposed model of MAE, MSE and RMSE were depicts in Figure 7, based on comparative analysis, the proposed had lower error values. The error can be evaluated in term of MAE, MSE and RMSE metrics, and their values are 0.0202, 0.0253 and 0.1591 respectively. In existing models had achieved higher error values, due to occurrence of higher values, the models become attained a slower convergence.

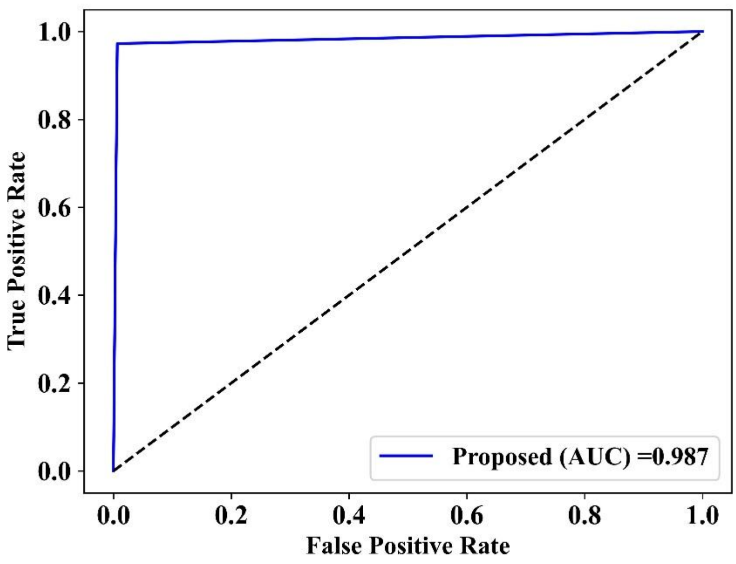


Figure 8: Performance of area under curve

International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 23s, 2025
https://theaspd.com/index.php

Figure 8 shown the graphical representation of area under curve (AUC). The AUC curve is plot with true positive rate and false positive rate at various classification threshold. The AUC signifies the overall performance of classifier. The proposed model exhibit performance at values of 0.987 respectively.

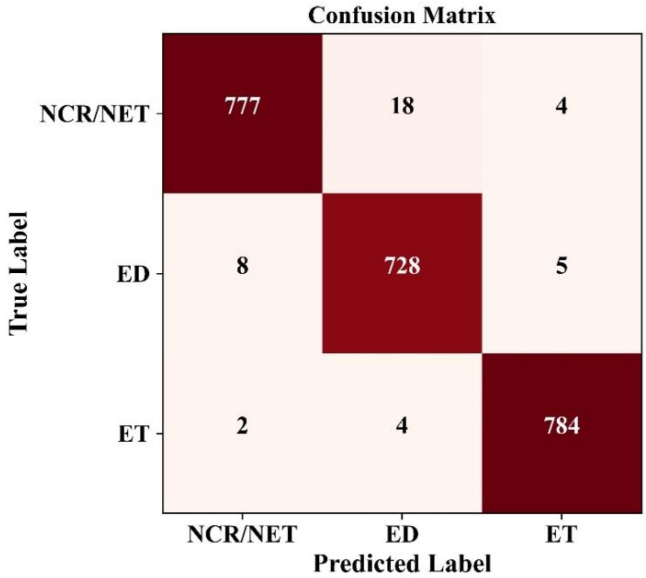
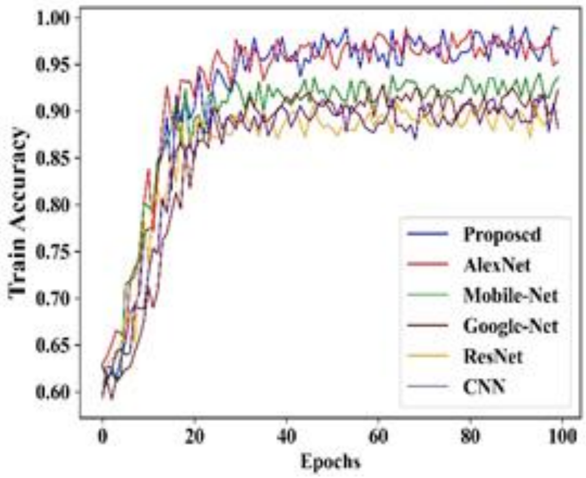
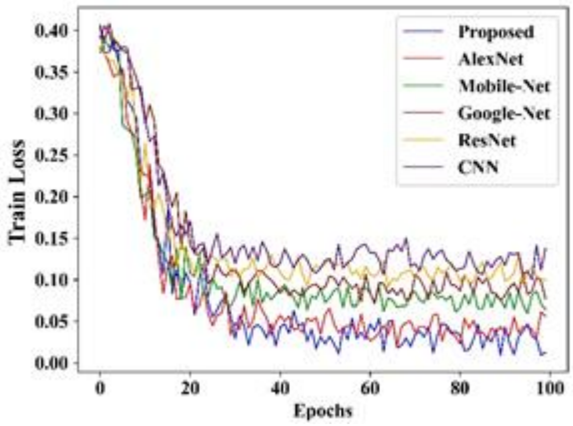


Figure 9: Confusion matrix

In BRATS 2020 dataset, prediction level of tumor is illustrated in Figure 9, which is explained in form of confusion matrix. Here class necrotic tumor core is correctly predicted as 777, perirtumoral edema (ED) is 728 and enhancing tumor (ET), predict 784 classes correctly.



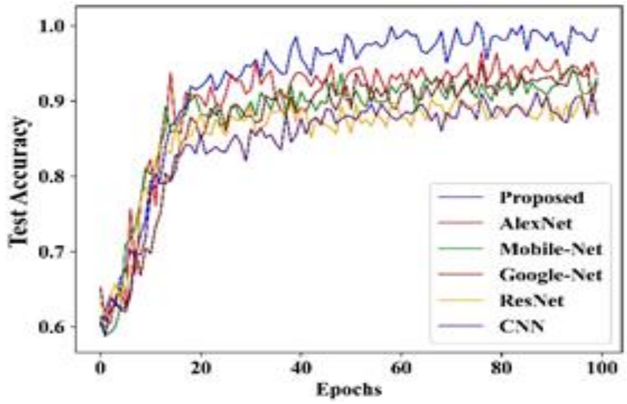
(a)



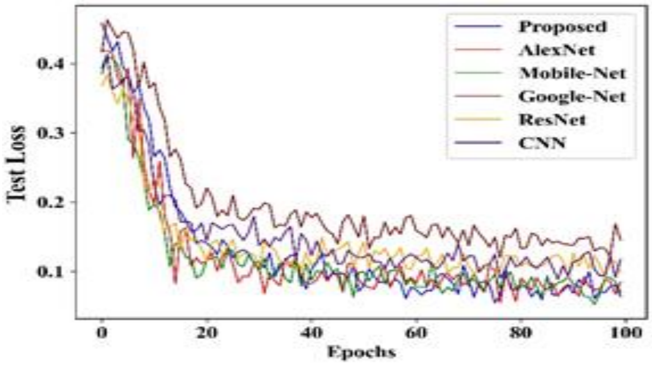
(b)

Figure 10 (a)-(b): Performance analysis of Training accuracy Loss curve

Figure 10 (a)-(b), shown the comparative analysis of existing and proposed model of training accuracy and loss curve. The evaluation of training accuracy is compared with existing and proposed model which is depicts in Figure 10 (a). In proposed range of training accuracy is high at value of 0.96 between epochs of 0 to 300, but existing had very low performance. However the comparison of existing and proposed model of training loss is illustrated in Figure 10 (b). In training loss, according to performance, the existing had very high losses, but proposed model achieved low loss values in the range of 0.04 between epochs of 0 to 300.



(a)

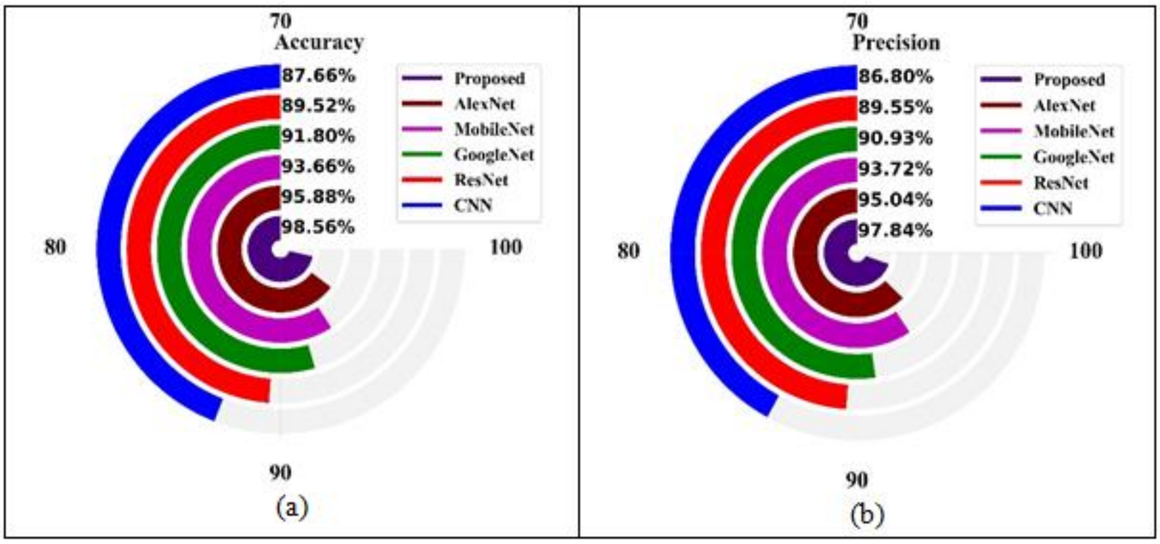


(b)

Figure 11 (a)-(b): Performance analysis of Testing accuracy and loss curve

The performance of existing and proposed model of Testing accuracy and loss curve is demonstrated in Figure 11 (a)-(b). Comparative analysis of testing accuracy with existing and proposed is depicts in Figure 11 (a). Based on analysis, proposed achieved high values at range of 0. 92 between respective epochs, but existing had achieved extremely poor performance. Similarly performance evaluation based on testing loss is compared with existing and proposed model, which shown in Figure 11 (b). Here proposed had testing loss is low at the range of 1.9, but existing had loss range is very high in the range of 0.39 between corresponding epochs of 0 to 300.

4.4 Performance of BRATS 2021 dataset



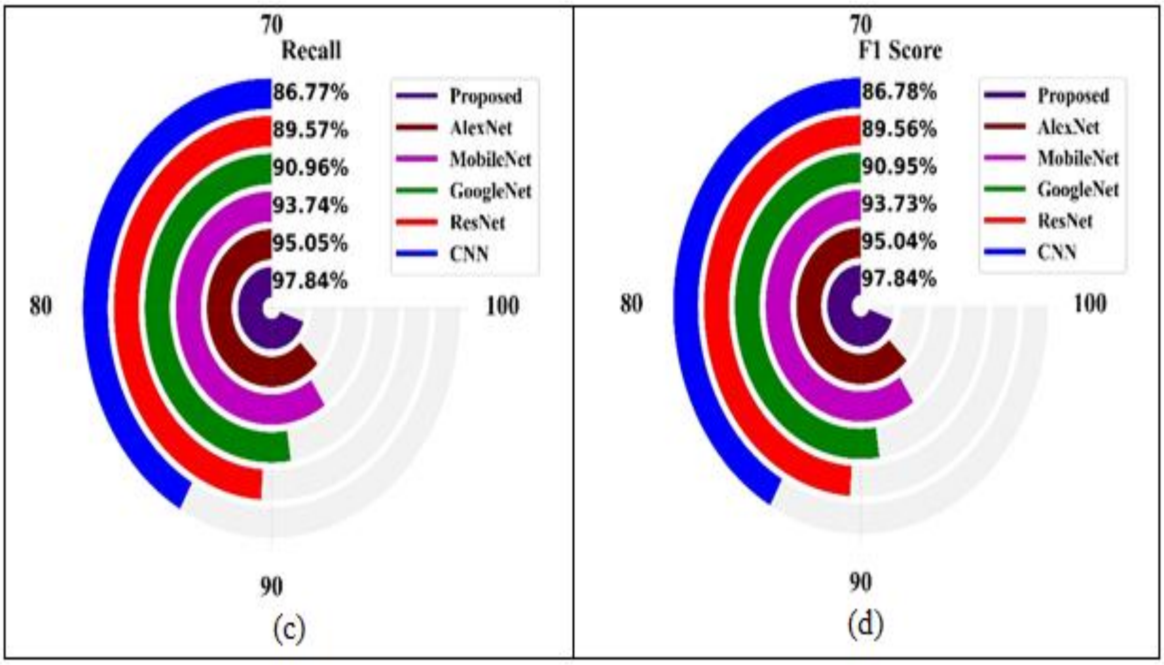


Figure 12 (a)-(d): Performance analysis of Accuracy, Precision, Recall and F1-score

The performance evaluation of existing and proposed model of accuracy, precision, recall and F1-score is shown in Figure 12 (a)-(b). Compare to existing, proposed had high accuracy rate at values of 98.56 % respectively. The proposed model done the classification effectively and reduce the computational problem. But in existing CNN and Alexnet model shown very low performance at value of 87.66 % and 95.88% respectively. However, these models contain deep network, it may assist with scaling issues. The performance, based on precision and recall of proposed model is high, at values of 97.84% and 97.84 % respectively. However in existing model had achieved extremely low performance. However, in proposed F1-score has achieved at high values of 97.84 % respectively. In existing F1-score become very low and models were consume more memory power and its cost of effectiveness also become high.

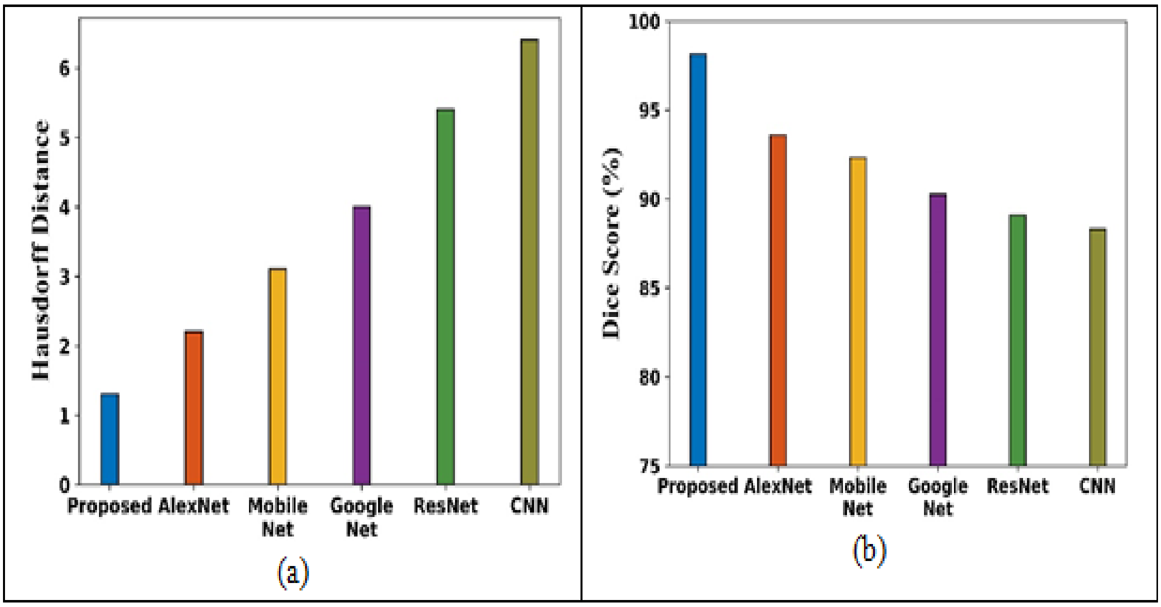


Figure 13 (a)-(b): Performance evaluation of Hausdorff Distance and Dice score

Hausdroff distance and dice score is compared with existing and proposed model, which is illustrated in Figure 13 (a)-(b). According to performance analysis, in proposed hausdroff distance and dice score has achieved better performance at values of 98.132% and 1.3031 respectively. However in existing model of CNN had low dice score performance at value of 88.30% respectively. Moreover CNN become difficult to interpret and class become present at imbalance state.

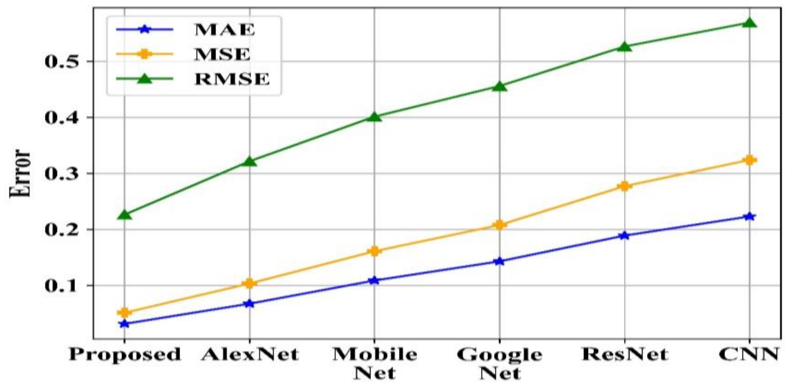


Figure 14: Comparative analysis of MAE, MSE and RMSE

The error metrics comparison of MAE, MSE and RMSE were illustrated in Figure 14. Based on comparison, the proposed model achieved low error value of MAE, MSE and RMSE at range of 0.0315, 0.0513 and 0.2265 respectively. In existing model had attained higher error values, it may cause less interactive for multi-classification.

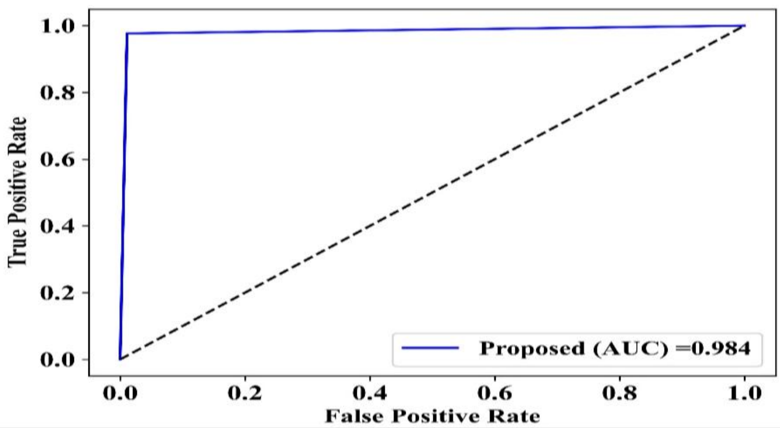


Figure 15: Representation of Area under curve

The probability of proposed model is mentioned in term of area under curve (AOC), the graphical representation of AOC is depicts in Figure 15. The graph can be chosen at positive and negative rate. Usually, the true positive rate has demonstrated the range of detection. In proposed model has achieved the range at 0.984 respectively.

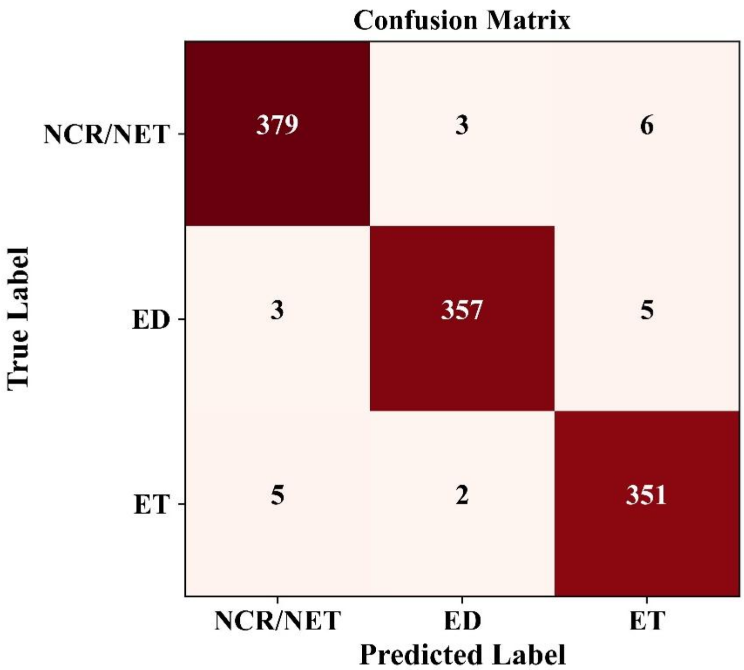


Figure 16: Confusion matrix

Figure 16 demonstrated the confusion matrix of BRATS 2021 dataset, here 379 classes for necrotic tumor core (NCR), is correctly prediction. The peritumoral edema (ED) is correctly predict as 357 classes and enhancing tumor (ET) were correctly predict as 351 classes. Similarly, the confusion matrix has also shown wrong class prediction.

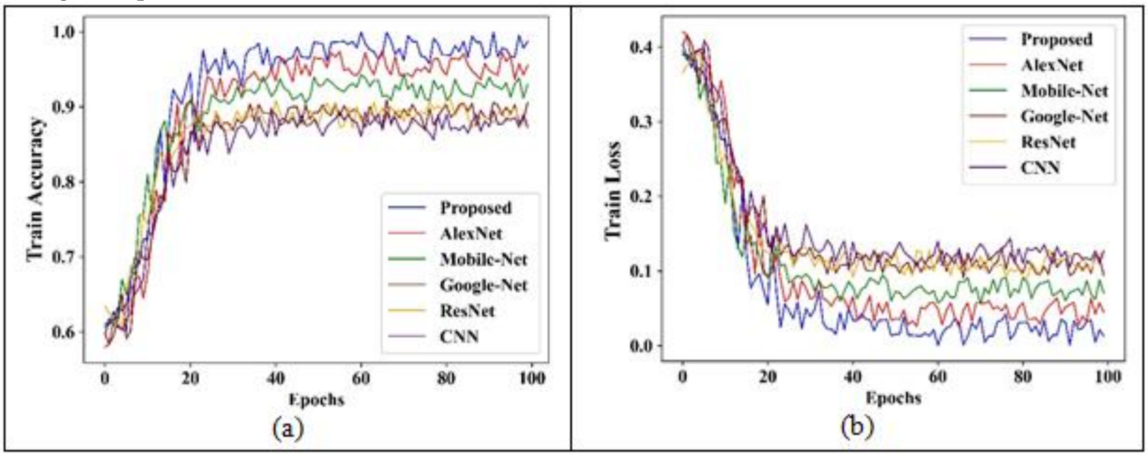


Figure 17 (a)-(b): Performance analysis of Training accuracy and loss curve

Comparison of existing and proposed model of training accuracy and loss curve is signifies in Figure 17 (a)-(b). The testing accuracy analysis with existing and proposed model is shown in Figure 17 (a), here existing had achieve low accuracy at 0.90, but comparatively proposed had high accuracy in the range of 0.95 with respective epochs. Figure 17 (b), represent the training loss for existing and proposed model. In proposed model had low loss testing, due to low loss value, classification reduce the memory utilization. However in existing model had high range of loss at 0.15 between corresponding epochs of 0 to 300.

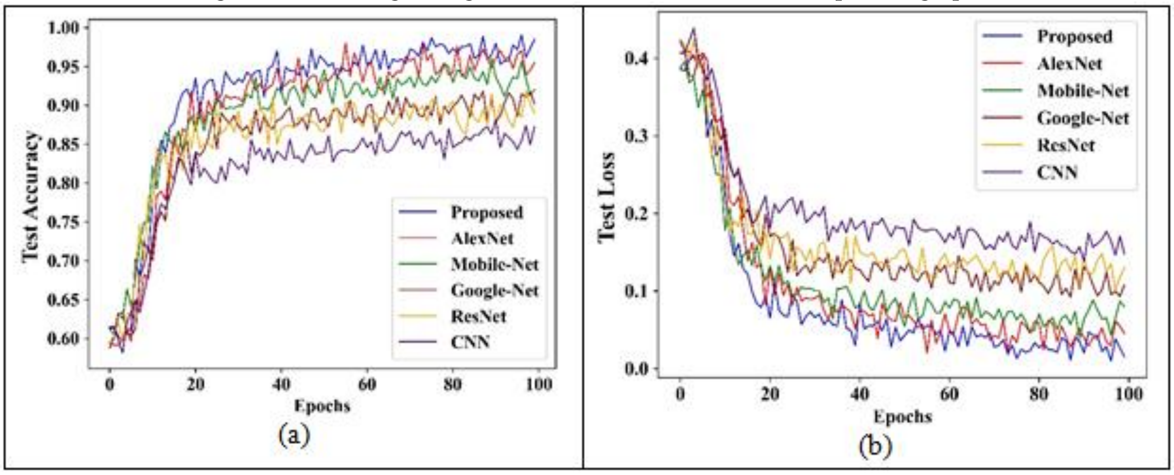


Figure 18 (a)-(b): Analysis performance for Testing accuracy and loss curve

Figure 18 (a)-(b) shown the comparison of existing and proposed model of testing accuracy and loss curve. The performance evaluation of existing and proposed model of testing accuracy is demonstrated in Figure 18 (a), here proposed model achieved accuracy at 0.97 between epochs of 0 to 100. However in existing, testing accuracy may extremely low at corresponding epochs. Comparative analysis of existing and proposed of testing loss curve is depicts in Figure 18 (b). The testing loss value for exiting model is very high, but proposed model attained, lesser loss value in the range of 0.0 9 between epochs of 0 to 100. Table 2 represent the comparison of BRATS 2020 dataset, Table 3 represent existing and proposed of BRATS 2021 dataset, Table 3 represent the comparison for BRATS 2020 dataset [29]

Table 2: Comparison of existing and proposed model of BRATS 2020 dataset

Metrics	Proposed	AlexNet	GhostNet	DenseNet	BiLSTM	CNN
Accuracy (%)	**98.83**	96.72	94.34	9289	90.14	88.79
Precision (%)	**98.22**	96.28	94.73	9257	90.43	88.44
Recall (%)	**98.24**	96.3	94.77	9259	90.46	88.46
F1 score (%)	**98.23**	96.29	94.75	9258	90.45	8845
MAE	**0.0202**	0.0442	0.091	0.1172	0.1768	0.2021
MSE	**0.0253**	0.0588	0.1279	0.1627	0.2592	0.2948

RMSE	**0.1591**	0.2425	0.3576	0.4033	0.5091	0.543
Dicescore (%)	**98.13298**	93.58968	92.32374	90.25862	89.09841	88.30822
Hausdroff distance	**1.303124**	2.203124	3.103124	4.003124	5.403124	6.403124

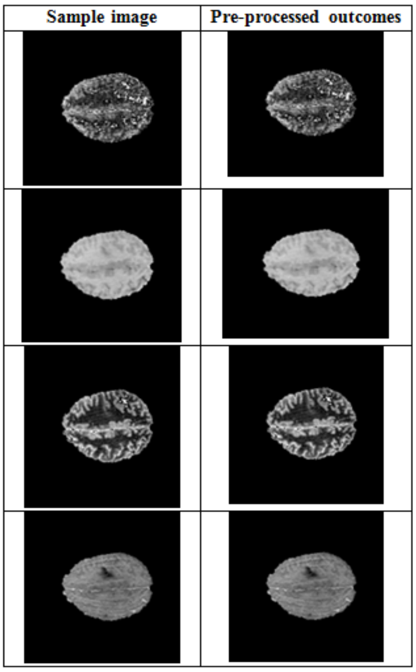
Table 3: Existing and proposed evaluation of BRATS 2021 dataset

Metrics	**Proposed**	**AlexNet**	**Ghost-Net**	**Dense-Net**	**BiLSTM**	**CNN**
Accuracy (%)	**98.56**	95.88	95.88	93.66	91.8	87.66
Precision (%)	**97.84**	95.04	95.04	93.72	90.93	86.8
Recall (%)	**97.84**	95.05	95.05	93.74	90.96	86.77
F1 score (%)	**97.84**	95.04	95.04	93.73	90.95	86.78
MAE	**0.0315**	0675	0.0675	0.1089	0.1431	0.2232
MSE	**0.0513**	0.1035	0.1035	0.1611	0.2079	0.324
RMSE	**0.2265**	0.3217	0.3217	0.4014	0.456	0.5692
Dicescore (%)	**98.13298**	93.58968	92.32374	90.25862	89.09841	88.30822
Hausdroff distance	**1.303124**	2.203124	3.103124	4.003124	5.403124	6.403124

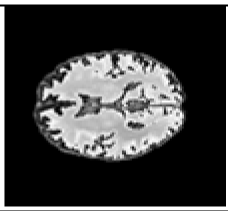
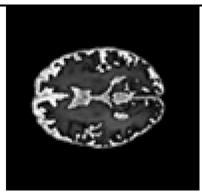
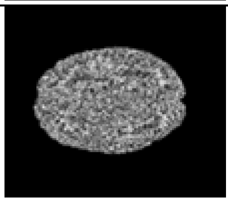
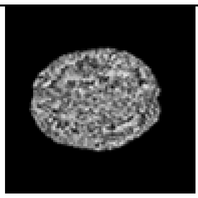
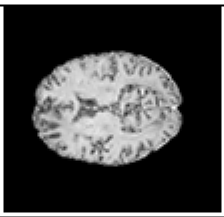
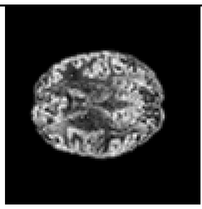
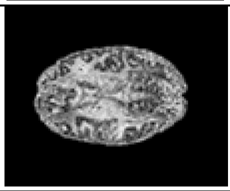
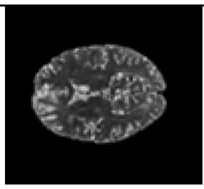
Table 4: Comparison for BRATS 2020 dataset

Models	**Accuracy**	**F1-score**
ResNet50	92.54	93.33
DenseNet 201	91.04	92.30
AlexNet	89.55	90.05
Inception V3	88.06	89.99
GoagleNet	71.64	66.03
Proposed	**98.83**	**98.23**

4.5 Visualization of sample and pre-processed image BRATS 2020 dataset



4.6 Visualization of sample and pre-processed image for BRATS 2021 dataset

Sample image	Pre-processed outcomes
	
	
	
	

4.7 DISCUSSION

In this section, the efficiency of the proposed model is thoroughly discussed. The performance of the existing model is compared with advanced proposed model. In existing, noted several limitations to classify the brain tumor such as complexity issues, feature present with low resolution of image. However the image with high resolution cannot handle on smaller dataset, it require more computational cost and utilize more memory. Moreover in existing prediction had occur higher error rate, meanwhile in existing model obtained considerable amount of training time for prediction [32-33]. The proposed model overcome all these issue and ensure the tumor classification as effectively. In proposed the advanced technique of Grp-DRcU-Net has identify the tumor with high efficiency and this model had reduce the network parameter and improve the training process. Similarly in proposed, the novel classifier model of Netr-HDCViT has tripartite attention which accurately focus towards the feature extraction and enhance the stability of model. However the novel model become fast inference and reduce the computational constraints. The performance of the proposed model is evaluated on two datasets and achieve comparatively high range of accuracy at 98.83 % and 98.56% respectively. However for evaluating performance, the proposed model is compared with several existing methods those methods attain certain limits like fitting and scalability issues. Based on comparative analytics, proposed model tackle all those limits and improve the classification of tumor analysis with better accuracy performance.

5. CONCLUSION

The research work is presented with advanced approach of hybrid knowledge distillation transformer model with improved U-Net model for brain tumor detection. For tumor detection, the images were pre-processed with Pix-TrMed filter and next taken the pre-processed image as for identification The identification technique of Grp-DRcU-Net model has make to identify the tumor effectively. Followed by identification, extract the features present in the image by use of Im-ResN model. This extraction technique has capture the local as well as global context images very well. According to selected feature, classify the classes with novel model of Netr-HDCViT. Here, the novel model contain transformative method of knowledge distillation and tripartite attention.

The framework of tripartite attention can acts as heart of the proposed model, it show an benchmark on multiclass brain tumor detection. By the way combination of tripartite attention with knowledge distillation has ability to process in complex medical image analysis. In this study, the special findings indicate that tripartite attention reduce the student and teacher losses and effectively showing its ability on distill knowledge. Meanwhile, the presence of transformer and network pruning has been enhance the model in term accuracy and efficiency. However pruning has reduce the model size and complexity, moreover, the proposed model experimental performance is demonstrated with two dataset namely BRATS 2020 and BRATS 2021. The BRATS 2020 achieve high performance in term accuracy of 98.83 %, precision of 98.22%, recall of 98.24% and F1-score of 98.23 % respectively. Similarly the BRATS 2021 also achieved better performance of accuracy, precision, recall and F1-score at values of 98.56%, 97.84%, 97.84% and 97.84% respectively. In future, research will add a contrast enhanced MRI to dataset, it may aid further differentiate of tumor classification. Also improve the explainability of the model's prediction, making more reliable and transparent tool for medical professionals.

REFERENCE

[1] Rasool, Novsheena, Javaid Iqbal Bhat, Niyaz Ahmad Wani, Naveed Ahmad, and Mohammed Alshara. "TransResUNet: Revolutionizing Glioma Brain Tumor Segmentation through Transformer-Enhanced Residual UNet." IEEE Access (2024).
[2] Ferdous, Gazi Jannatul, Khaleda Akhter Sathi, Md Azad Hossain, Mohammed Moshiul Hoque, and M. Ali Akber Dewan. "LCDEiT: A linear complexity data-efficient image transformer for MRI brain tumor classification." IEEE Access 11 (2023): 20337-20350.
[3] ZongRen, Li, Wushouer Silamu, Wang Yuzhen, and Wei Zhe. "DenseTrans: multimodal brain tumor segmentation using swin transformer." IEEE Access 11 (2023): 42895-42908.
[4] Anaya-Isaza, Andrés, Leonel Mera-Jiménez, Lucía Verdugo-Alejo, and Luis Sarasti. "Optimizing MRI-based brain tumor classification and detection using AI: A comparative analysis of neural networks, transfer learning, data augmentation, and the cross-transformer network." European Journal of Radiology Open 10 (2023): 100484.
[5] Wang, Jian, Si-Yuan Lu, Shui-Hua Wang, and Yu-Dong Zhang. "RanMerFormer: Randomized vision transformer with token merging for brain tumor classification." Neurocomputing 573 (2024): 127216.
[6] Saeedi, Soheila, Sorayya Rezayi, Hamidreza Keshavarz, and Sharareh R. Niakan Kalhori. "MRI-based brain tumor detection using convolutional deep learning methods and chosen machine learning techniques." BMC Medical Informatics and Decision Making 23, no. 1 (2023): 16.
[7] Mahmud, Md Ishtyaq, Muntasir Mamun, and Ahmed Abdelgawad. "A deep analysis of brain tumor detection from mr images using deep learning networks." Algorithms 16, no. 4 (2023): 176.
[8] Ibrahim, Rahmeh, Rawan Ghnemat, and Qasem Abu Al-Haija. "Improving Alzheimer's disease and brain tumor detection using deep learning with particle swarm optimization." AI 4, no. 3 (2023): 551-573.
[9] Asad, Rimsha, Saif Ur Rehman, Azhar Imran, Jianqiang Li, Abdullah Almuhaimeed, and Abdulkareem Alzahrani. "Computer-aided early melanoma brain-tumor detection using deep-learning approach." Biomedicines 11, no. 1 (2023): 184.
[10] Aggarwal, Mukul, Amod Kumar Tiwari, M. Partha Sarathi, and Anchit Bijalwan. "An early detection and segmentation of Brain Tumor using Deep Neural Network." BMC Medical Informatics and Decision Making 23, no. 1 (2023): 78.
[11] Patil, Suraj, and Dnyaneshwar Kirange. "Ensemble of deep learning models for brain tumor detection." Procedia Computer Science 218 (2023): 2468-2479.
[12] Abdusalomov, Akmalbek Bobomirzaevich, Mukhriddin Mukhiddinov, and Taeg Keun Whangbo. "Brain tumor detection based on deep learning approaches and magnetic resonance imaging." Cancers 15, no. 16 (2023): 4172.
[13] Venmathi, A. R., S. David, Ennam Govinda, K. Ganapriya, R. Dhanapal, and A. Manikandan. "An automatic brain tumors detection and classification using deep convolutional neural network with VGG-19." In 2023 2nd International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA), pp. 1-5. IEEE, 2023.
[14] Rajendran, Surendran, Suresh Kumar Rajagopal, Tamilvizhi Thanarajan, K. Shankar, Sachin Kumar, Najah M. Alsubaie, Mohamad Khairi Ishak, and Samih M. Mostafa. "Automated segmentation of brain tumor MRI images using deep learning." IEEE Access 11 (2023): 64758-64768.
[16] Jain, Sachin, and Vishal Jain. "Novel approach to classify brain tumor based on transfer learning and deep learning." International Journal of Information Technology 15, no. 4 (2023): 2031-2038.
[16] Hossain, Shahriar, Amitabha Chakrabarty, Thippa Reddy Gadekallu, Mamoun Alazab, and Md Jalil Piran. "Vision transformers, ensemble model, and transfer learning leveraging explainable AI for brain tumor detection and classification." IEEE Journal of Biomedical and Health Informatics 28, no. 3 (2023): 1261-1272.
[17] Aloraini, Mohammed, Asma Khan, Suliman Aladhadh, Shabana Habib, Mohammed F. Alsharekh, and Muhammad Islam. "Combining the transformer and convolution for effective brain tumor classification using MRI images." Applied Sciences 13, no. 6 (2023): 3680.
[18] Liu, Yu, Yize Ma, Zhiqin Zhu, Juan Cheng, and Xun Chen. "TransSea: Hybrid CNN-Transformer with Semantic Awareness for 3D Brain Tumor Segmentation." IEEE Transactions on Instrumentation and Measurement (2024).
[19] Reddy, C. Kishor Kumar, Pulakurthi Anaghaa Reddy, Himaja Janapati, Basem Assiri, Mohammed Shuaib, Shadab Alam, and Abdullah Sheneamer. "A fine-tuned vision transformer based enhanced multi-class brain tumor classification using MRI scan imagery." Frontiers in Oncology 14 (2024).
[20] Pacal, Ishak. "A novel Swin transformer approach utilizing residual multi-layer perceptron for diagnosing brain tumors in MRI images." International Journal of Machine Learning and Cybernetics (2024): 1-19.
[21] Hatamizadeh, Ali, Vishwesh Nath, Yucheng Tang, Dong Yang, Holger R. Roth, and Daguang Xu. "Swin unetr: Swin transformers for semantic segmentation of brain tumors in mri images." In International MICCAI brainlesion workshop, pp. 272-284. Cham: Springer International Publishing, 2021.
[22] Aggarwal, Mukul, Amod Kumar Tiwari, M. Partha Sarathi, and Anchit Bijalwan. "An early detection and segmentation

of Brain Tumor using Deep Neural Network." BMC Medical Informatics and Decision Making 23, no. 1 (2023): 78.

[23] Aboussaleh, Ilyasse, Jamal Riffi, Adnane Mohamed Mahraz, and Hamid Tairi. "Inception-UDet: an improved U-Net architecture for brain tumor segmentation." Annals of Data Science 11, no. 3 (2024): 831-853.

[24] Gunasekara, Shanaka Ramesh, H. N. T. K. Kaldera, and Maheshi B. Dissanayake. "A systematic approach for MRI brain tumor localization and segmentation using deep learning and active contouring." Journal of Healthcare Engineering 2021, no. 1 (2021): 6695108.

[25] Tiwari, S. (2022). Usage of drones as a medical applications. International Journal of Health Sciences, 6(S1), 11643–11655. https://doi.org/10.53730/ijhs.v6nS1.7854

[26] Dubey S., Shukla O., and Tiwari S., "Analysis of Application of Natural Language Processing in Artificial Intelligence," International Journal of Mechanical Engineering, vol. 7, no. 5, pp. 419- 421, 2022.

[27] Yu, Helong, Xianhe Cheng, Chengcheng Chen, Ali Asghar Heidari, Jiawen Liu, Zhennao Cai, and Huiling Chen. "Apple leaf disease recognition method with improved residual network." Multimedia Tools and Applications 81, no. 6 (2022): 7759-7782.

[28] Chong, You Li, Chin Poo Lee, Shahrin Zen Muhd-Yassin, Kian Ming Lim, and Ahmad Kamsani Samingan. "TransKGQA: Enhanced Knowledge Graph Question Answering With Sentence Transformers." IEEE Access (2024).

[29] Çinar, Ahmet, and Muhammed Yildirim. "Detection of tumors on brain MRI images using the hybrid convolutional neural network architecture." Medical hypotheses 139 (2020): 109684.

[31] https://www.kaggle.com/datasets/awsaf49/brats20-dataset-training-validation

[32] https://www.kaggle.com/datasets/dschettler8845/brats-2021-task1

[33] Esakkirajan, S., T. Veerakumar, Adabala N. Subramanyam, and C. H. PremChand. "Removal of high density salt and pepper noise through modified decision based unsymmetric trimmed median filter." IEEE Signal processing letters 18, no. 5 (2011): 287-290.