Texton Tri-alley Separable Feature Merging (TTSFM) Capsule Network for Brain Tumor Detection

Vivian Akoto-Adjepong, Obed Appiah, Peter Appiahene, Patrick Kwabena Mensah
Department of Computer Science and Informatics-University of Energy and Natural Resources, Sunyani, Ghana

***Abstract*—Brain tumors represent one of the most perilous and lethal forms of tumors in both children and adults. Early detection and treatment of such malignant disease types may reduce the mortality rate. However, manual procedures can be used to diagnose such disorders, and this process necessitates a careful, in-depth analysis which is prone to errors, tedious for health professionals, and time-consuming. Therefore, this research aims to design a Texton Tri-alley Separable Feature Merging (TSFM) Capsule Network based on dynamic routing, suitable for the automatic detection of brain tumors. The TSFM Capsule Network's Texton layer helps to extract important features from the input image, and the separable convolutions coupled with the use of fewer filters and kernel sizes help to reduce the time for training, the size of the model on disk, and the number of trainable parameters generated by the model. The model's evaluation results on the brain tumor dataset consisting of four classes show better performance than the traditional capsule network, and are comparable to the state-of-the-art models, with an overall accuracy of 97.64%, specificity of 99.24%, precision of 97.43%, sensitivity of 97.45%, f1-score of 97.44%, ROC rate of 99.50%, PR rate of 99.00%. The components and properties of the proposed model make the model deployable on devices with low memory like mobile devices. This model with better performance can assist physicians in the diagnosis of brain tumors.**

***Keywords*—Texton; separable convolutions; capsule neural network; dynamic routing; brain tumor; brain tumor detection**

I. INTRODUCTION

Brain tumor is among the most fatal and dangerous tumors in both children and adults [1]. Brain and spinal cord tumors are assemblages of abnormal cells that have multiplied uncontrollably inside the brain or spinal cord. There will be 25,050 diagnoses of malignant brain and spinal tumors by 2023 in both men and women[2].

Medical imaging plays a vital role in the diagnosis, monitoring of tumor progression, and treatment of tumors. Magnetic Resonance Imaging (MRI) is the preferred technique for imaging due to its non-ionizing nature. It offers significant insights into the characteristics, dimensions, form, and positioning of brain tumors.

Manually evaluating MRIs is laborious and error-prone, hence an Artificial intelligence (AI)-driven system that operates automatically is required to aid in medical diagnosis. Techniques rooted in machine learning, like support vector machines, have been utilized to aid in accurately detecting medical conditions [3]. Nevertheless, the outcomes of these approaches fell short of established benchmarks, and the process of extracting features is notably time-intensive. To tackle these challenges, deep learning techniques like convolutional neural networks (CNNs) were embraced to enhance the process of extracting features. Remarkably, CNNs demonstrated a level of performance that is comparable to that of human experts.

Despite CNN's strong achievements, the study found specific constraints including the need for extensive datasets, high computational demands [4], translational invariance [5], and adherence to particular criteria for optimal feature selection [6]. In the field of health, obtaining a voluminous dataset poses a significant hurdle, compounded by a scarcity of skilled annotators and privacy issues [4]. Consequently, to mitigate the overfitting of CNNs on these limited datasets, methods of data augmentation are employed. However, it should be noted that these data augmentation techniques are both time-consuming and labor-intensive [7].

Capsule Network (CapsNet) was introduced to tackle the issues of CNN [8]. In contrast to CNNs, CapsNet does not necessitate extensive datasets, and is resistant to uneven class distributions and spatial orientation changes. These properties of CapsNet render it appropriate for medical image diagnosis. However, CapsNets do have their own set of limitations[9]. They exhibit suboptimal performance on complex images, and those with diverse backgrounds, and try to account for every element in an image. As a result of these properties, the performance of the network may suffer when dealing with detailed malignant images.

In order to further improve CapsNets textural, color, and spatial recognition capabilities, this paper adopts CapsNets dynamic routing algorithm and implements a Texton layer [10], separable convolutions, and a max-pooling layer. This allows CapsNet to decide on which features are essential and the coupling coefficients that need to be decreased in enhancing the hierarchical relationship of closely related capsules. The Texton Tri-alley Separable Feature Merging (TTSFM) Capsule Network model proposed helps to address the crowding problem in CapsNet and performs better than the traditional CapsNet and some other models found in literature on brain tumor detection. The proposed model exhibits better convergence speed and can generalize well on unknown data, hence can serve as an intelligent tool, assisting physicians in diagnosing and administering appropriate

treatments for brain tumors. Fig. 1 depicts the adopted workflow for the proposed work.

Most of the models found in literature performed dataset balancing, data segmentation, data augmentation, and thorough data preparation, before model fitting. This study utilized raw datasets without any data augmentation and data preprocessing to evaluate the proposed model's effectiveness with natural or raw data since data augmentation or segmentation might not be practical in medical emergency situations. Also, this study offers detailed visual representations of image regions that capture the focus of specific parts of our model, evaluation performance on imbalanced datasets using ROC and Precision-Recall (PR) curves, clusters of features at the class capsule layer to assess the model's effectiveness, and the model's transparency and understandability was enhanced by reconstructing input images.

The contributions of this study are:

1) A fast, robust, and low-parameterized TTSFM CapsNet, that has efficient feature extraction capabilities is proposed.

2) Separable convolutions are employed to reduce the size and trainable parameters of the model.

3) A comparative analysis was conducted to assess the proposed model with other CapsNet models.

4) The study presented a comprehensive visual representation of the outputs of layers to help offer notable contributions to the explainable artificial intelligence field.

The study is organized as follows: Related works are presented in Section II. Methodology is presented in Section III. Section IV deals with results and discussion and Section V deals with the conclusion and future works of the study.

II. Related Works

Manual diagnosis in the medical field is prone to error, tedious for health professionals, and time-consuming. These limitations led to the employment of algorithms for predicting and detecting radiomic medical conditions. For instance, Gao et al., [11] utilized both 2D and 3D convolutional neural networks (CNNs), the researchers employed these networks to categorize individuals as having tumors, no tumors, or Alzheimer's disease based on CT scans. They were able to attain an accuracy level of 87.6%. A hybrid approach employing CNN and Neutrosophy (NS-CNN), was proposed by Özyurt et al [12]. The approach was used to categorize benign or malignant segmented tumor regions from brain tumor images. The accuracy of the suggested model was 95.62 %. Sajjad et al. [13] presented a modified (CNN) based multi-grade system for classifying brain tumor grades. The model's accuracy was 90.67 %. In order to solve the classification challenge for brain tumors, Ayadi et al. [14] suggested a new model that makes use of the CNN sequential model. The model has several layers and was designed to categorize MRI brain cancers. The model had a 94.74 % accuracy rate. A CNN model was proposed by Badža and Barjaktarovic [15] for classing three distinct forms of brain tumors. The suggested model, which has a straightforward architecture akin to traditional CNN, accurately classified 96.56 % of the brain tumor MRI images in the dataset. Also, Afshar et al. [16], introduced a boosted capsule network (also known as BoostCaps), that makes use of boosting approaches' capacity to accommodate poor learners, by steadily boosting the models. The BootsCaps architecture, according to the results, classified brain tumors with an accuracy of 92.45%. DCNet and DCNet++ were suggested by Phaye et al.[17]. By substituting densely connected convolutions for the typical convolutional layers in the two suggested models, the CapsNet was modified. On Brain Tumor Dataset, the two models were assessed and achieved a validation accuracy of 93.04 % and 95.03%, respectively. A capsule network for automatic brain tumor classification that achieves a 92.65% accuracy was proposed by Goceri. This network includes three fully connected layers and utilizes an expectation-maximization (EM)-based dynamic routing algorithm to extract important features from images [18]. In order to increase the focus of CapsNet, Afshar and colleagues suggested an improved CapsNet architecture for classifying brain tumors that incorporates the tumor coarse boundaries as additional inputs. The validation accuracy for the model was 90.89% [19]. According to Adu and friends, an improved CapsNets with several convolutional layers and dilation to preserve image resolution and boost classification accuracy was proposed. The proposed system can guarantee an increase in CapsNets focus by inputting segmented tumor regions within the structure. This model's performance obtained an accuracy of 95.54 % [20]. Some researchers presented the BayesCap, a Bayesian CapsNet architecture that can offer both the mean forecasts and entropy as a gauge of uncertainty in forecasting. According to the findings, accuracy can be increased by filtering out uncertain forecasts. The model's maximum accuracy was 73.9 % with a CI of: (73.5%, and 74.4%) [21].

All the existing models performed well on the various datasets. But for medical image diagnosis, there is a need for a more robust and efficient model for better diagnosis, hence this study aims to propose an improved, fast, low-parameterized, and robust Capsule Network which incorporates Texton and Separable convolutions for effective feature extraction and better classification of brain tumor diseases. Most of the studies mentioned above performed dataset balancing, data segmentation, data augmentation, and thorough data preparation, before model fitting.

III. Methodology

This section presents the methodologies employed to attain our goal of developing a deployable CapsNet that has an effective ability to extract features efficiently with lesser parameters and size on disk. Fig. 1 shows the proposed methods block diagram for the automatic classification of brain tumor types.

A. Capsule Network

The structure of the baseline CapsNet on which the proposed model is based is found in Fig. 2.

B. Texton Detection

The notion of Texton involves the identification of clusters of shapes within an image that possess a shared characteristic. Julesz further developed this concept [22], placing emphasis on the significance of measuring the distances between texture elements when calculating gradients of Textons. Textures emerge only when neighboring elements are in proximity, and the scale of the elements impacts the surrounding region. Larger elements oriented in a particular direction can slightly impede the initial, instinctive differentiation. Gradients of Texton are only present at the borders of textures, so utilizing a smaller element size, such as 2x2, can enhance the distinction of textures. The approach of Multi Texton Detection (MTD), utilized to extract information regarding edges and colors, involves the use of six distinct types of Texton (T_1, T_2, T_3, T_4, T_5, and T_6) on a 2x2 grid (shown in Fig. 3) to identify textons. A Texton is generated by the grid when the two shaded pixels share the same value. By systematically shifting the 2x2 block across the image C (x, y), textons can be detected in a stepwise manner. If a texton is identified, the original pixel values are preserved; otherwise, the block is disregarded. The resultant image containing textons is represented as T (x, y) [10], as illustrated in Fig. 4.

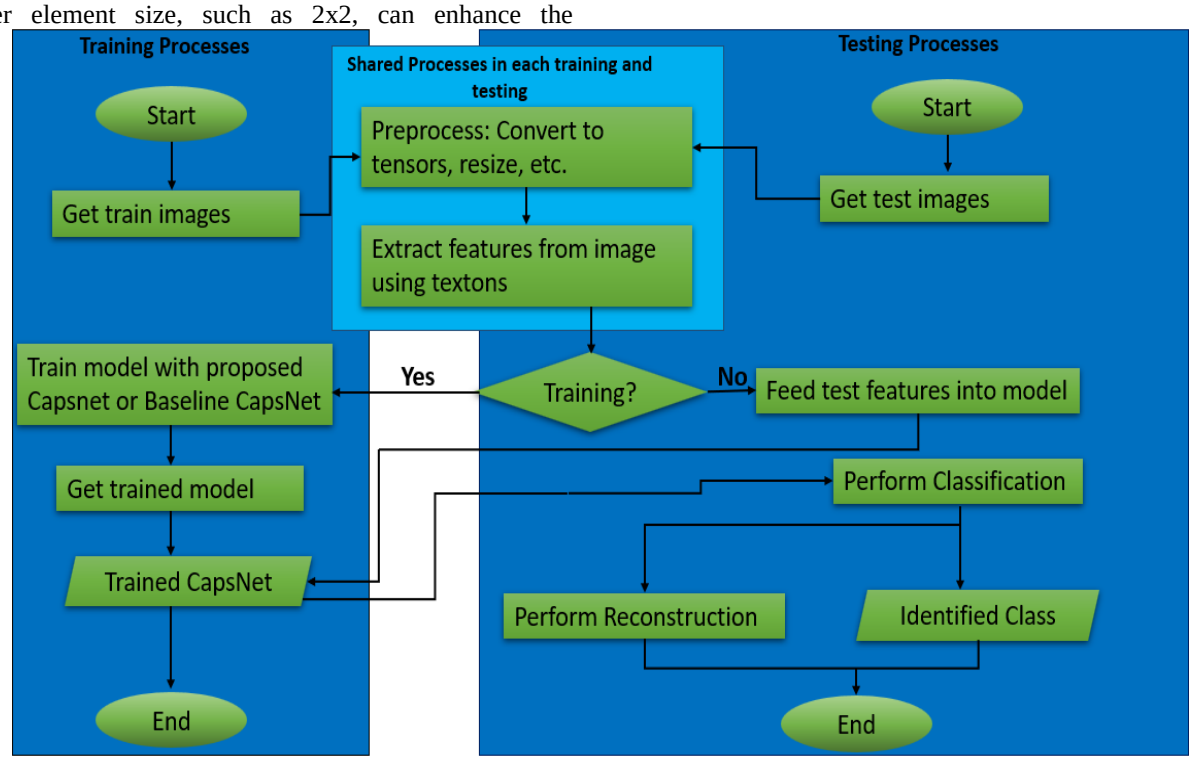

Fig. 1. Workflow diagram of the study.

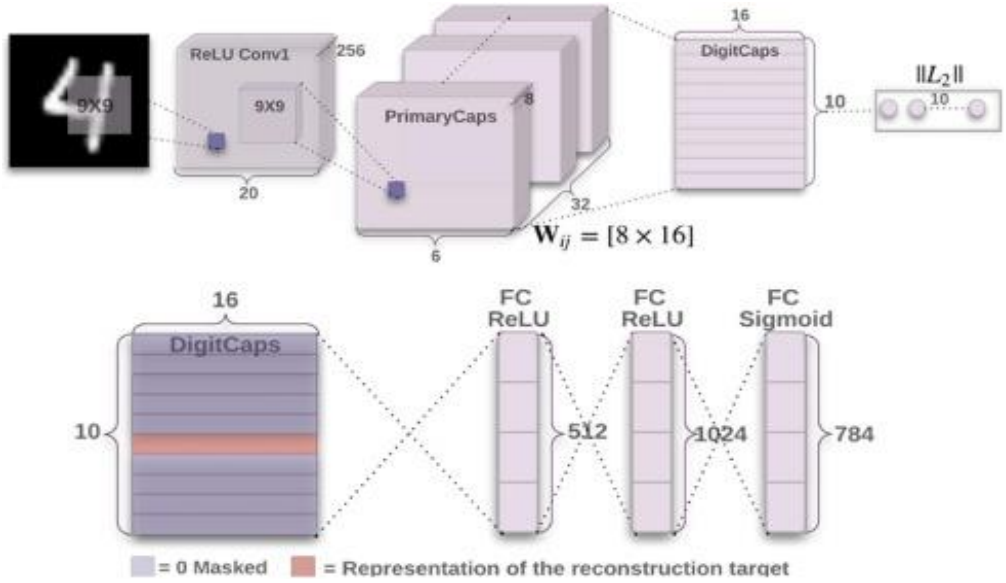

Fig. 2. Architecture of the baseline capsule network model.

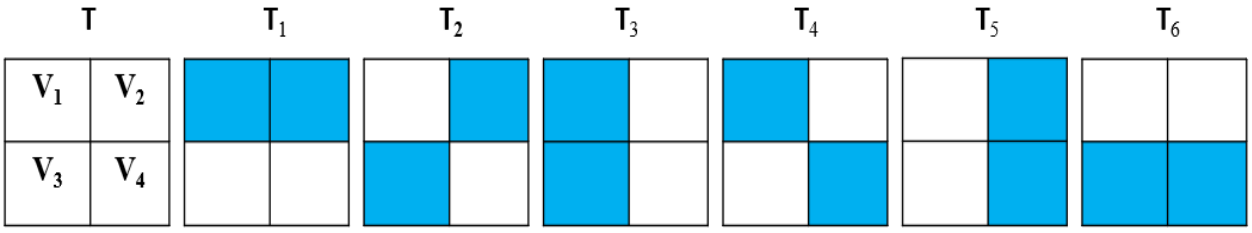

Fig. 3. Six Texton types used in Texton detection process: (T) 2x2 grid.

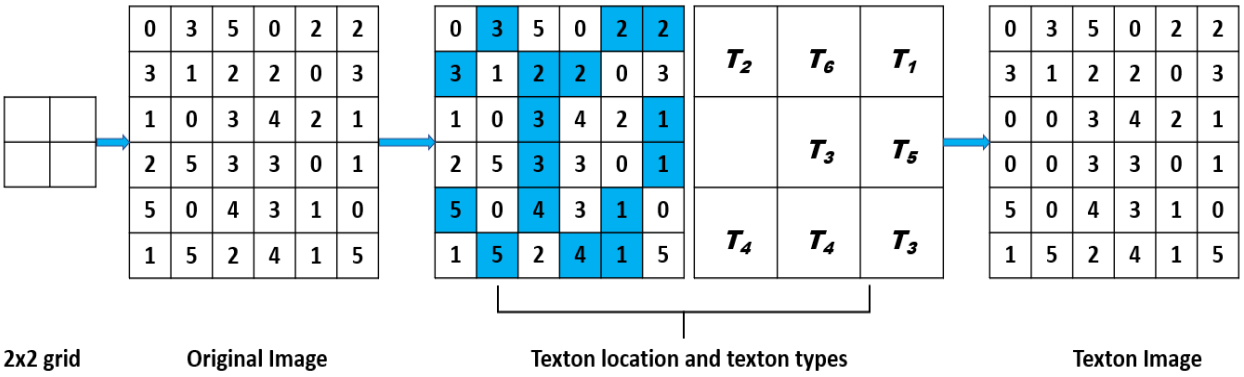

Fig. 4. Illustration of the Texton detection process.

C. Depthwise Separable Convolution

Two separable convolutions types exist in separable convolutional neural networks. These are depthwise separable convolutions (DSC) and spatial separable convolutions (SSC). DSC is adopted for this study, and can be viewed as grouped convolutions, similar to the concept of 'inception modules' employed in the design of the Xception architecture [23]. It relies on spatial convolution, which operates separately on each input channel. After the spatial convolution, a pointwise convolution (PC) is executed, which involves a standard convolution using 1 × 1 windows. This leads to the creation of a new channel space as a result of projecting the channels calculated during the depthwise convolution (DC). The mathematical expression for depthwise convolution (DC) and pointwise convolution (PC) is presented as follows:

$$\mathrm{PC}(\mathrm{W},\mathrm{y})_{(\mathrm{i,j})} = \sum_{\mathrm{m}}^{\mathrm{M}} \mathrm{W_m} \cdot \mathrm{y}_{(\mathrm{i,j,m})} \quad (1)$$

$$DC(W,y)_{(i,j)} = \sum_{k,l}^{K,L} W_{(k,l)} \odot y_{(i+k,\ j+l)} \quad (2)$$

$$DSC(W_p,W_d,y)_{(i,j)} = PC_{(i,j)}\left(W_P, DC_{(i,j)}(W_d,y)\right) \quad (3)$$

W_p and W_d represent the inputs used for pointwise and depthwise convolutions in the above equations, respectively. The operator ⊙ within Eq. (2) pertains to the element-wise multiplication. Consequently, the fundamental idea underpinning depthwise separable convolutions involves splitting the feature extraction process carried out by regular convolutions across a unified "space-cross-channels domain" into two distinct stages: spatial pattern learning and channel fusion. This approach represents a generalization when dealing with convolution operations on 2D or 3D inputs having both relatively independent channels and closely interconnected spatial positions.

D. Proposed Model

Fig. 5 shows the proposed Texton Tri-alley Separable Feature Merging (TTSFM) CapsNet model that employs Texton, separable convolution, traditional convolutions, max pooling, dropout and reconstruction layers. The Texton layer is used to extract important texture and edge features from the input image. The output features from the Texton layer are processed by three different separable convolutions (each having 32 filters, kernel size of 2x2, depth multiplier of 1, depthwise and pointwise initializers of “ones” and a stride of 1) followed with batch normalization and max pooling, contributing to reduced number of parameters, model size, computational time and complexity of the model, as can be seen at alley_1_conv1, alley_2_conv1, and alley_3_conv1. The feature map from alley_1_conv1 serves as input in into alley_1_conv2, and the feature maps from alley_2_conv1, and alley_3_conv1 are merged and serves as input into alley_2_conv2, and alley_3_conv2. The feature maps from alley_1_conv2, alley_2_conv2, and alley_3_conv2 (all conv2 layers employs 64 filters, kernel size of 3x3, and a stride of 1) are then concatenated and sent as input into a dropout layer, followed with a batch normalization layer. This feature map is sent as input into the primary capsule layer consisting of 32 channels with eight dimensions, a kernel size of 3x3, and a stride of two. Features from this primary capsule are then sent to the TumorCaps layer (by employing dynamic routing algorithm) for classification. This TumorCaps consist of the total classes number in 16D capsules. The output of TumorCaps is directed to the reconstruction layer, which works on rebuilding the characteristics acquired from the TumorCaps. The features are then transferred to the decoder layer within the capsule, which decodes the properties of the entity. This decoder is composed of three layers of fully connected neurons, with counts of 512, 1024, and 3072 respectively. The Texton, max pooling, and convolution layers

help to extract important features from the input images using separable convolutions results in reduced number of parameters, model size, computational time and complexity of the model.

E. Dataset Description

Brain Tumor: The dataset comprises 7022 MRI scans of the human brain, which are grouped into four categories: (1) glioma, (2) meningioma, (3) pituitary, and (4) no tumor. To make the dataset more manageable, the images were resized to 32 × 32 × 3 and redistributed using 70:20:10 leave-out approach. This publicly available dataset can be found at: https://www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset.

F. Experimental Setup

The proposed model was developed and assessed using Keras, Python via Anaconda, and employed the TensorFlow backend on a 64-bit Windows computer. The hardware configuration encompassed an NVIDIA GeForce RTX 2080 SUPER GPU having 8GB of dedicated GPU memory along with 32GB of system RAM. During the training stage, Adam optimizer was employed with a learning rate set to 0.001, and training operations were executed using batches of 100 samples. In order to ensure optimal training progress, the model achieving the highest performance was saved during the training iterations. The evaluation of loss was conducted using the margin loss, represented as L_k in Eq. (4), with specific details provided below:

$$L_k = T_k \, max(0, m^+ - ||v_k||)^2 + \lambda(1 - T_k) \, max(0, ||v_k|| - m^-)^2 \quad (4)$$

where, T_k is 1 when class k is active and 0 otherwise. Hyper-parameters λ, m^-, m^+ are set during the learning process.

G. Performance Evaluation Measures

The following metrics were employed in this study for the purpose of classification:

Validation Accuracy: Calculates the proportion of accurately classified classes from the total number of classes. The attained overall validation accuracy for the entire set of experiments is reported.

Loss: Evaluates the variance between the model's predictions and the actual labels. This assessment employs the margin loss during testing.

Confusion Matrix: Assist in providing a thorough examination of the tally of correctly and incorrectly categorized images. Factors like True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) are employed to evaluate diverse measurements such as precision, accuracy, specificity, sensitivity (recall), and additional indicators.

Precision (P): The proportion of accurately detected positive instances compared to the overall number of predicted positive instances.

Recall (R) or Sensitivity: the proportion of accurately detected positive instances in relation to the overall count of positive instances within the dataset.

Specificity: The proportion of negative instances have been accurately recognized in relation to the overall count of negative instances present in the dataset.

F1-Score: The Mean that combines precision and recall in a harmonic manner.

Area under the curve (AUC): The model's performance is assessed on datasets where classes are imbalanced or unevenly distributed by creating Receiver Operating Characteristic (ROC) and precision-recall (PR) curves [24]-[25]. Higher AUC values are favored compared to their smaller equivalents.

Clustering: We utilize t-distributed stochastic neighbor embedding to acquire and examine the clusters within the class capsule layer of the models.

IV. Results And Discussions

In this section, we present the outcomes of our experiments and demonstrate the favorable performance of the model when tested on the brain tumor dataset in comparison with the baseline Capsule Network [5]. To bolster confidence and ensure the reliability of the model's outcomes, we employed and meticulously executed various evaluation methods. These techniques included evaluating metrics such as classification accuracy and loss, specificity, sensitivity, precision, F1-Score, number of parameters, Area Under the Curve (AUC) for both the Receiver Operating Characteristic Curve (ROC) and Precision-Recall (PR) curves. We also trained a traditional capsule network [5] using the same dataset and compared its results with our model's performance using the aforementioned metrics.

A. Performance Evaluation

Graphs presented in Fig. 6 illustrate the accuracy and loss trends for both CapsNet models: the proposed and the baseline model. The accuracy and loss graphs during training and validation reveals that the proposed model outperform the baseline CapsNet model, exhibiting superior and consistent accuracy with quicker convergence. It is important to highlight that while accuracy is widely used to assess classification algorithms, it may not be suitable for evaluating medical images due to their small size and significant class imbalance [26]. Despite its limitations, accuracy can offer an overview of overall system.

Fig. 7 displays the ROC and PR curves for both the proposed and baseline models. Analyzing the data from these curves, it becomes evident that the performance of the proposed model surpasses that of the baseline model. This shows that the proposed model performs better on small and imbalanced datasets like medical images [27] than the baseline model.

Fig. 8 depicts confusion matrices illustrating the accurate and erroneous image identifications. The results presented in Table I highlight that the proposed model outperformed the CapsNet baseline by exhibiting fewer misclassifications.

Hence resulted in better per class accuracy, specificity, sensitivity, precision and F1-Score for each class, as compared to the baseline Capsule Network model.

B. Ablation Study

Conducting ablation experiments involves analyzing the elements of the model that significantly influenced its performance [28][29]. The model's layers are systematically removed in succession to assess their effects on the overall model's performance. As depicted in Table II, the model's performance shows a significant improvement through the integration of the Texton and max-pooling layers.

C. Number of Parameters and Size on Disk

A number of models found in literature expand their width and depth in order to enhance their performance on complex images. This causes a surge in the number of parameters. For example, ResNet50 [30], AlexNet [31], and VGG16 [32], among others, generate parameter counts of 23 million, 60 million, and 138 million respectively. The intricacy of a model directly correlates with the parameters it generates, resulting in a substantial computational load that strains the resources of a system. Consequently, this poses a constraint on the feasibility of deploying such models on devices with limited memory, such as mobile phones. The comparison of the parameters of the models as well as size on disk is found in Table III. It can be seen that the size of the model on disk is small and less parameters were generated by the proposed model. This makes the proposed model suitable for deployment on mobile devices.

D. Model Interpretability

The inner workings of deep learning models are often labeled as black boxes. In order to rely on and employ these models for important functions, such as in the field of healthcare, it is essential that both the operations within the models and the results they produce are explainable. Through the utilization of saliency maps, mathematical models, activation maps, and similar techniques, explainable neural networks [33][34] and model interpretability [29] approaches aid in revealing insights into the operations occurring within the inner layers of deep learning models.

TABLE I. PERFORMANCE METRICS ON THE BRAIN TUMOR DATASET FOR THE PROPOSED AND BASELINE CAPSNET MODELS

Model (Dataset)	Class	TP	FP	TN	FN	Precision	Sensitivity	Specificity	Accuracy	F1-Score	Data Size
Baseline (Brain Tumor)	0	271	16	995	29	0.9443	0.9033	0.9842	96.57%	0.9235	300
	1	283	30	975	23	0.9042	0.9248	0.9702	95.96%	0.9144	306
	2	296	6	1005	4	0.9801	0.9867	0.9941	99.24%	0.9834	300
	3	405	4	902	0	0.9902	1	0.9956	99.70%	0.9951	405
Proposed (Brain Tumor)	0	287	10	1001	13	0.9663	0.9569	0.9872	98.25%	0.9616	300
	1	289	12	993	17	0.9601	0.9444	0.9832	97.79%	0.9522	306
	2	299	9	1002	1	0.9708	0.9967	0.9990	99.24%	0.9836	300
	3	405	0	906	0	1	1	1	100%	1	405

TABLE II. ABLATION STUDY RESULTS

Layers	Validation accuracy %
-texton	95.04
-alley_1_conv1	96.49
-alley_2_conv1	96.11
-alley_3_conv1	96.11
-alley_1_conv2	97.48
-alley_2_conv1	97.41
-alley_3_conv1	97.41
+ all layers	97.64

TABLE III. COMPARISON OF PARAMETERS OF MODELS AND SIZE ON DISK

Model	Trainable Parameters	Non-Trainable Parameters	Size on disk
Baseline CapsNet model	10,127,104	0	38.6MB
Proposed model	4,834,532	960	18.5MB

E. Visualization of Activation Maps and Clusters

Here, comparison of activation maps and clusters from the proposed and baseline models are done. This help to know the model that extract more important features from input images. Proposed model features from the Texton layer, extracted by one of the first separable convolutions is shown in Fig. 9. *1st row image 1* and baseline model features extracted by the convolution layer is shown in Fig. 9. In *2nd row image 1*, insufficient features were extracted. This exhibit that, the baseline convolution layer alone is not enough to extract important features. This inability of the convolution layer of the baseline model affected the primary capsule layer since it did not extract more important necessary to make differentiation between capsules, whereas the proposed model convolution layer extracted better edge and textural features from the Texton layer, hence its primary capsule produced better activation maps as seen in Fig. 9 *1st row image 2* than the baseline PC activation maps seen in Fig. 9 *2nd row image 2*. These visualizations of layers help to improve the understandability of the inner workings of the black box models and contributes to explainable Artificial Intelligence [35][36][37].

The technique of t-distributed stochastic neighbor embedding (tsne) [38] [39], was employed to visually represent the distinctness of clusters formed within the class capsule layer of the models. The suggested model displays noticeable groupings in contrast to the clusters formed by the baseline model. While a few outliers are evident in both the suggested and baseline model clusters, the outliers in the suggested model remain relatively close to their respective clusters. This highlights the effective discriminatory capability

of the suggested model in comparison to the baseline model as it can be seen in Fig. 10.

F. Prediction and Reconstruction

The process of determining the likely class of an input image and determining if there is a strong likelihood for that categorization is achieved through the application of a reconstruction method. In light of this, this research showcases reconstructed images of Brain Tumor using the decoder network for both models. The images generated by the proposed model exhibit slightly improved visual quality and demonstrate higher class identification and emphatic probabilities per class compared to the images produced by the baseline model, as observed in Fig. 11.

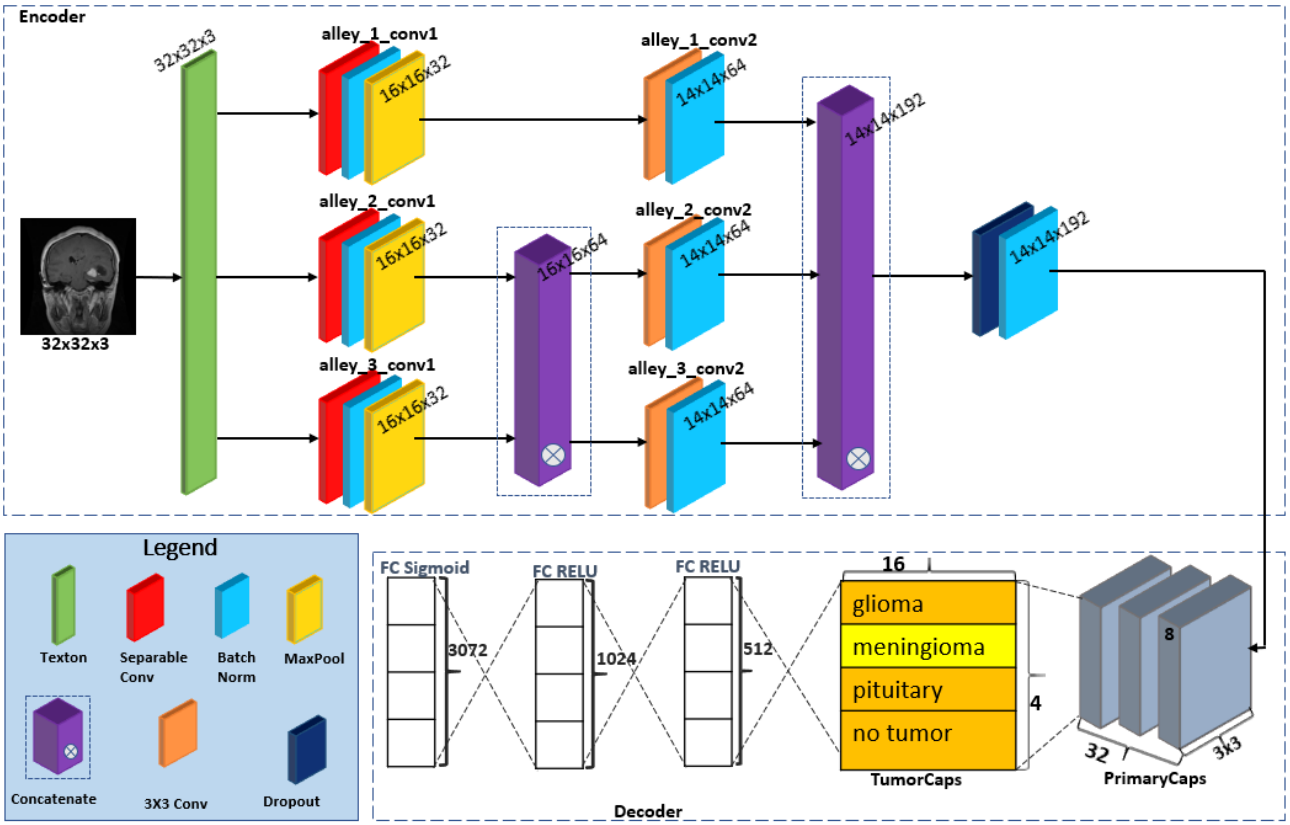

Fig. 5. Architecture of the proposed CapsNet model.

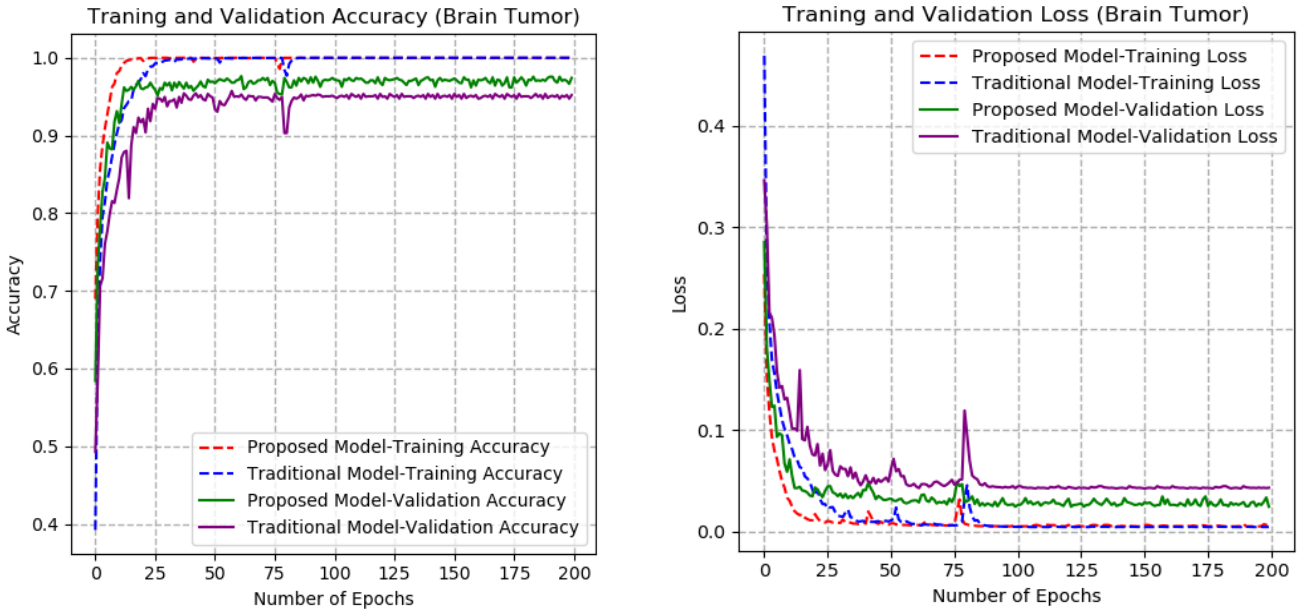

Fig. 6. Accuracy and Loss graphs of the proposed and baseline CapsNet models.

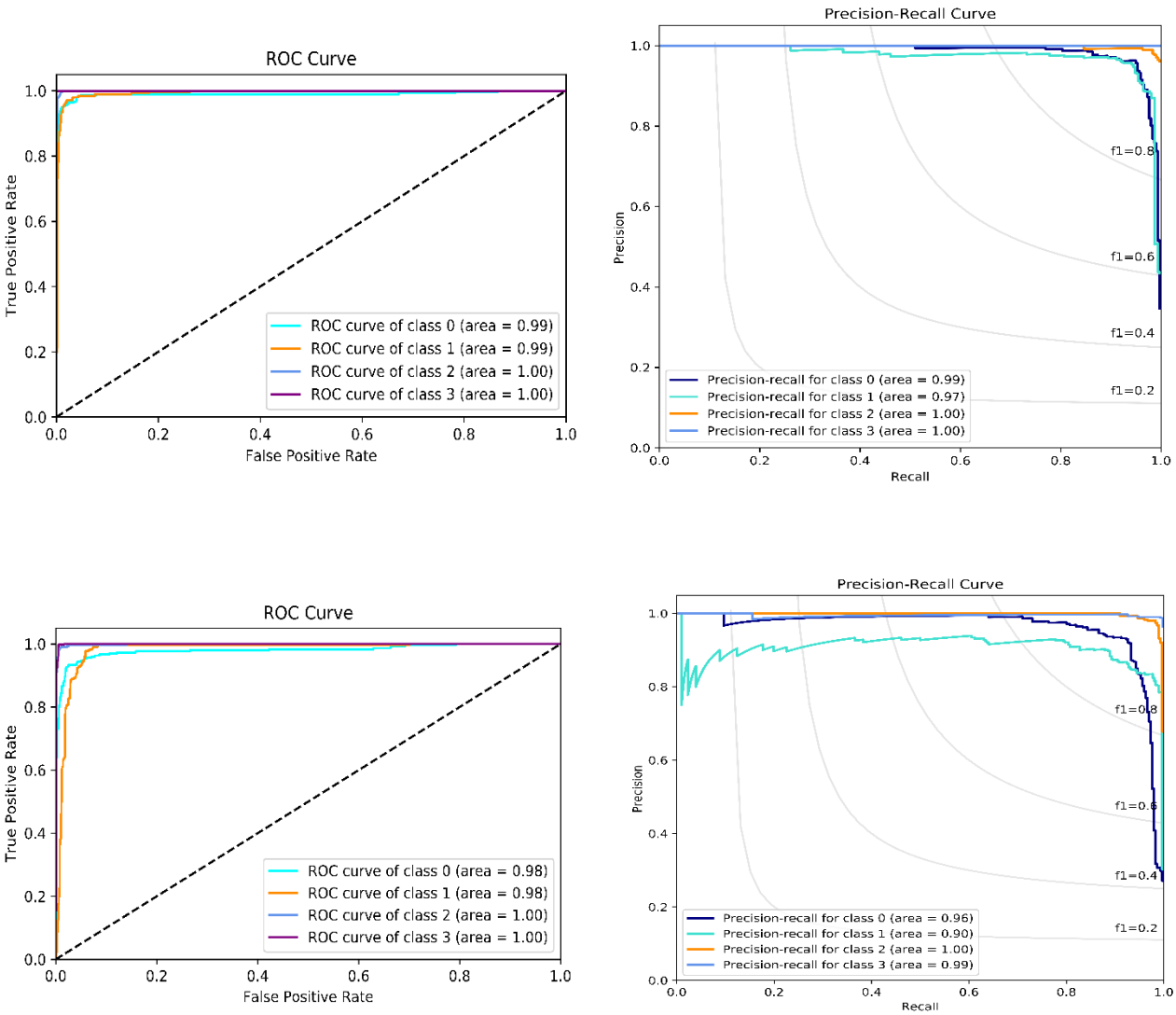

Fig. 7. ROC and PR curves for the (**1st row**) proposed and (**2nd row**) baseline models.

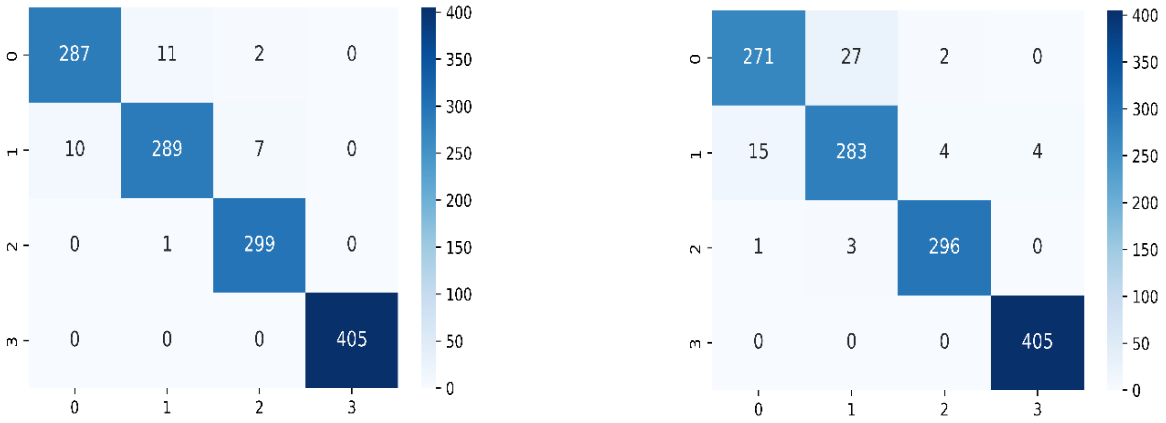

Fig. 8. Confusion Matrices of (**left**) proposed and (**right**) baseline models.

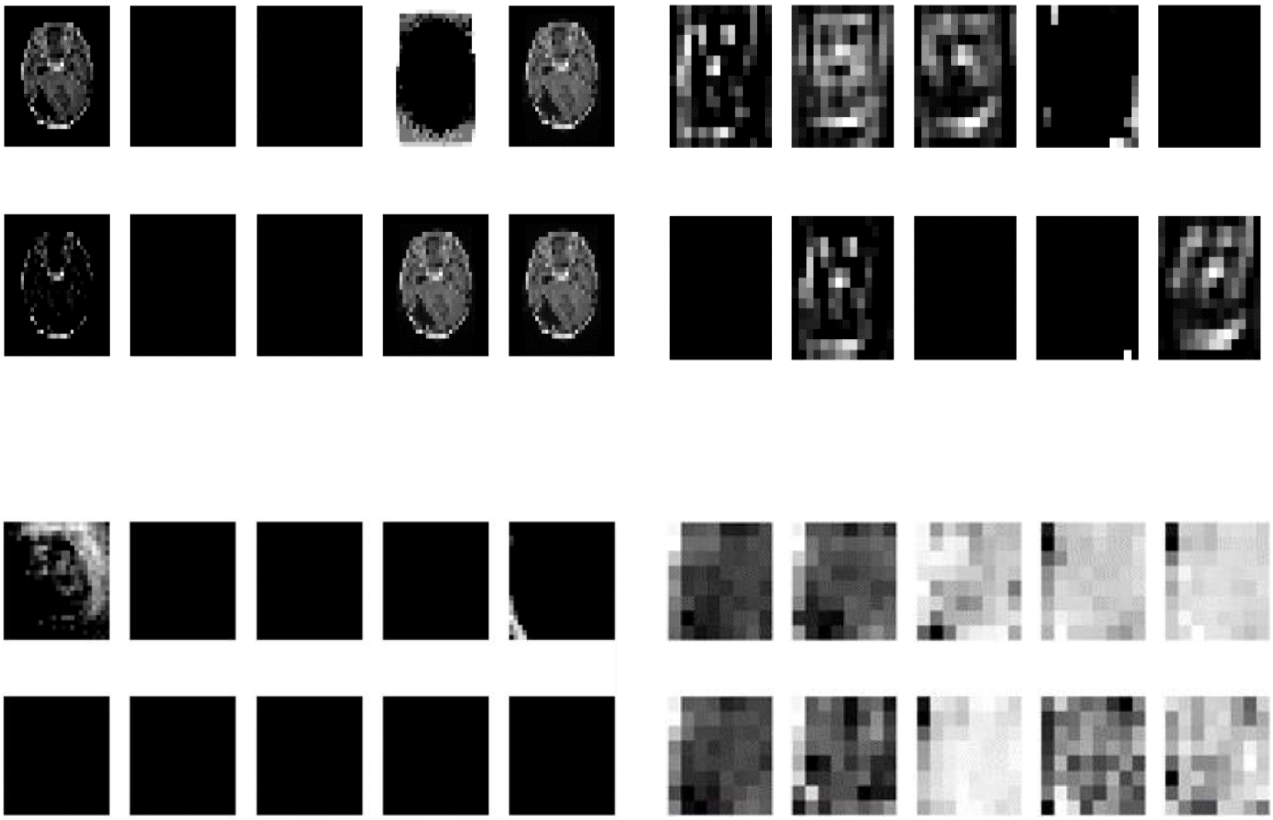

Fig. 9. Activation maps from the proposed and baseline convolution and primary capsule layers.

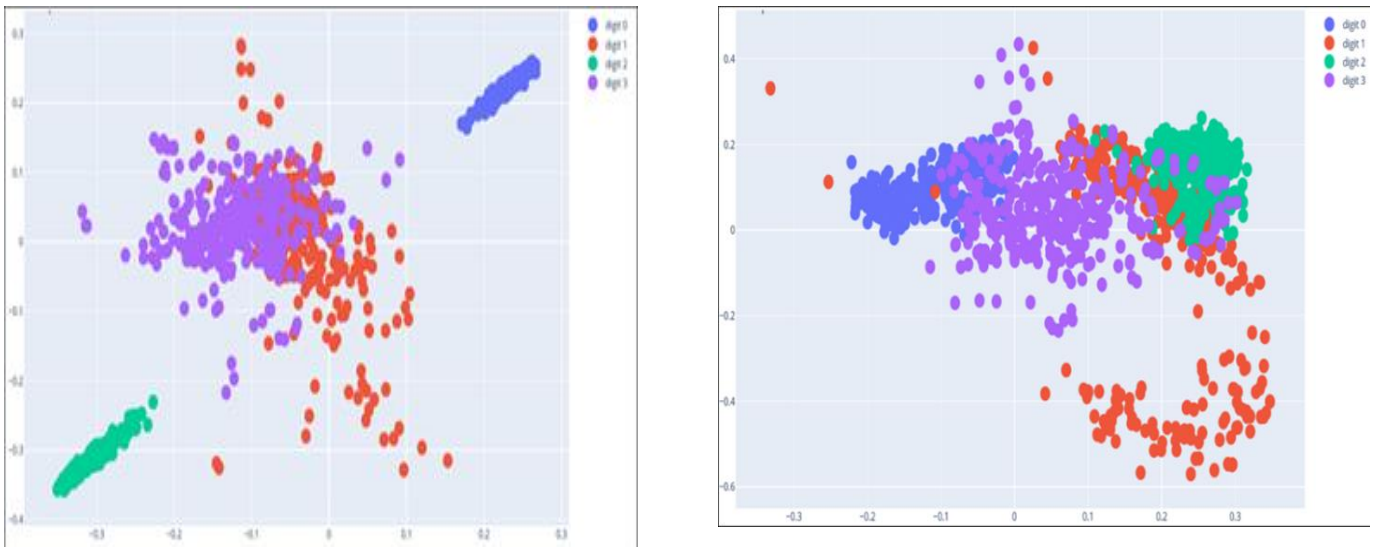

Fig. 10. Clusters obtained by the (**left**) proposed and (**right**) baseline models class capsules.

G. Comparison of Results

To demonstrate the effectiveness of our novel approach, we conducted a performance comparison between our model and cutting-edge models on brain tumor datasets. Our modifications primarily center around the structure of capsule network, specifically focusing on dynamic routing. Although our main emphasis was on dynamic routing, we extended our investigation to encompass multiple routing techniques. The outcomes, as detailed in Table IV, indicate that our model's performance matches that of the current state-of-the-art capsule network models. The commendable performance achieved by our proposed model in medical image diagnosis can be attributed to its adeptness in extracting pertinent information from diverse images, which contributes to its capability in achieving accurate results.

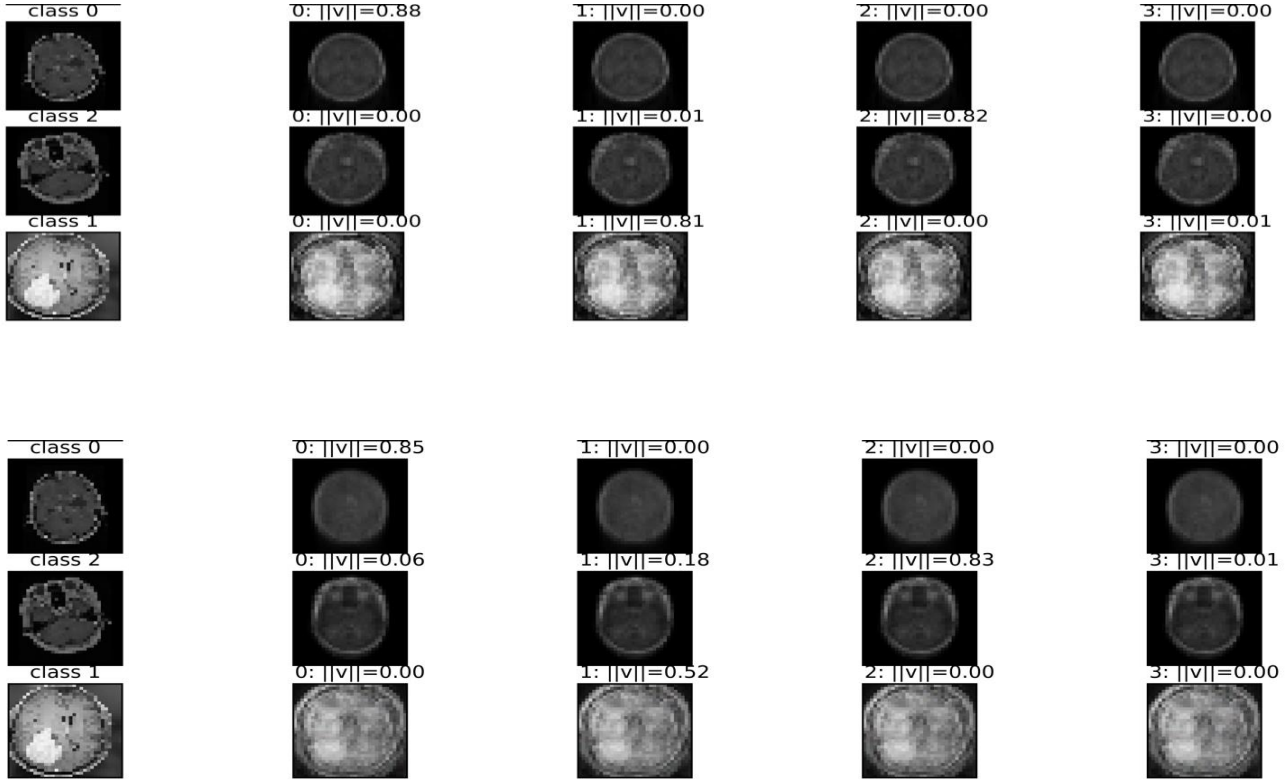

Fig. 11. Reconstructed images (top) proposed and (down) baseline model.

TABLE IV. PROPOSED MODEL AND PREVIOUS WORKS COMPARISON ON BRAIN TUMOR DATASET

CapsNet Methods	Validation accuracy (%)
Baseline [5]	95.73
BoostCaps [16]	92.45
DCNet and DCNet++ [17]	93.04 and 95.03
MLAF-CapsNet[40]	93.40 and 96.60
Vimal Kurup et al.[41]	92.60
Afshar et al. [19]	90.89
Dilated CapsNet. [20]	95.54
BayesCaps[21].	73.9
Proposed Model	**97.64**

V. CONCLUSION AND FUTURE WORKS

This study introduced a novel architecture that utilizes less time to train with less parameters, small size on disk, and proficient feature extraction capabilities, named Texton Tri-alley Separable Feature Merging (TTSFM) CapsNet, utilizing a capsule network approach, aimed at the detection of brain tumors. Texton layer helps to extract important features from input image and the separable convolutions coupled with the use of less filters and kernel sizes resulted in using less amount of time for training, small size on disk, and a smaller number of trainable parameters. These components and properties lead to the appreciable performance of the proposed model, making the model deployable on devices with lower memory like mobile devices. We went on to enhance the model's interpretability and practical usability by conducting thorough analyses, including extensive visualization of layer activation maps, examination of feature clusters, and performing ablation study.

In future, our focus will be on improving the performance of the model, and conducting in-depth experiments using medical datasets to advance the field of explainable artificial intelligence (XAI). Our objective is to remove all uncertainties from the outcomes of the models, ensuring that both professionals and other users can trust the reliable application of these models in disease diagnosis.

REFERENCES

[1] D. N. Louis et al., "The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary," Acta Neuropathol., vol. 131, no. 6, pp. 803–820, Jun. 2016, doi: 10.1007/s00401-016-1545-1.

[2] K. D. Miller et al., "Cancer treatment and survivorship statistics, 2022," CA. Cancer J. Clin., vol. 72, no. 5, pp. 409–436, 2022, doi: 10.3322/caac.21731.

[3] A. F. M. SAIF, C. SHAHNAZ, W.-P. ZHU, and M. O. AHMAD, "Abnormality Detection in Musculoskeletal Radiographs Using Capsule Network," IEEE Access, vol. 7, pp. 81494–81503, 2019, doi: 10.1109/ACCESS.2019.2923008.

[4] N. Tajbakhsh et al., "Convolutional Neural Networks for Medical Image Analysis: Full Training or Fine Tuning?," IEEE Trans. Med. Imaging, vol. 35, no. 5, pp. 1299–1312, 2016, doi: 10.1109/TMI.2016.2535302.

[5] S. Sabour, N. Frosst, and G. E. Hinton, "Dynamic Routing Between Capsules," Appl. Biosaf., vol. 22, no. 4, pp. 185–186, 2017, doi: 10.1177/1535676017742133.

[6] X. Zhang et al., "Real-time gastric polyp detection using convolutional neural networks," PLoS One, vol. 14, no. 3, pp. 1–16, 2019, doi: 10.1371/journal.pone.0214133.t005.

[7] A. Singh, S. Sengupta, and V. Lakshminarayanan, "Explainable deep learning models in medical image analysis," J. Imaging, vol. 6, no. 6, pp. 1–19, 2020, doi: 10.3390/JIMAGING6060052.

[8] G. E. Hinton, A. Krizhevsky, and S. D. Wang, "Transforming auto-encoders," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 6791 LNCS, no. PART 1, pp. 44–51, 2011, doi: 10.1007/978-3-642-21735-7_6.

[9] M. Kwabena Patrick, A. Felix Adekoya, A. Abra Mighty, and B. Y. Edward, "Capsule Networks – A survey," J. King Saud Univ. - Comput. Inf. Sci., vol. 34, no. 1, pp. 1295–1310, 2022, doi: 10.1016/j.jksuci.2019.09.014.

[10] G.-H. Liu, L. Zhang, Y.-K. Hou, Z.-Y. Li, and J.-Y. Yang, "Image retrieval based on multi-texton histogram," Pattern Recognit., vol. 43, no. 7, pp. 2380–2389, Jul. 2010, doi: 10.1016/j.patcog.2010.02.012.

[11] X. W. Gao, R. Hui, and Z. Tian, "Classification of CT brain images based on deep learning networks," Comput. Methods Programs Biomed., vol. 138, pp. 49–56, 2017, doi: 10.1016/j.cmpb.2016.10.007.

[12] F. Özyurt, E. Sert, E. Avci, and E. Dogantekin, "Brain tumor detection based on Convolutional Neural Network with neutrosophic expert maximum fuzzy sure entropy," Meas. 2019, vol. 147, 2019, doi: 10.1016/j.measurement.2019.07.058.

[13] M. Sajjad, S. Khan, K. Muhammad, W. Wu, A. Ullah, and S. Wook, "Multi-grade brain tumor classification using deep CNN with extensive data augmentation," J. Comput. Sci., vol. 30, pp. 174–182, 2019, doi: 10.1016/j.jocs.2018.12.003.

[14] W. Ayadi, W. Elhamzi, I. Charfi, and M. Atri, "Deep CNN for Brain Tumor Classification," Neural Process. Lett., vol. 53, no. 1, pp. 671–700, 2021, doi: 10.1007/s11063-020-10398-2.

[15] M. M. Badža and M. c Barjaktarovic, "Classification of Brain Tumors from MRI Images Using a Convolutional Neural Network," Appl. Sci., 2020.

[16] P. Afshar, K. N. Plataniotis, and A. Mohammadi, "BoostCaps : A Boosted Capsule Network for Brain Tumor Classification," 2020 42nd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc., pp. 1075–1079, 2020.

[17] S. S. R. Phaye, A. Sikka, A. Dhall, and D. Bathula, "Dense and Diverse Capsule Networks: Making the Capsules Learn Better," Comput. Vis. Pattern Recognit., pp. 1–11, May 2018, [Online]. Available: http://arxiv.org/abs/1805.04001

[18] E. Goceri, "CapsNet topology to classify tumours from brain images and comparative evaluation," IET Image Process., vol. 14, no. 5, pp. 882–889, 2020, doi: 10.1049/iet-ipr.2019.0312.

[19] P. Afshar, A. Mohammadi, and K. N. Plataniotis, "Brain tumor type classification via capsule networks," 2018 25th IEEE Int. Conf. Image Process., 2018.

[20] K. Adu, Y. Yu, J. Cai, and N. Tashi, "Dilated Capsule Network for Brain Tumor Type Classification Via MRI Segmented Tumor Region," Proceeding IEEE Int. Conf. Robot. Biomimetics Dali, China, December 2019, no. December, pp. 942–947, 2019.

[21] P. Afshar, A. Mohammadi, and K. N. Plataniotis, "BayesCap : A Bayesian Approach to Brain Tumor," IEEE Signal Process. Lett., vol. 27, pp. 2024–2028, 2020.

[22] B. Julesz, "Texton Gradients: The Texton Theory Revisited," vol. 251, pp. 245–251, 1986.

[23] F. Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions," Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2017, pp. 1251-1258, 2017, doi: 10.4271/2014-01-0975.

[24] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size," Comput. Vis. Pattern Recognit., pp. 1–13, 2016, [Online]. Available: http://arxiv.org/abs/1602.07360

[25] C. Zhang, S. Bengio, M. Hardt, B. Recht, and O. Vinyals, "Understanding deep learning (still) requires rethinking generalization," Commun. ACM, vol. 64, no. 3, pp. 107–115, 2021, doi: 10.1145/3446776.

[26] F. Provost, T. Fawcett, and R. Kohavi, "The case against accuracy estimation for comparing induction algorithms," Int. Conf. Mach. Learn., p. 445, 1998.

[27] P. Singla and P. Domingos, "Discriminative training of Markov logic networks," Proc. Natl. Conf. Artif. Intell., vol. 2, pp. 868–873, 2005.

[28] R. Meyes, M. Lu, and T. Meisen, "Ablation Studies to Uncover Structure of Learned Representations in Artificial Neural Networks," Proc. Int. Conf. Artif. Intell. (pp. 185-191). Steer. Comm. World Congr. Comput. Sci. Comput. Eng. Appl. Comput. (WorldComp)., pp. 185–191, 2019.

[29] R. Meyes, M. Lu, C. W. De Puiseau, and T. Meisen, "Ablation Studies in Artificial Neural Networks," Comput. Vis. Pattern Recognit., pp. 1–19, 2019.

[30] K. He, X. Zhang, S. Ren, and J. Sun, "Deeper neural networks are more difficult to train.," Comput. Vis. pattern Recognit., vol. 37, no. 50, pp. 1951–1954, Dec. 2016, [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1002/chin.200650130

[31] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," Proc. Adv. Neural Inf. Process. Syst. 25 (NIPS 2012), pp. 1–1432, 2012, doi: 10.1201/9781420010749.

[32] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc., pp. 1–14, 2015.

[33] G. Ras, N. Xie, M. Van Gerven, and D. Doran, "Explainable Deep Learning: A Field Guide for the Uninitiated," J. Artif. Intell. Res., vol. 73, pp. 329–397, Jan. 2022, doi: 10.1613/jair.1.13200.

[34] P. Angelov and E. Soares, "Towards explainable deep neural networks (xDNN)," Neural Networks, vol. 130, pp. 185–194, Oct. 2020, doi: 10.1016/j.neunet.2020.07.010.

[35] W. Samek and K.-R. Müller, "Towards Explainable Artificial Intelligence," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 11700 LNCS, 2019, pp. 5–22. doi: 10.1007/978-3-030-28954-6_1.

[36] A. Shahroudnejad, P. Afshar, K. N. Plataniotis, and A. Mohammadi, "IMPROVED EXPLAINABILITY OF CAPSULE NETWORKS: RELEVANCE PATH BY AGREEMENT," in 2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP), Nov. 2018, pp. 549–553. doi: 10.1109/GlobalSIP.2018.8646474.

[37] D. Gunning, M. Stefik, J. Choi, T. Miller, S. Stumpf, and G.-Z. Yang, "XAI—Explainable artificial intelligence," Sci. Robot., vol. 4, no. 37, Dec. 2019, doi: 10.1126/scirobotics.aay7120.

[38] P. Hajibabaee, F. Pourkamali-Anaraki, and M. A. Hariri-Ardebili, "An Empirical Evaluation of the t-SNE Algorithm for Data Visualization in Structural Engineering," in 2021 20th IEEE International Conference on Machine Learning and Applications (ICMLA), Dec. 2021, pp. 1674–1680. doi: 10.1109/ICMLA52953.2021.00267.

[39] S. Arora, W. Hu, and P. K. Kothari, "An Analysis of the t-SNE Algorithm for Data Visualization," vol. 75, no. 2008, pp. 1–8, Mar. 2018, [Online]. Available: http://arxiv.org/abs/1803.01768

[40] K. Adu, Y. Yua, J. Caia, P. K. Mensah, and K. Owusu-Agyemang, "MLAF-CapsNet: Multi-lane atrous feature fusion capsule network with contrast limited adaptive histogram equalization for brain tumor classification from MRI images," J. Intell. Fuzzy Syst., 2021, doi: 10.3233/JIFS-202261.

[41] R. Vimal Kurup, V. Sowmya, and K. P. Soman, "Effect of Data Pre-processing on Brain Tumor Classification Using Capsulenet," ICICCT 2019 – Syst. Reliab. Qual. Control. Safety, Maint. Manag., pp. 110–119, 2020, doi: 10.1007/978-981-13-8461-5_13.