Journal of Theoretical and Applied Information Technology
30th June 2025. Vol.103. No.12

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

EVALUATING PREDICTIVE CERTAINTY IN AI MODELS FOR ACCURATE BRAIN TUMOR DETECTION

[1]GOKAPAY DILIP KUMAR, [2]SHAIK THASEENTAJ, [3]KUNDA SURESH BABU, [4]JELDI ASHA JYOTHI, [5]NUTHALAPATI KAMALA VIKASINI, [6]PREM SWARUP MALLIPUDI

[1]Research Scholar, School of Computer Science & Engineering, VIT-AP University, Amaravati, Near Vijayawada City, Andhra Pradesh, India.
[2]Assistant Professor, Department of Computer Science & Engineering, KL University (KLEF), Green Fields, Vaddeswaram, Guntur District, Andhra Pradesh, India.
[3]Associate Professor, Department of CSE, Narasaraopeta Engineering College (Autonomous), Narasaraopet, Palnadu(Dist)-522601, Andhra Pradesh, India.
[4]Assistant Professor, Department of Computer Science & Engineering, Vignan Bharathi Institute of Technology (VBIT), Aushapur (V), Ghatkesar (M), Medchal Dist, Hyderabad, Telangana, India.
[5]Assistant Professor, Swarna Bharathi Institute of Science & Technology, Pakabanda Street, Khammam - 507 002, Telangana, India.
[6]Research Scholar, School of Computer Science & Engineering, VIT-AP University, Amaravati, Near Vijayawada City, Andhra Pradesh, India.

E-Mail: [1]gokapay.22phd7117@vitap.ac.in, [2]thaseentaj235@gmail.com, sureshkunda546@gmail.com, [4]jaldiasha@gmail.com, [5]vikasini574@gmail.com, [6]prem.24phd7009@vitap.ac.in,

Corresponding Author: Shaik Thaseentaj[2]

ABSTRACT

Brain tumors are highly dangerous and often life-threatening, significantly affecting patients' overall health and quality of life. This research explores prediction certainty, an underrepresented area, as existing research focuses on accuracy. This study highlights establishing a correlation between a model's loss value and greater certainty in predictions. Along with conventional performance metrics such as precision and recall, this work emphasizes the critical role of loss value. To assess the reliability and effectiveness of artificial intelligence models, including CNN, ResNet-50, XceptionNet, and a proposed model (integrating advanced layers), were tested. The study prioritized loss values for accurate detection of tumor cases, minimizing false negatives. The models are effective for real-time tumor detection due to their low loss values and efficient runtimes. Experimental results showed the following metrics on testing: CNN achieved a loss of 0.35 and 68.60% accuracy; ResNet-50 achieved improved performance with a loss of 0.17; and the proposed model achieved 90% accuracy with superior recall and runtime. The study concludes that while accuracy is important, the certainty in predictions plays a significant role in reliable tumor diagnosis. Given the global shortage of specialized medical professionals, the proposed approach addresses this gap by providing timely and accurate cancer detection tools, contributing effectively to healthcare systems and medical education, enabling future AI applications to be used effectively in clinical practice.

Keywords: *Brain Tumor Detection, Artificial Intelligence, Cnn, Resnet-50, Xceptionnet.*

1. INTRODUCTION

1.1. Background

The high fatality rate associated with brain tumors makes them one of the most severe diseases [1], and their prevalence continues to rise. In 2024, approximately 19,000 new brain tumor cases were projected according to Cancer Statistics 2024 [2]. The increasing prevalence of technology, especially the extensive use of mobile devices, has been linked to a growing concern among younger generations, with brain cancer identified as one potential negative outcome [3]. Significant factors contributing to this issue include exposure to ionizing radiation and genetic susceptibility [4].

Figure 1 illustrates the ionizing radiation emitted by various devices.

The formation of a brain tumor can be categorized into four separate stages. In the initial

Journal of Theoretical and Applied Information Technology
30th June 2025. Vol.103. No.12

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

stage, the tumor shows a slow and steady progression.. And it may be fully remedied by entirely removing the afflicted region. That is a pilocytic astrocytoma. During the second stage, cancer has the potential to progress to more severe grades, but the pace of proliferation is relatively slow. These tumours might develop despite the patient's adherence to therapy. During stage three, the cancer spreads more rapidly to the adjacent tissues. These tumours need adjuvant chemotherapy and radiotherapy after surgery since they cannot be eradicated alone by surgical intervention. Stage four is the most dangerous phase and rapidly spreads. This type may use blood arteries to accelerate growth [6]. Timely identification and classification of brain abnormalities and tumors are essential for successful treatment and improved patient outcomes. [7]. Detecting brain tumours is tough due to several factors such as tumour shape, size, appearance, location, scanning settings, and modalities [8]. Traditional and intelligence strategies are employed to accomplish this objective. Traditional methods such as Leksell Gamma Knife, Gamma Knife (GK), and Radioactive rays may diagnose lesions, but they need human intervention and are time-consuming. To identify brain tumours, many medical imaging methods like CT, MRI, and PET-CT are used. MRI is a non-invasive method that employs magnetic fields and microwave pulses to visualize internal structures of the body. Three primary MRI techniques—FLAIR, T1-weighted, and T2-weighted imaging—are utilized for the diagnosis of brain tumors. Proper identification of tumor-affected areas through MRI is crucial for accurate assessment [9].

The research community has begun exploring the certainty of AI model predictions in medical imaging. Studies like [10] have showed the effectiveness of hybrid DL approaches for brain tumor classification, showcasing the potential of advanced AI techniques in real-time diagnostics.

The complexity of MRI images prevents the human visual system from detecting minute changes.

Recently, researchers developed CAD tools to help Imaging professionals in accurate diagnosis. The Leksell Gamma Knife can diagnose tumours, although brain necrosis may affect outcomes. Efficient ML is needed to address this challenge. In [11], authors presented an RF classifier-voxel clustering algorithm method. In [12], Researchers have applied unsupervised FCM clustering algorithms to partially automate the separation of lesion volumes, although Leksell Gamma Knife diagnostics remain time-intensive. For brain tumor segmentation, methods such as K-means, Fuzzy K-means, GMM, and GHMRF are recommended [13].

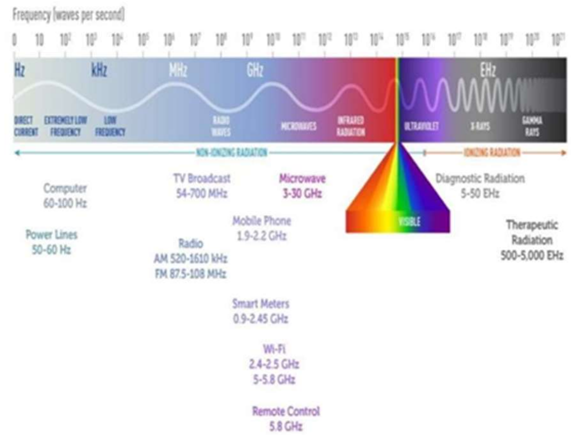


Figure 1: Electromagnetic waves and the devices that emit these waves [14]

1.2. Investigative Contributions

Early identification and categorization of brain tumours is a vital field in medical imaging. This research aids in determining the most suitable treatment approach to potentially save patients' lives. While ML methods work well for MRI tumour detection, DL models are producing far better results. This research intends to provide the following to help readers understand how DL methods perform. This study introduces a novel approach to brain tumor detection by combining accuracy with interpretability through certainty scoring based on loss metrics. Unlike prior models focused solely on correctness, our method offers detailed prediction confidence, enhancing reliability in clinical settings. Experiments across multiple architectures validate this unique contribution.

1. The proposed study utilizes various pretrained deep learning (DL) algorithms to detect and classify brain tumor based on MRI scans.
2. The proposed approach research study was conducted on over one hundred papers sourced from diverse sources such as ScienceDirect, Springer, IEEE, and others.
3. The research study focused on the research gap “certainty in the model responses [15]” by emphasizing the loss value, recall and running time of the models.
4. To incorporate DL techniques into the current pedagogical framework, a research gap analysis was conducted.
5. A research study highlights that “fully automated methods for brain tumor detection are currently lacking [13].” In this work, we will illustrate how to develop a completely

automated system by utilizing predictions from various models.

The structure of the paper is as continues with: Section I presents the introduction, comes after Section II, which discusses literature review. The article has a distinct part that encompasses a range of works classified under the domains of DL present in Section III. The different findings/results are presented in Section IV. The document concludes with a final section, which is then followed by a list of references. The paper reviews models like CNN, ResNet, and EfficientNet, noting their focus on accuracy. Unlike these, our study uniquely evaluates model certainty by analyzing loss values and certainty scores. This approach quantifies prediction trustworthiness, providing a new evaluation perspective essential for high-risk medical diagnostics such as brain tumor detection.

1.3. Article Structure

This work presents structured to address both the technical and practical aspects of evaluating AI models for brain tumor detection, with an emphasis on prediction certainty. Following the introduction, which outlines the motivation and context of the study, we delve into the detailed literature review and current research in ML and DL applications in brain tumor categorization. The research gap is then explored, highlighting the limitations of existing models in terms of predictive certainty. In the subsequent sections, we present the proposed methodology, which emphasizes the correlation between loss values and model certainty, using experiments with pre-trained models like CNN, PROPOSED MODEL, ResNet, and XceptionNet. The results section focuses not only on accuracy metrics but also on how loss values impact model confidence in predictions. The discussion interprets these findings in light of their practical applications, especially in real-time medical diagnostics. In conclusion, the article emphasizes the importance of incorporating AI models into healthcare, highlighting the need to balance predictive reliability with exceptional accuracy, and suggests potential pathways for future research advancements. The study began after reviewing over 100 articles from major databases, identifying a gap in brain tumor detection research. Most existing models emphasize accuracy but overlook prediction certainty. This absence of loss-certainty evaluation motivated the research, aiming to enhance diagnostic reliability in clinical decision-making.

2. REVIEW OF RELATED WORK

This section offers a summary of the background material, definition, and relevant research in the topic of brain tumours. The Research Deficiency and Related Work sections reveal that existing methods overlook model prediction confidence. Our study addresses this gap by introducing certainty scores derived from loss functions, enhancing the safety and reliability of AI in healthcare. The connection between literature gaps and our contributions is clearly established throughout the paper.

2.1. Brain Tumour

A brain tumor is a mass or collection of irregular cells located within the brain. The cranial structure, which encompasses the cerebral organ, has a high degree of rigidity. Any proliferation inside such a confined area might lead to concerns. Brain tumour may be separated into malignant (cancerous) and benign (noncancerous) categories. Benign or malignant tumors may lead to intracranial hypertension as they develop, resulting in increased pressure inside the skull. This may result in cerebral impairment, and it has the potential to be fatal. Although benign tumors aren't aggressive, they may nevertheless affect other parts of the brain as they grow. Just Around 30% of brain tumors are cancerous, whereas 70% are completely benign. This corresponds a significant disparity. Hundred and twenty distinct categories of brain tumors are known to the age. The most common types are those affecting the meninges, which influence both the brain and spinal cord. The glioma tumor is caused by astrocytes, which are another name for glial cells. This kind grows at a slower rate and poses less of a threat. Pituitary tumors are another kind of brain tumor that may develop in the Pituitary part of the brain [16]. Biopsy, the study of cerebrospinal fluid (CSF), and X-ray imaging are some of the ways using which medical experts find brain tumours. During a biopsy, a small piece of the affected tissue is surgically removed and analyzed for diagnostic purposes. With a biopsy, only 49.1% of cases are detected [17]. Inflammation and excessive bleeding are consequences of surgical operations that patients have to endure [18]. CSF is a transparent fluid that is used by medical experts for brain tumor detection. Similar to biopsy, it also comes with some side effects, i.e., allergic reactions and excessive bleeding, etc. [19]. X-rays have the potential risk of increasing the growth of cancer cells as shown in Figure 1.

Medical imaging techniques have

substantially reduced the risk factor for the patients. Various methods were introduced for getting medical images, i.e., X-ray, MRI, CT scanning, and ultrasound. MRI gained popularity due to its non-invasiveness, and high quality without exposing the patients to harmful radiation [20]. CAD is used along with MRI by medical professionals. CAD improved using artificial intelligence robustness with ML and DL algorithms [21]. For instance, the hybrid DL-based approaches for brain tumor classification [10] have proven effective results in enhancing accuracy and certainty in prediction. Furthermore, recent developments in CNN-based frameworks have also contributed to The differentiation of stages in Alzheimer's disease [22], underlining the versatility of AI models in healthcare applications.

2.2. Active Research

ML algorithms heavily rely on feature selection and extraction, with good features being essential for the excellent performance of ML classifiers. Feature extraction and feature selection thus became major tasks in ML [23]. Abir et al. used Discrete Cosine Transform (DCT) for preprocessing The study utilized images along with GLCM for feature extraction, employing a PNN for image classification, resulting in an accuracy of 83.3% [24]. In another study, GLCM was also used for feature extraction, with Random Forest serving as the classifier. Chi-square and t-tests were applied to validate the experiment's results [25].

Vidyarthi et al. suggested using ML to categorize brain tumor into various groups based on their class. Real-life datasets were used, featuring five categories and a large set of features from six fields. To aid feature selection, the CVM was proposed. The classification accuracies achieved were 88.43% for KNN, 92.5% for mSVM, and 95.86% for the other method for Neural Networks (NN) in multi-class prediction using the suggested method. The NN predictor used a diverse set of traits to achieve an accuracy of 95.66% [26].

Jaeyong Kang etal. developed classification methods for brain tumor using deep features and ML classifiers. Deep features were extracted from brain MRI images through Transfer-based learning and models trained beforehand deep CNNs. The most effective features were organized and presented to classifiers. Their strategy was evaluated using three brain MRI datasets, and they found that deep feature ensembles remarkably improved performance. The SVM using the RBF kernel showed exceptional performance, especially with larger datasets [27].

Asiri et al. applied six ML classifiers to a dataset of 253 images. The data was imbalanced, comprising 98 images of healthy controls and 155 images of tumor patients. They extracted 2058 features from the images and used various classifiers, including Random Forest models, Naive Bayes approach, Neural Network techniques, SVM, as well as Decision Trees. SVM achieved the highest accuracy at 95.3% [28].

A fine-tuned version of an EfficientNet-B0 pre- trained model was employed in one study to identify brain tumors [16]. Images were enhanced using a three-step preprocessing imaging approach.

Another method to reduce overfitting was data augmentation, which added distinct traits to make the data more robust for the model to learn from. A validation accuracy of 98.87% was achieved, with the authors claiming that their model outperformed Frontline models like VGG16, Inceptionv3, XceptionNet, ResNet50, as well as InceptionResNetV2 in terms of precision. The AUC value for the proposed approach was 0.988 [16]. While the model performed well on performance measures, generalization remains limited as the dataset used was only one. Additionally, hybrid approaches like the one proposed by Raza et al. [10] show the growing potential of DL methods, combining several models for even better classification accuracy. Other works like that of Shahwar et al. [29], using hybrid classical-quantum neural networks, have proven promising for automating the detection of diseases, such as Alzheimer's.

This research indicates that DL approaches, especially those based on CNNs have shown great potential as tools for classifying brain tumors as well as other neural diseases, providing high accuracies and potential realtime applicability [22]. Further, the implementation of AI for complex tasks such as multiclass motor imagery has demonstrated strong performance in healthcare-related classifications [30].

2.3. Research Deficiency

Despite significant advancements in ML and DL approaches for detecting brain tumor and classification, key challenges remain unaddressed. Existing research, as detailed in subsections 1.1 and 2.1, has largely focused on achieving high accuracy in tumor classification but has often neglected an equally critical factor: the certainty of these predictions. While models like CNN, EfficientNet,

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

and others have demonstrated high accuracy in identifying and classifying brain tumors [16][10], there is little emphasis on how certain the models are when making these predictions, especially in high-stakes medical contexts where false positives and false negatives can have severe consequences. This gap between accuracy and predictive certainty presents a major concern in real-time clinical applications.

Furthermore, the current methodologies rely heavily Based on performance indicators like precision, recall, as well as accuracy [20][25], which, although important, do not capture the model's confidence in its predictions. This is particularly relevant when dealing with cases of overlapping symptoms differentiating benign from malignant tumors, along with other neural diseases like Alzheimer's [22]. The expanding research on DL applications in healthcare continues to highlight the need for frameworks that not only classify accurately but also provide interpretable and reliable certainty levels in their predictions [30]. Additionally, most studies—including those that use pre-trained models or hybrid approaches—often rely on small, limited datasets, which reduce the generalizability of their findings [16][10]. This limitation makes it difficult to deploy these models across different institutions with varying imaging protocols. Moreover, existing methods focus on optimizing model performance through feature selection and augmentation [21][22], but few explore how models behave under uncertain or ambiguous conditions, such as edge cases that do not conform to standard patterns [29].

Given these challenges, our research fills a critical gap by emphasizing not only accuracy but also the certainty of AI models in detecting brain tumors. We introduce a novel approach that correlates loss values with prediction confidence, demonstrating that lower loss values lead to more reliable predictions. As shown in subsection 2.2, while XceptionNet achieves high accuracy, it does not sufficiently account for the loss function as a metric for certainty [10][22]. By focusing on both loss and recall, our research aims to reduce the number of false negatives, ensuring that no tumor case is falsely categorized as healthy, a critical factor for real-time clinical decision-making.

Our contributions, as outlined in subsection 1.2, address the following research gaps:

We propose an approach correlating prediction certainty with model loss, bridging the current gap in evaluating AI model reliability.

We present an in-depth analysis of how lower loss values can enhance the certainty and confidence of model predictions, making them suitable for Dynamic applications.

Our study demonstrates importance of focusing on both recall and loss, Due to these metrics together provide a more holistic view of model performance in high-stakes environments like healthcare.

Thus, our work not only builds on the existing body of research but also moves beyond accuracy metrics, focusing on the reliability and certainty of Predictions to facilitate the successful application of AI models in clinical workflows

3. METHODOLOGY AND MATERIALS

This section describes the material and methods used in this research and the repository information. The paper explicitly states its problem: existing brain tumor detection models prioritize accuracy while ignoring prediction certainty, a key factor in clinical diagnostics. It explores whether loss-based certainty scores can effectively measure model confidence, how architectures differ in certainty, and if integrating certainty improves trust in AI-driven medical decisions. For the classification task, we have used the CNN and transfer learning approach, employing Resnet101, XceptionNet, and VGG19 pre-trained architectures. Fig 2 describes the block diagram of the methodology of the current study.

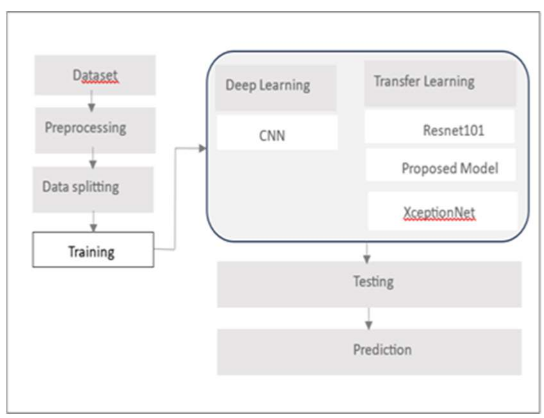


Figure 2: Block diagram of the current study

3.1. Dataset:

Utilizing DL techniques to enhance health diagnosis

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

yields significant and influential solutions. The WHO suggest that a comprehensive brain tumor diagnosis contains the identification of the tumor, determination of its location, and categorization based on factors such as malignancy, grade, and type. This study is on using MRI to pinpoint brain tumors. The primary objectives are to detect the presence of the tumor, categorize it based on its grade and kind, and identify its specific location inside the brain. This approach has been validated with a single model to classify brain MRI images across multiple tasks, rather than employing separate models for each task. The CNN is capable of doing multi-task classification, specifically to classify and detect tumors. Brain tumor location may be identified using a CNN-based model that segments the tumor.The dataset is taken from a well- known data repository" Kaggle". It combines the following three datasets:figshare ,SARTAJ dataset,Br35H,The present dataset includes MRI scans of 7,023 human brains, divided into four distinct classes:glioma, meningioma, no tumour, and pituitary. No images depicting tumour classes The data was collected from the Br35H dataset. There is an issue with the categorization of glioma class photographs in the SARTAJ dataset. This problem became apparent via the analysis of other researchers' work and the training of various models. Consequently, the doubtful images from the dataset folder are replaced with images obtained from the figshare dataset.

3.2. Descriptions of the classes

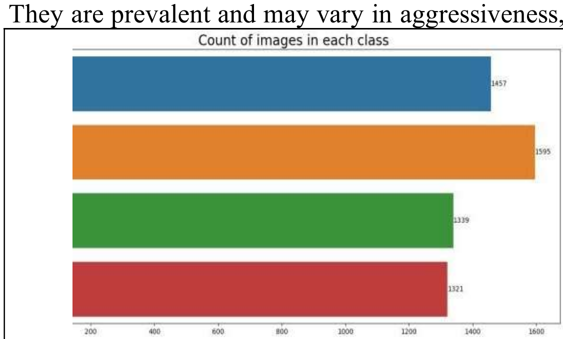
Gliomas are brain tumors that arise from glial cells. They are prevalent and may vary in aggressiveness, ranging from low-grade, less aggressive forms to high-grade, very malignant types.
Prompt and accurate categorization is crucial for efficient therapy.

Meningiomas are non-cancerous tumors originating from the meninges, the brain and spinal cord's protective coverings, but their location and size can cause significant challenges. Accurate detection is essential for successful patient management.

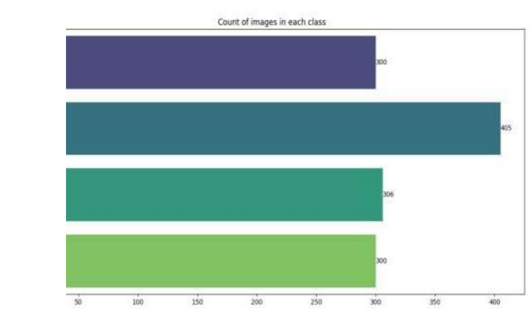
No Tumor: This category comprises MRI pictures that exhibit no discernible signs of tumor development. Accurately recognizing pictures without tumors is crucial to prevent needless treatments and preserve a high level of diagnostic precision.

Pituitary tumors develop within the pituitary gland, situated at the base of the brain.Dysregulation of hormones caused by these tumors may result in systemic complications, underscoring the need for their identification for effective medical care.

The testing dataset is divided into both validation and testing portions. In training set, number of images are 5712, and there are 656 images both in the testing and validation datasets after removing the unwanted images. Figure 3 shows the count of images in the training and testing datasets. Figure 4 Shows MRI scans from the different classes utilized in this research.



a Training dataset
)



b Testing Dataset
)

Figure 3: The number of images per class in both the training and testing datasets

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

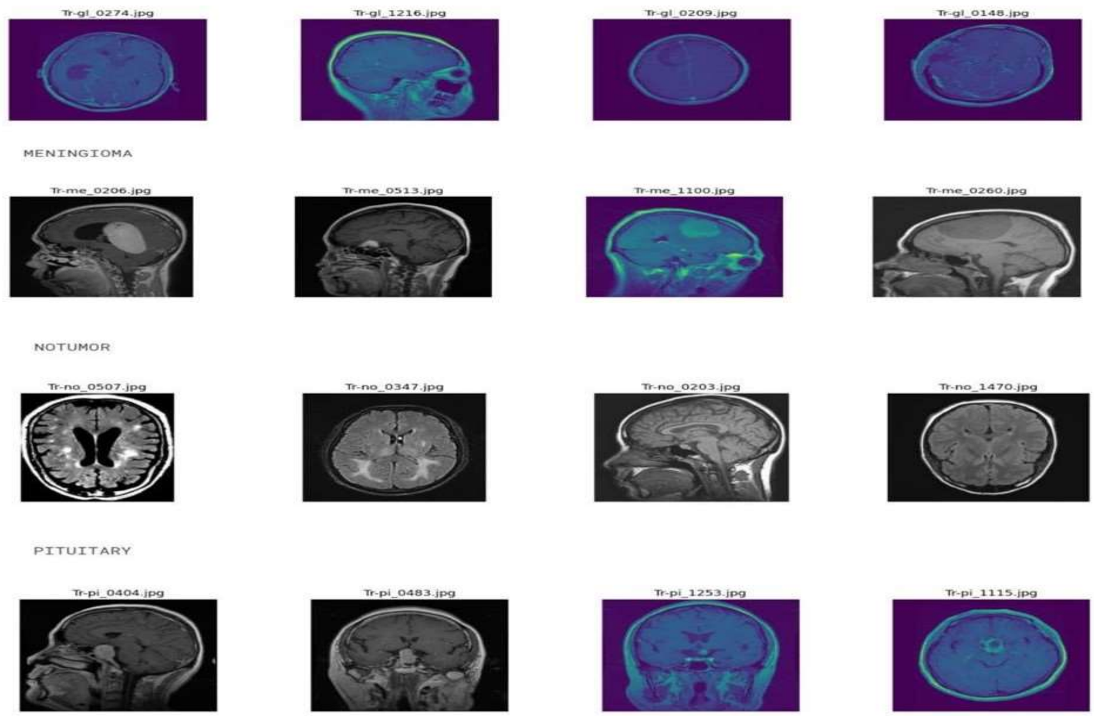


Figure 4: MRI images of distinct classes

3.3. Preprocessing

In preprocessing step data augmentation is done. Data augmentation technique artificially enlarges the training dataset by creating variations of the original data. It entails creating little alterations for the dataset, such as flipping, rotating, and shearing the data, as shown in Figure 5.

The dimensions and luminosity of pictures significantly influence the model's performance, since increased brightness enhances features, as seen in Figure 6.

3.4. AI technologies

3.4.1. Deep Learning

DL is a specialized branch within ML. Within the realm of DL, the process of extracting and selecting features is not done manually. Instead, DL algorithms are used, which rely on a large amount of data to do this task. This research uses a convolutional neural network.

3.4.2. Convolutional Neural Network

CNN was composed of multiple layers of neural networks. Due to its robustness, it is extensively used in the field of image processing [31]. In a CNN model, the image is sent in the model. On which various filters are applied. The image undergoes convolution with these filters, creating a feature map for each one [32]. These feature maps are also called convolution layers. The convolutional layer is the essence of this model, and based on this model is named as CNN. Besides this, the pooling layer was used, that is used to pool the features having maximum importance, it is done by taking the average or by picking the mx value, etc from the features layer.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

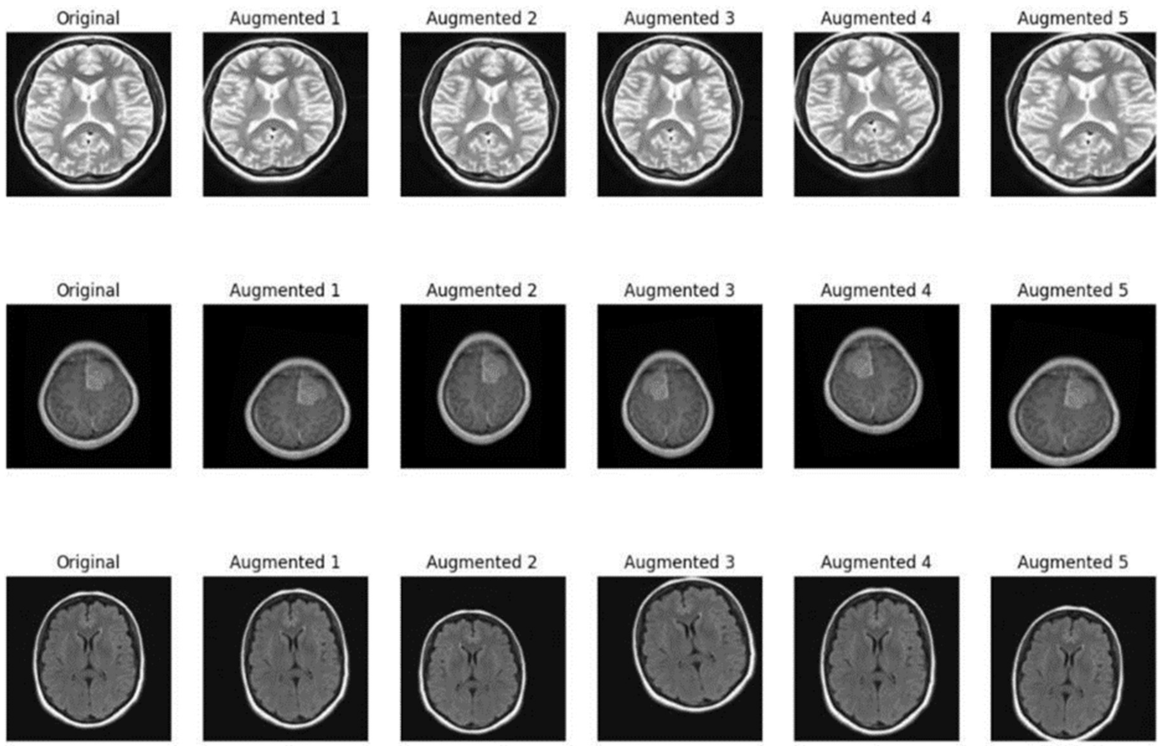
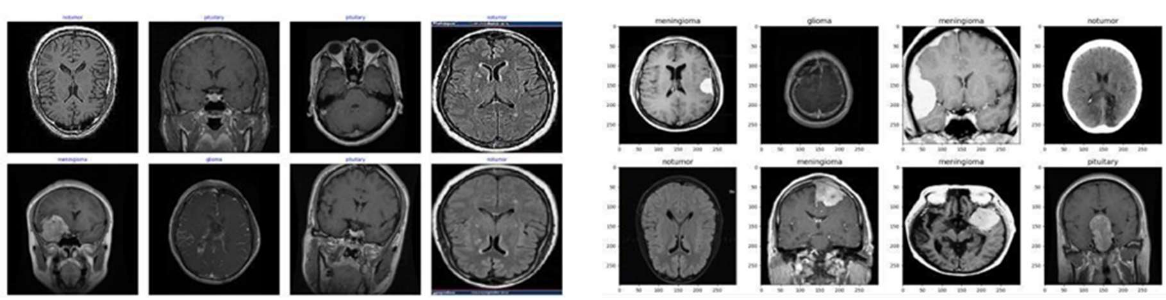


Figure 5: Augmented mages



(a) Before **(b) After**

Figure 6: Images after adjusting brightness and size

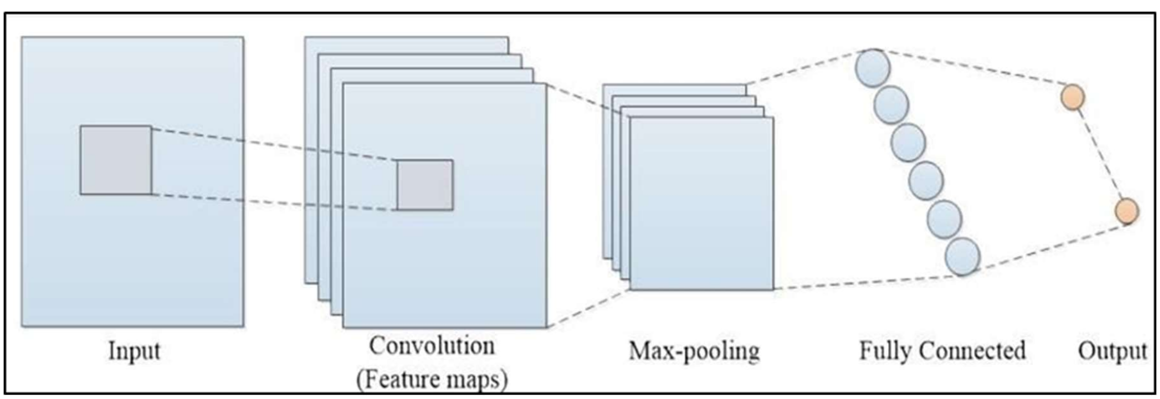


Figure 7: Simplest model of CNN [33]

3.4.3. Transfer Learning

TL is a strategy that is leveraged to transfer the knowledge learned by a model on a large dataset, for solving a similar problem, but with a smaller dataset. It reduces time and effort but also provides the same efficiency and speed as a state-of-art algorithm. Transfer learning is mostly used for problems of complex nature, i.e., image processing and natural

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

language processing, etc, by elevating the training methods. In this study, three different transfer learning models are employed. All the images are of size (299 x 299).

Transfer learning has become very popular recently in medical image diagnosing.

3.4.4. ResNet101

Within the ResNet (Residual Networks) family of CNN architectures, ResNet-101 was created to overcome training issues with deep neural networks (DNNs). ResNet-101 is a well-known deep and effective network for image categorization that was developed by Microsoft Research Asia. ResNet-101 is a significant variation of 101 layers, whereas other ResNet architectures ResNet-18 and ResNet- 32, for example—show different depths. ResNet dealt with the main problem of degradation in deep neural networks in an efficient manner. Network accuracy quickly reaches a saturation point and subsequently declines with increasing depth. The decline can be attributed to challenges in training optimization rather than overfitting. To resolve the issue of vanishing gradients, ResNet used Residual Blocks, which allow data to be transmitted directly over skip links. A Bottleneck Residual Block is a type of ResNet residual block. This block's architecture is seen in Figure 8.

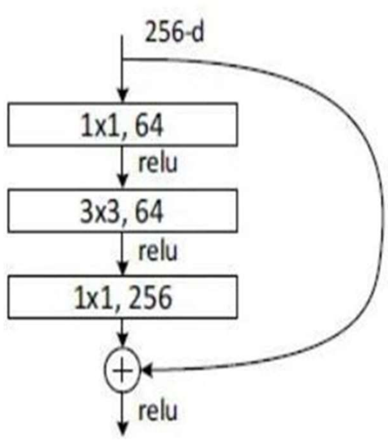


Figure 8: The bottleneck architecture of ResNet50 [35]

The ReLU activation function is applied following each convolutional and batch normalization layer. It permits only positive values to pass, introducing non-linearity to the network. This non-linearity is crucial for enabling the network to capture and learn complex patterns within the data.

3.4.5. Proposed Model (VGG19)

VGG19 was the deep CNN framework that represents a step forward from VGG16. There are nineteen layers in all, sixteen convolutional layers as well as three completely connected layers.Complex architecture of VGG19 permits it to derive features from visual input by using 3×3 convolutional filters, which makes complex patterns and features visible. By reducing the spatial size of the input data, maxpooling layers lower computer complexity. In order to enable forecasts derived from the abstract data gathered by the convolutional layers, the final layers are densely connected. VGG19 utilizes the ReLU activation function to handle nonlinearity. VGG19 is a commonly used standard in computer vision for picture classification. Despite its simplicity and depth, it is outperformed in terms of efficiency and performance by contemporary designs such as ResNet and Inception. After obtaining representations from the pre-trained VGG19 network, we have used customized fully connected layers for classification.The structure of the proposed model is shown in Figure 9.

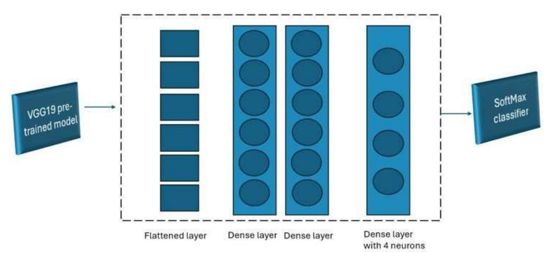


Figure 9: The architecture of the proposed model

Figure 10 illustrates the customized layers of the proposed model. The prediction from VGG19 the model that has been pre-trained is fed into the proposed model. The dimensions of the images are configured to 224x224x3. The image has a dimensions of 224 pixels in width as well as height, with three color channels. Flattened layer converts input data into a single-dimensional tensor The fully connected layers receive this tensor as input.The Dense () function consists of two densely connected layers, every layer including 4096 neurons. ReLU functions as a activation function. The output layer is the final layer, including four neurons. The output layer must contain the number of neurons equivalent to the count of classes that are in the dataset. The output layer utilizes the Softmax function as its activation function.All layers of VGG19 are frozen, preventing the pre-trained model from resuming training. The proposed model employs VGG19 as a feature extractor, extended with customized layers for classification. A summary of the proposed model

Journal of Theoretical and Applied Information Technology
30th June 2025. Vol.103. No.12

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

is shown in Figure 10.

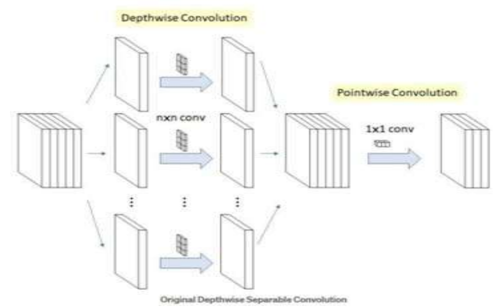
Layer (type)	Output Shape	Param #
input_layer (InputLayer)	(None, 224, 224, 3)	0
block1_conv1 (Conv2D)	(None, 224, 224, 64)	1,792
block1_conv2 (Conv2D)	(None, 224, 224, 64)	36,928
block1_pool (MaxPooling2D)	(None, 112, 112, 64)	0
block2_conv1 (Conv2D)	(None, 112, 112, 128)	73,856
block2_conv2 (Conv2D)	(None, 112, 112, 128)	147,584
block2_pool (MaxPooling2D)	(None, 56, 56, 128)	0
block3_conv1 (Conv2D)	(None, 56, 56, 256)	295,168
block3_conv2 (Conv2D)	(None, 56, 56, 256)	590,080
block3_conv3 (Conv2D)	(None, 56, 56, 256)	590,080
block3_conv4 (Conv2D)	(None, 56, 56, 256)	590,080
block3_pool (MaxPooling2D)	(None, 28, 28, 256)	0
block4_conv1 (Conv2D)	(None, 28, 28, 512)	1,180,160
block4_conv2 (Conv2D)	(None, 28, 28, 512)	2,359,808
block4_conv3 (Conv2D)	(None, 28, 28, 512)	2,359,808
block4_conv4 (Conv2D)	(None, 28, 28, 512)	2,359,808
block4_pool (MaxPooling2D)	(None, 14, 14, 512)	0
block5_conv1 (Conv2D)	(None, 14, 14, 512)	2,359,808
block5_conv2 (Conv2D)	(None, 14, 14, 512)	2,359,808
block5_conv3 (Conv2D)	(None, 14, 14, 512)	2,359,808
block5_conv4 (Conv2D)	(None, 14, 14, 512)	2,359,808
block5_pool (MaxPooling2D)	(None, 7, 7, 512)	0
flatten (Flatten)	(None, 25088)	0
fc1 (Dense)	(None, 4096)	102,764,544
fc2 (Dense)	(None, 4096)	16,781,312
predictions (Dense)	(None, 4)	16,388

Total params: 139,586,628 (532.48 MB)
Trainable params: 119,562,244 (456.09 MB)
Non-trainable params: 20,024,384 (76.39 MB)

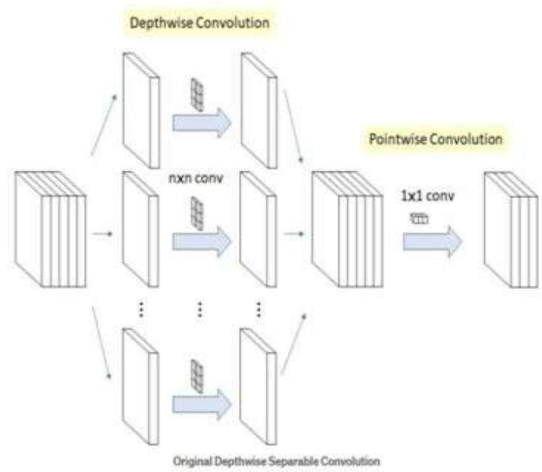
Figure 10: Summary of the proposed model

3.4.6. XceptionNet

XceptionNet is the advanced version of Inception Net. Figure 11 the difference between the two architectures. In the modified version, following pointwise convolution, depthwise convolution is used. In the original architecture of Inceptionv3, at first channel-wise spatial convolution is performed followed by 1 x 1 convolution [36]. Whereas in XceptionNet, 1 x 1 convolution is carried out first, then channel-wise spatial convolution is applied. In the Inception model except for the first operation, all the operations have nonlinearity. Whereas in XceptionNet process does not include intermediate non-linearity in depth-wise convolution layers. On "ImageNet" dataset XceptionNet has shown better performance than VGG16, Resnet152, and Inceptionv3 [37].



(a) InceptionNetv3



(b) XceptionNet

Figure 11: XceptionNet comparison with the original Inception Net. [38]

4. RESULTS AND DISCUSSION

In this section, the results of our experiments will be presented. The Discussion and Conclusion sections critically assess the model, highlighting VGG19's strong performance with low loss. However, overfitting in models like ResNet101 and XceptionNet is noted. The use of Kaggle MRI datasets may affect real-world applicability, emphasizing the model's strengths while acknowledging limitations in generalizability and robustness. This paper goes beyond incremental advances by introducing certainty-aware evaluation in clinical AI. Using models like CNN and VGG19, it uniquely correlates loss with prediction certainty, revealing cases where high accuracy masks low confidence.

4.1. Hardware and Software

Hardware: For performing experiments, we used 11th Gen Intel(R) Core(TM) i5-1135G7 having 2.42 GHz hard disk and 16 GB RAM. It is a 64-bit operating system, x64-based processor, with 11th-generation windows.

Software: We used Anaconda platform and Python 3.0 for performing the experiments.

Performance Metrics: the performance metrics that we are using in this analysis include accuracy, precision, recall and loss functions. Accuracy means ration between correct predictions to the total predictions made. Precision means ratio between positive prediction to all the predictions. Recall is also known as sensitivity, it was the ratio between true positive to all actual positive instances. High recall is more important than high precision. High recall means never missing a case of tumor that should be predicted. Loss is the error, that the model makes while prediction. Loss means the average error made by the model. The following Figure 12 shows how these metrics can be calculated. One DL model is used in this study, to compare its performance with transfer learning models.

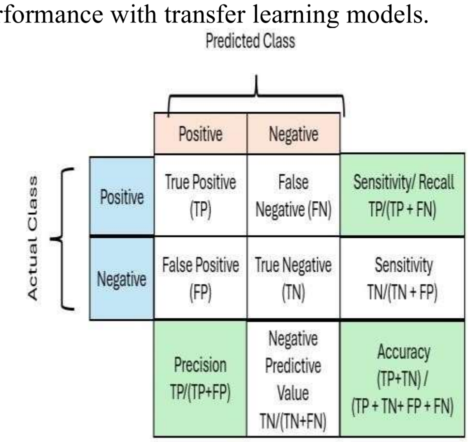


Figure 12: Confusion matrix with precision, recall, and accuracy [39]

4.2. Experiment_1 Convolutional Neural Network:

In the CNN scenario, we have used five blocks of convolutions 2D. The structure of the CNN model was shown in Figure 13. It can be seen that 5 convolutional blocks are used. In the first 3 blocks, there are 2 convolutional layers and one max pooling. In the 4th as well as 5th blocks, there are 3 convolutional layers and one max pooling layer. In the next step, layers are flattened before using dense layers, that are densely connected layer. The time taken to train one epoch is 5 mins, 2 sec.

To present a clear picture we found not only focused on validation accuracy but also on validation precision, validation recall, and validation loss. Validation accuracy is the one that is achieved by the model by using a validation dataset. We used the test dataset for prediction purposes. And checked which model correctly predicts the images, belonging to a certain class. A validation loss of 0.38 is observed for the CNN model. We found a correlation between the loss and prediction certainty of the model. If the model has a greater loss, then it will not be certain in making predictions on the unknown dataset. Less loss shows better efficiency of the model. The lower value of the loss function shows how well the model can predict and how much the model can miss the actual values. So lesser loss value is the best one. In this study, we are focusing on the loss value of benchmark DL models and various transfer learning models to see which model can predict better. The hyperparameters used in the CNN model are represented in Table 1 . From Figure 14, it can be seen that the training accuracy of the CNN model is 96% as well as the validation accuracy is 91%. Values of other performance metrics are given in Table 2. A precision of 0.95 during training means 95% of the images are correctly learned by the model, but 5% images are not correctly learned, i.e. model predicted them as tumors but they were not tumor images. A recall of 0.95 means that 95% of the images are correctly learned, whereas 5% of the images are predicted as healthy but they were tumor images. The recall is more important than precision because a missing tumor is more dangerous than a false prediction of the tumor during validation, a loss of 0.37 was observed that means the model is well trained but it needs improvements. It shows the average error made by the model during validation.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

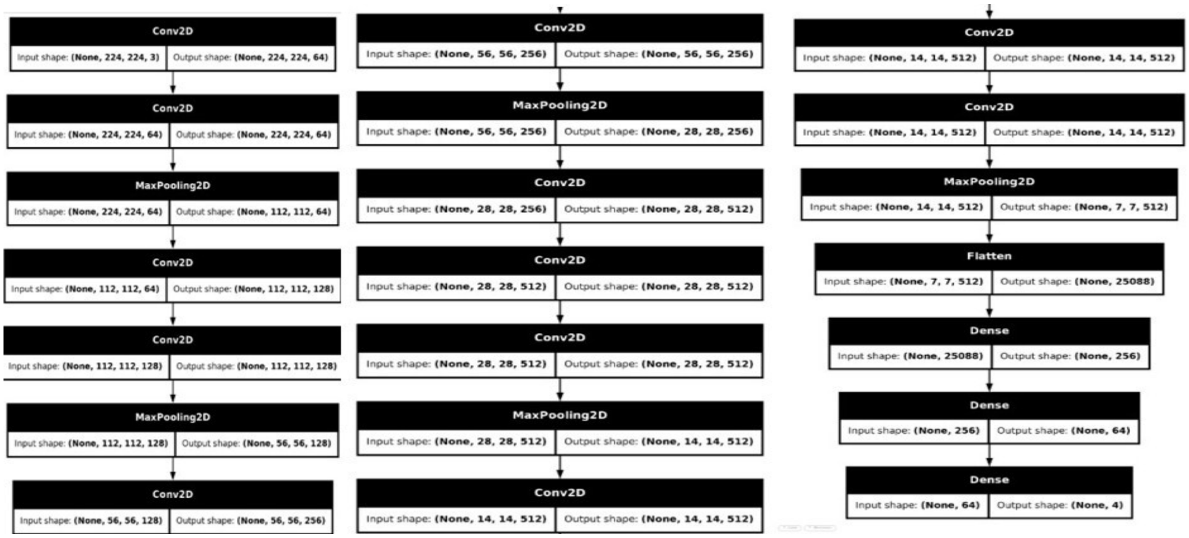


Figure 13: CNN model's architecture used in this study

Table 1: Hyperparameters used in the CNN model

Hyperparameters	Values of the hyperparameters
Size of batch	16
Optimizer	Adam
Number of epochs	10
Evaluation criteria	Cross entropy loss
Learning rate	0.001

Table 2: Performance metric values for training, testing, and validation datasets for CNN

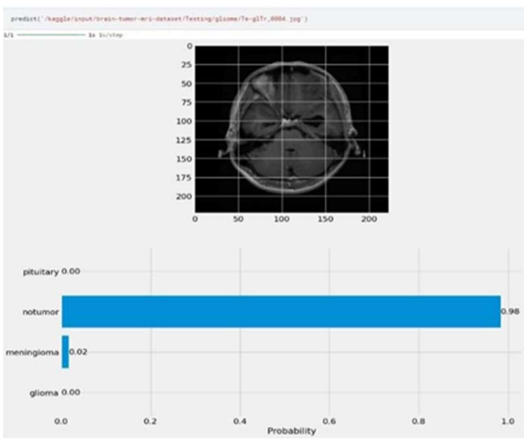
Dataset	Accuracy	Loss	Precision	Recall
Training	0.95	0.12	0.95	0.95
Validation	0.90	0.37	0.90	0.90
Testing	0.91	0.30	0.91	0.91

Ground truth: Glioma Model's prediction: No tumor Misclassified

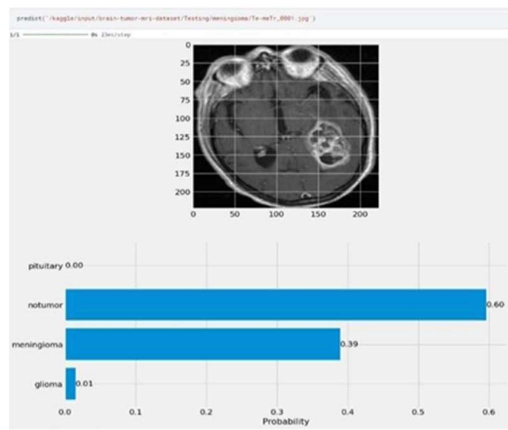
Ground truth: Meningioma Model's prediction: No tumor (60%) probability), Meningioma (30% probability) Very low confidence

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

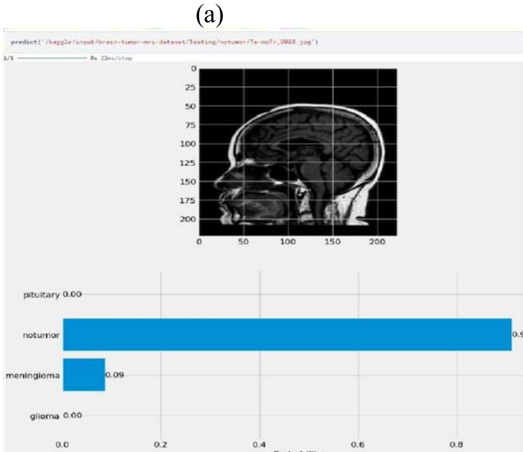
Figure 14: Training and validation accuracies, precision, recall, and loss values plots for CNN



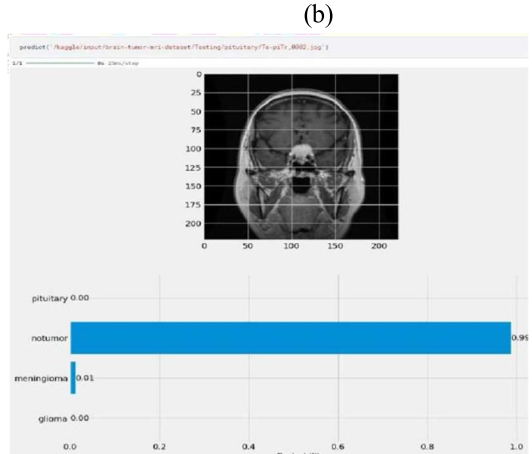
(a)



(b)



(c)



(d)

Ground truth: No tumor Model's prediction: No tumor

Correct prediction

Ground truth: Pituitary Model's prediction: No tumor

Misclassified

Figure 15: CNN model's predictions for the classes a) Glioma, b) Meningioma, c) No Tumor d) Pituitary

Journal of Theoretical and Applied Information Technology
30th June 2025. Vol.103. No.12

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

From Figures 15, we can see that the CNN model is not confident in predicting the images. So accuracy alone is not enough in predicting medical images. There must be certainty in the response. From the test results of the CNN model, we found that two out of four are wrongly predicted. One is predicted with very low confidence, i.e., 30%, and only one out of four images is predicted correctly.

4.3. Transfer Learning

For the rest of the experiments, transfer learning is used. for this purpose, we used three popular pretrained models, Resnet101, VGG19 (for the Proposed Model, and XceptionNet. The hyperparameters used for the transfer learning experiments are described in Table 3.

Table 3: Hyperparameters used in transfer learning models

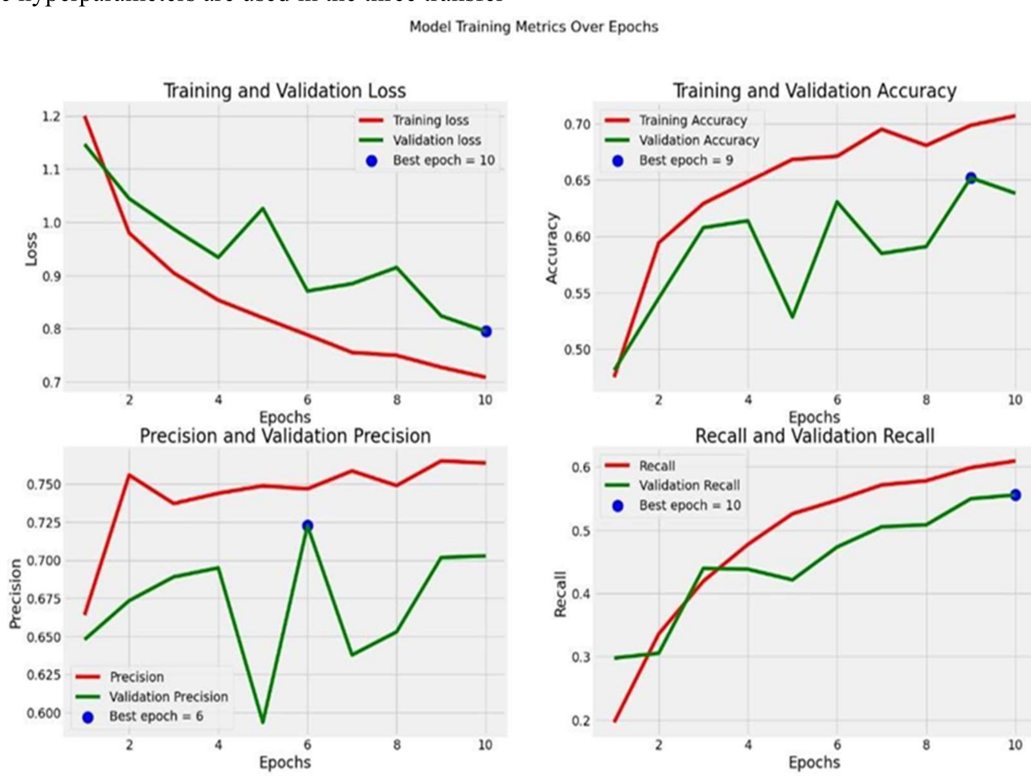
Hyperparameters	Values of the hyperparameters
Size of batch	32
Optimizer	Adamx
Number of epochs	10
Evaluation criteria	Cross entropy loss
Learning rate	0.001

These hyperparameters are used in the three transfer learning models that are used in this study.

4.4. Experiment_2 Resnet101

A Residual Neural Network lets hundreds of layers train easily. It is considered one of the best pre-trained models due to its robustness in computer vision especially for image classification. This is the reason, that we included this model in our research study. Resnet provided a novel way for dealing with the vanishing gradient problem, i.e., Overlooking one or several layers which is called an identity shortcut.

For using CNN in this study, the data preprocessing and augmentation is done as initial experiment. A CNN model IS build based on the pretrained ResBet101 model. For this Tensorflow's kera library is used. Since the transfer learning approach is used, the knowledge gained by ResNet101 after training on the "imagenet" dataset is utilized to build a new model to the present learning issue, i.e., brain tumor detection. The input shape that is defined for this model is 299 x 299 x 3. Resnet101 model was loaded without the top classification layer, so we can add our own layers according to the issue at hand. Weights of the models are initialized, which the model learned after training on "imagenet" dataset. Resnet101 is called and weights are initialized, we say this as the base model. This base model has information about the learned features.



Journal of Theoretical and Applied Information Technology
30th June 2025. Vol.103. No.12
© Little Lion Scientific
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Figure 16: training and validation accuracies, precision, recall, and loss values plots forResnet101

The information from the base model are transmitted through the global mean pooling layer. It takes global mean of all the input data, to avoid variability in the input data. A dense layer with 512 neurons is used, together with ReLU as activation function. In this layer all the features are lined up in a vector, having 512 neurons. This layer helps in learning from the features, learned by Resnet101. To prevent overfitting, the dropout layer is appliedTo prevent overfitting, the dropout layer is applied. The last dense layer has neurons equal in number to the classes, also known as the output layer. In this layer, Softmax is utilized as the activation function instead of ReLU . Softmax ensures that sum of probabilities should be equal to 1. The final model in order to learning brain tumor detection from MRI scans is made, by combining base model (Resnet101) and custom output layers. All the layers of the ResNet model are frozen so their weights may not update during the model's training. This is done to reduce costs in terms of time and complexity. we can soundly say that in transfer learning Resnet101 is used as a feature encoder. Custom classification layers are added to perform classification tasks. The performance of the model is displayed in Figure 16. From Figure 16, it can be seen that the accuracy during training of the Resnet101 model is 72.81% as well as the validation accuracy is 64.89%. Values of other performance metrics for training, validation, and testing datasets are given in Table 4.

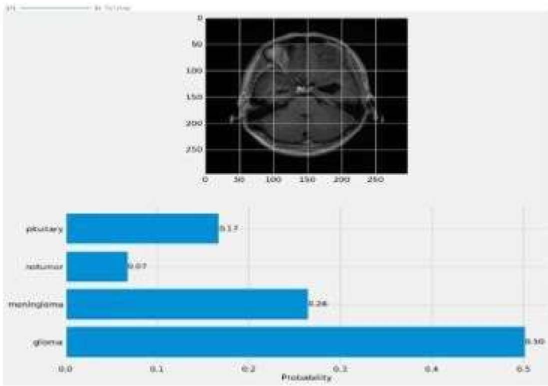
Table 4: Performance metric values for training, testing, and validation datasets for Resnet101.

Dataset	Accuracy	Loss	Precision	Recall
Training	72.78	0.66	0.78	0.64
Validation	65.39	0.79	0.70	0.55
Testing	68.52	0.74	0.75	0.57

The model's accuracy for correct predictions during training is 72.78%. The validation and testing accuracies were lower, i.e., 65% and 68%. The loss of the model increased from training to validation and then on testing. An increase in the loss value shows that the model is not learning correctly. Out of all the predictions the model made during training, only 78% were correct, and this value drops for the validation dataset, i.e. 70%. For the testing dataset out of all the predictions the model made only 75% were correct. Out of all the positive cases, the model was able to correctly identify positive cases up to 64% during training, 55% during validation, and 57% during testing. The model is performing relatively better on training than validation and testing. This leads to the fact that the model shows signs of overfitting to the training data. The same collection of images is used for Resnet101, to check its prediction, and it is found that 3 out of 4 are misclassified as shown in Figure 17. Only one instance that is correctly classified but it is very low probability, i.e., the model was not certain about the correct prediction. It shows that this model is not good for the identifying brain tumors through MRI scans.

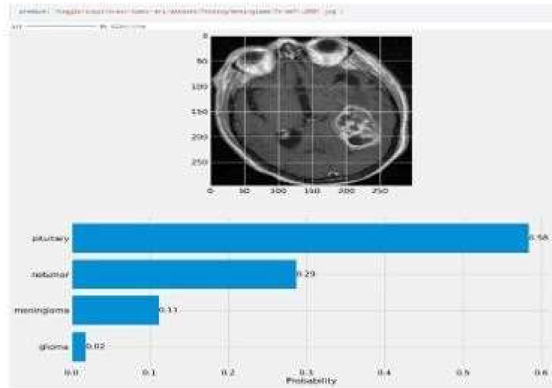
Journal of Theoretical and Applied Information Technology
30th June 2025. Vol.103. No.12

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195



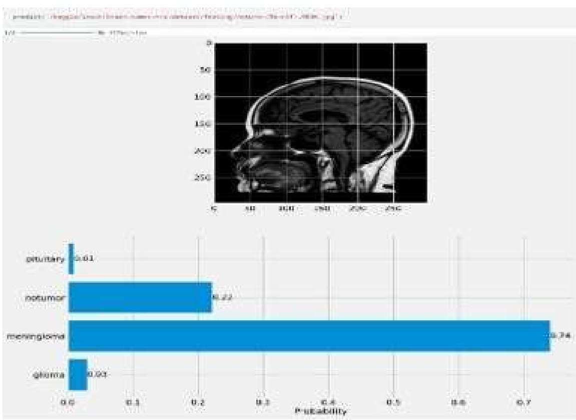
(a)

Ground truth: Glioma
Model's prediction: Glioma (50% probability), Meningioma (26% confidence), No tumor (7% probability), Pituitary (17 % probability)
Misclassified



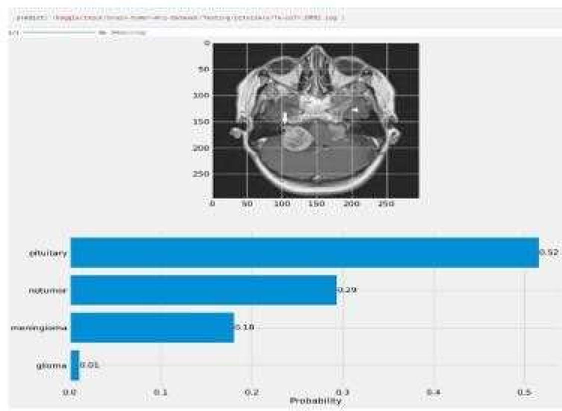
(b)

Ground truth: Meningioma
Model's prediction: Pituitary (58% probability), No Tumor (29% probability), Meningioma (11% probability), Glioma (2% probability)
Misclassified



(c)

Ground truth: No tumor
Model's prediction: No Tumor (22% probability), Meningioma (74% probability), Glioma (3% probability)
Misclassified



(d)

Ground truth: Pituitary
Model's prediction: Pituitary (52% probability), No Tumor (29% probability), Meningioma (18% probability), Glioma (1 % probability)
Correctly predicted with very low probability, 52%

Figure 17: Resnet101 model's predictions for the classes a) Glioma, b) Meningioma, c) No Tumor d) Pituitary

4.5. Experiment_3 Proposed Model (VGG19)

VGG19 is an expanded iteration of VGG16, with 19 layers as indicated by its name. There are sixteen convolutional layers, three completely linked layers, and five pooling layers. It used 3 x 3 filters in all convolutional layers. The images undergo convolution with various filters, resulting the generation of feature maps. The first 2 layers include 64 filters, deriving in the generation of 64 feature maps post-convolution. Similarly, the subsequent two convolutional layers include 128 filters, and the next four layers comprise 256 filters. The last four convolutional layers include 512 filters each. The feature maps are pooled to extract the most salient features, ensuring that critical information is preserved. The architecture is quite simplistic. It takes the form of a sequence of convolutional layers, each followed by ReLU activations and max pooling layers. At end, there are three entirely linked layers named fc6, fc7, and fc8.Terminal layer is output layer, with softmax as its activation function. VGG19 has 144 million parameters; yet, it incurs significant training time costs.

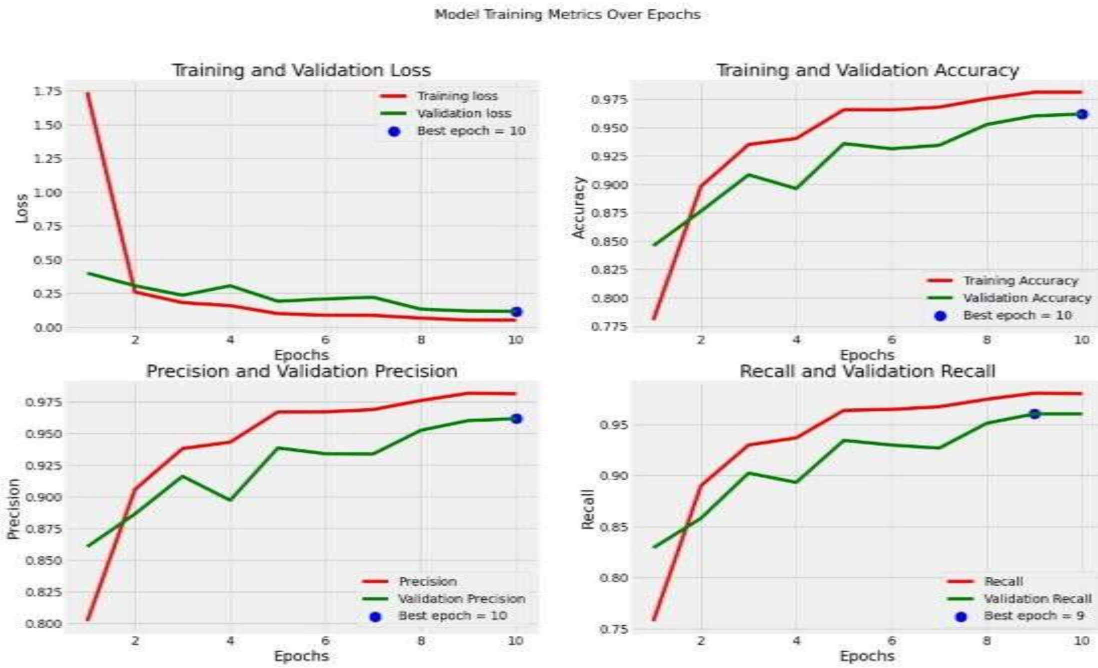


Figure 18: training and validation accuracies, precision, recall, and loss values plots. for Proposed Model

The proposed model's customized layers are shown in Figure 10. It takes as input the results produced by the VGG19 pre-trained model. The images are set to 224 × 224 x 3 dimensions. With three colour channels (RGB), the picture has a height and width of 224 pixels. A one-dimensional tensor is created from the input by use of the flattened layer. The fully connected layers use this tensor as input. There are a total of 4096 neurons in each fully connected. ReLU was used as an activation function in the entirely linked layers. The output layer, comprising four neurons, is the last layer. The number of neurons in the output layer must match the total number of classes in the dataset. The output layer uses the softmax function as its activation mechanism. The proposed model uses VGG19 to extract features, with additional custom layers added for classification. Figure 18 shows the performance of the proposed model.

The aforementioned results demonstrate model's exceptional functionality, achieving 99% training accuracy as well as 96% accuracy of the validation data. The loss is much lower than in prior models, with training loss of 0.01 as well as validation loss of 0.11. The loss value indicates that the model has learned well, and per our hypothesis, it will also make predictions with optimal confidence and certainty. Additional performance metric data are included in the Table 5. A 99.60% accuracy indicates that the model has acquired a profound understanding of the training data. A loss value of 0.017 indicates that model exhibits high confidence in its predictions on training data. Precision as well as recall indicate that the model has accurately predicted all classes. The model's accuracy on the validation dataset decreased somewhat, perhaps indicating overfitting. A 97% accuracy on the unseen data (test data) is the true benchmark. The loss value on the test data is 0.08, indicating the model's confidence in its predictions. This confidence is also shown in Figure 19. It depicts that the model has accurately predicted 3 out of 4 cases. It is noteworthy that the model has predicted the actual classes with 100% and 99% confidence, while also misclassifying with 98% confidence. The findings indicate a link between lower loss value and increased accuracy in the model's predictions.

Table 5: Performance metric values for training, testing, and validation datasets for Proposed Model

Dataset	Accuracy	Loss	Precision	Recall
Training	99.60	**0.017**	99.60	99.58
Validation	95.73	0.1118	95.73	95.67
Testing	97.05	0.08	97.04	96.56

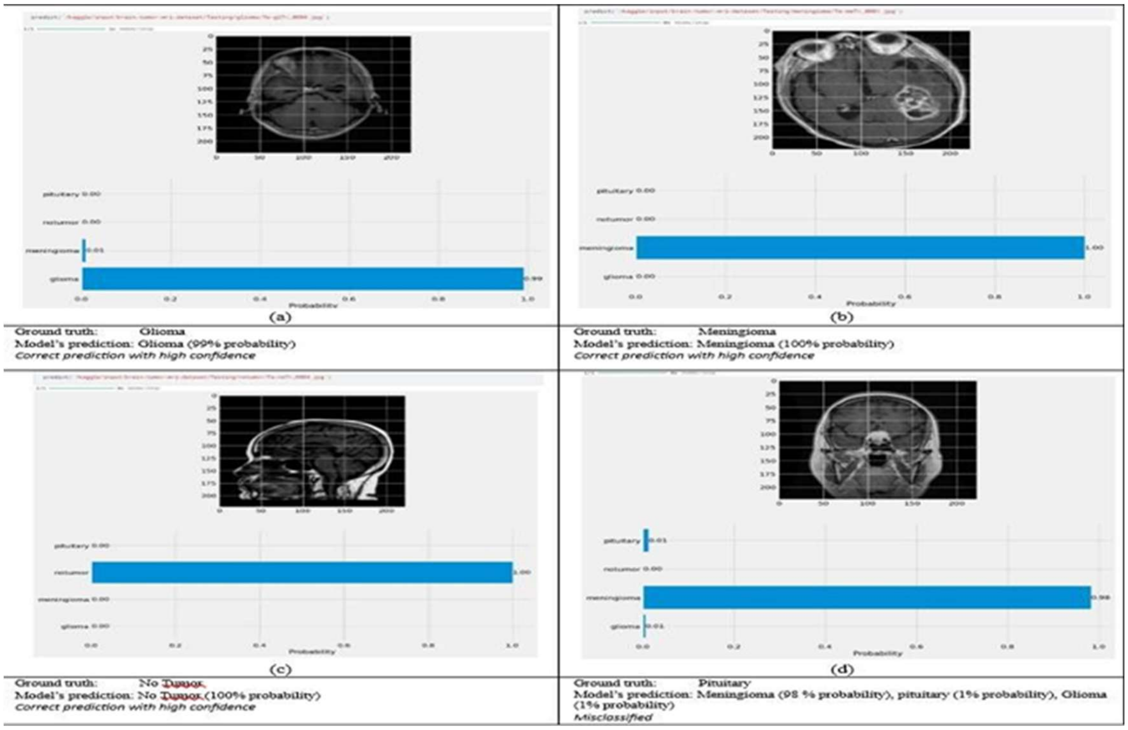


Figure 19: Proposed model's predictions for the classes a) Glioma, b) Meningioma, c) No Tumor d) Pituitary

4.6. Experiment_4 XceptionNet

XceptionNet is the extreme version of InceptionNetv3. It can capture diverse features from the images. It contains 71 layers and is pre-trained on the "ImageNet" dataset that has 1000 classes and millions of images.

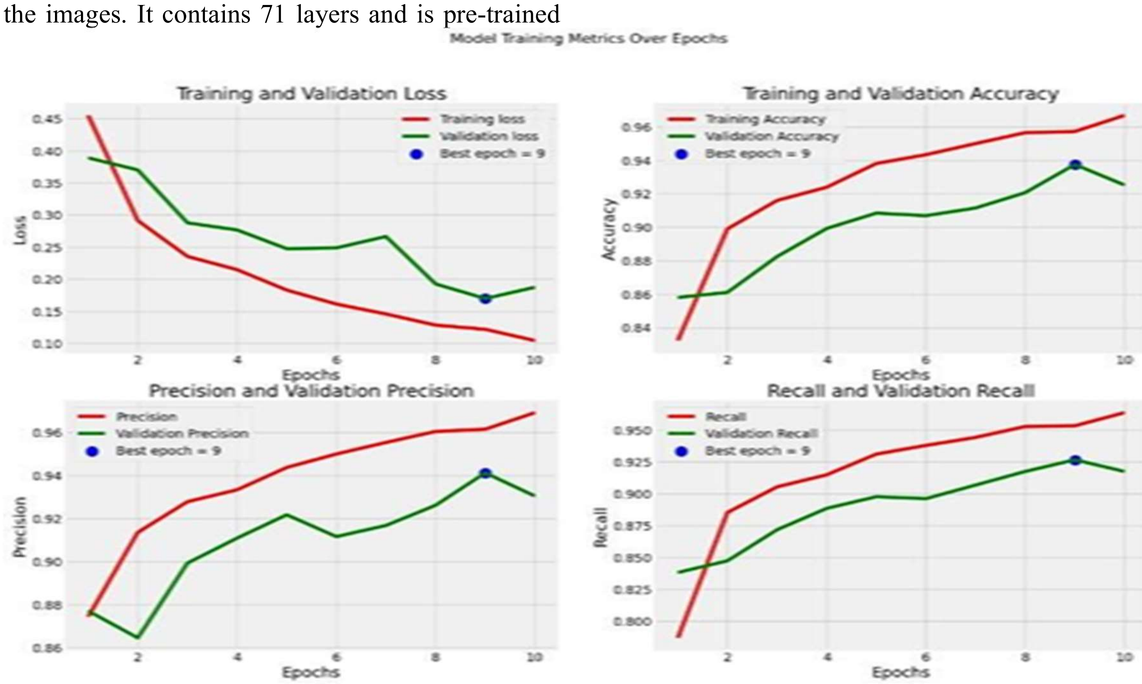


Figure 20: training and validation accuracies, precision, recall, and loss values plots. For XceptionNet

It performs depth-wise filtering on each map and then reduces their dimensions using a 1x1 convolution. The XceptionNet model is loaded without its top classification layers and is pre-trained on the "ImageNet" dataset, which contains millions of images. This model learns the weights during the training process. So when model is loaded, the weights of the current model are started with pre-trained weights obtained from the "ImageNet" dataset. This model is loaded as the base model. Global average pooling layers are added to reduce spatial dimensions. Two dense layers are used in the custom model, a layer with 512 neurons utilizing the ReLU activation function is followed by a dropout layer, and another layer serves as the output with 4 neurons.One neuron is used for one class. The base layers of XceptionNet are set to freeze so the model doesn't train those layers again while training the custom model. Figure 20 illustrates the performance of the XceptionNet model.

As shown in Figure 20, It is evident that the training accuracy of XceptionNet is quite impressive. XceptionNet attained an accuracy of 96.64% alongside a training loss of 0.10.The validation accuracy of model is 92.21% and the validation loss is 0.19.Substantial difference among training and validation values is present. That indicates that model is overfitting training data. The values of other performance metrics are given in the table 6.

Table 6: Performance metric values for training, testing, and validation datasets for XceptionNet

Dataset	Accuracy	Loss	Precision	Recall
Training	96.57	**0.098**	96.87	96.21
Validation	92.39	0.18	93.05	91.88
Testing	93.83	0.19	94.33	93.27

An accuracy of 96.57% indicates that model is doing well on training data. A loss of 0.098 indicates that model has appriciatabily assimilated the training data, exhibiting few errors. Both accuracy and recall exhibit favorable values, indicating that the model accurately identifies positive and negative situations with equilibrium. The accuracy of the validation dataset decreases somewhat, indicating a bit overfitting; nonetheless, the loss value of 0.18 demonstrates that the model generalizes with few errors. The testing accuracy of 93.83% indicates that the model performs well on testing data as well. The loss for the testing dataset is 0.19, which is almost identical to that of the validation dataset that demonstrates the model's stability. The accuracy and recall exhibit commendable levels, namely 94.33% and 93.27%, respectively. The model shows accurate predictions for both positive and negative situations. There exists a negligible trade-off among false positives and false negatives.

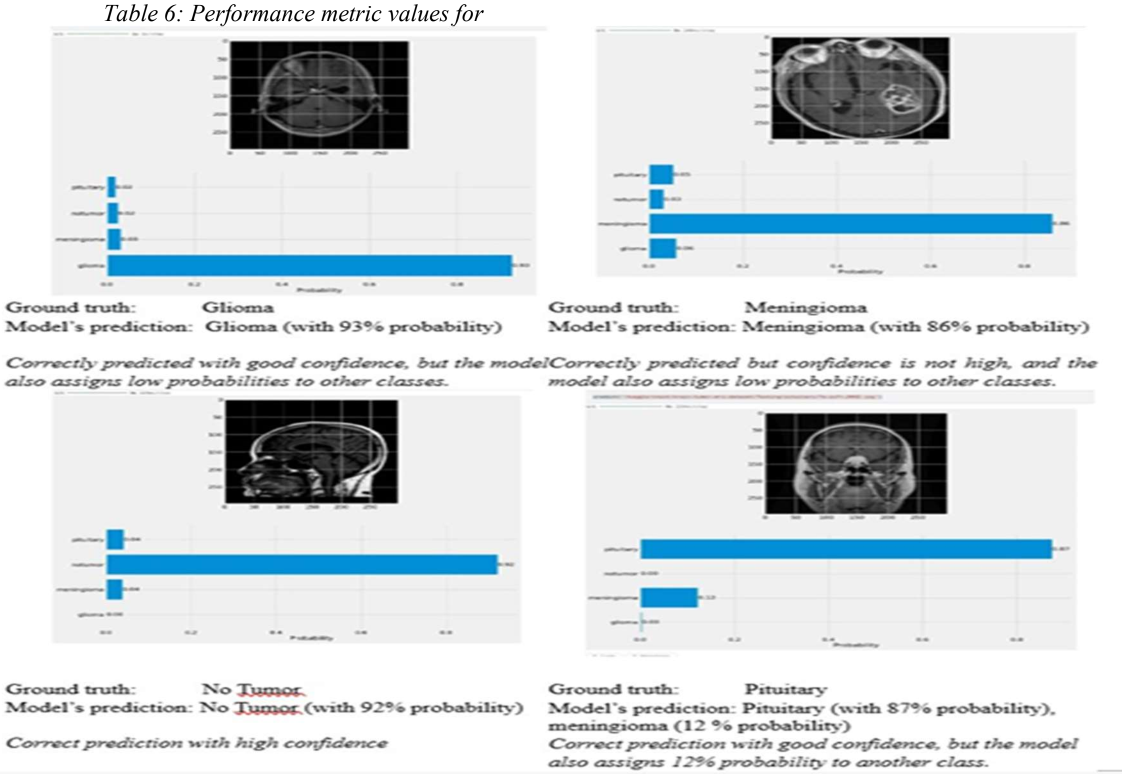


Figure 21: XceptionNet model's predictions for the classes a) Glioma, b) Meningioma, c) No Tumor d) Pituitary

Journal of Theoretical and Applied Information Technology
30th June 2025. Vol.103. No.12

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Figure 21 demonstrates that model accurately predicted entire four classes, in contrast to other models, which gave at least one erroneous prediction. The minimal loss value during training tells that model has accurately learned four classes, committing just a few errors in the process. The model's loss value of 0.098 is rather high compared to the Proposed Model's training loss value of 0.017. The model demonstrates proficient learning with few errors; nonetheless, its confidence is inferior to that of Proposed Model. The facts presented in the predictions corroborate this statement.

The model has accurately predicted the instances, but not with very high confidence; in this context, high confidence refers to a 100% probability. The impact of loss value on the model's predictions is evident. Thus, only depending on the model's accuracy is insufficient. We must consider the loss value while implementing the model in a real-time context. In future work, an ensemble of VGG19 and XceptionNet may be developed to enable the model to accurately predict all classes with high confidence. The training time of one epoch of the models used in this study is given in the Table 7.

Table 7: Training time for different models

Model	Time (Hr:Min)
CNN	02:25
XceptionNet	12:05
Resnet50	03:15
Proposed Model	03:30

4.7. Discussion

This subsection presents the results of the experiments conducted. All models demonstrate accuracies exceeding 90%, except Resnet101. The outcomes of the top four models are presented in the subsequent Figure 22. The line in Figure 22 illustrates the loss of the models. The recall value of Resnet101 shows a significant decline from the training phase to the validation phase which shows Resnet is not effectively generalizing to unseen data. Figure 22 depicts that the value of the loss is inversely proportional to the other performance metrics. The training dataset for Proposed Model exhibits the lowest loss value, while the validation dataset for Resnet101 shows the highest loss point. It is important to observe that when the loss value is elevated, the model's recall value tends to be lower compared to other performance metrics. The recall of XceptionNet is also low in comparison to the precision. it shows the model is unable to achieve a balance between false positives and false negatives. The sole example in which both precision and recall exhibit good values is the Proposed Model. It also demonstrates the lowest loss value, thereby achieving the best prediction results. The loss value indicates the number of training instances that are misclassified by the model throughout the training process. A lower loss value indicates the improved efficiency of the model. Figure 22 illustrates the performance metric values for training, validation, and testing across the four models utilized in the study.

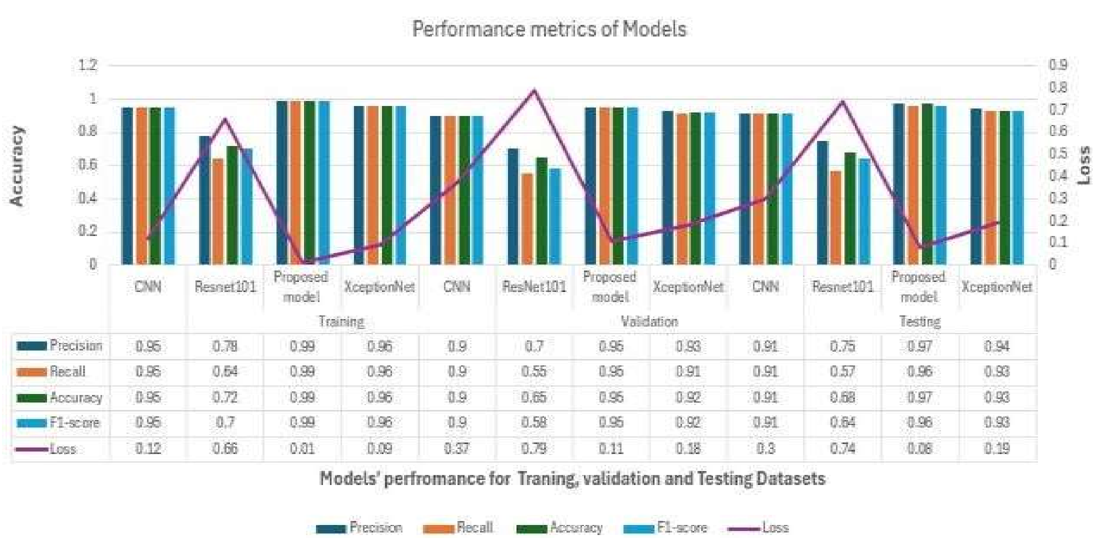


Figure 22: Performance metrics Precision, Recall, F1-score, Accuracy, and Loss of four models used in the study

The performance of different models for the four classes of the dataset is given in Table 8. The following values are added by using the classification report of the models. The report and

Journal of Theoretical and Applied Information Technology
30th June 2025. Vol.103. No.12

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

confusion matrix both are generated based on the test dataset. The count of instances of each class in the test dataset is shown in Figure 23.

Table 8: Performance of the models CNN, Resnet101, Proposed Model, and XceptionNet for four classes of dataset

Model	Class	Precision	Recall	F1score	Accuracy
CNN	Pituitary	0.97	0.98	0.97	0.92
	Nontumor	0.93	1	0.96	
	Meningioma	0.83	0.88	0.85	
	Glioma	0.98	0.78	0.87	
Resnet101	Pituitary	0.66	0.87	0.75	0.69
	Nontumor	0.85	0,77	0.81	
	Meningioma	0.50	0.50	0.50	
	Glioma	0.73	0.57	0.64	
Proposed Model	Pituitary	0.98	0.99	0.99	0.97
	Nontumor	1	1	1	
	Meningioma	0.95	0.92	0.93	
	Glioma	0.94	0.95	0.95	
XceptionNet	Pituitary	0.91	1	0.96	94.36
	Nontumor	0.98	1	0.99	
	Meningioma	0.96	0.81	0.88	
	Glioma	0.92	0.95	0.93	

Table 8 shows that CNN best learned the class “no Tumor” and it is also shown from the prediction Figure 14 of the CNN model. The high precision value of Glioma, 0.98 shows that the model can correctly predict glioma cases, but the low recall 0.78 shows that the model misses a lot of glioma cases. The loss of 0.3 shows that the model is not certain in the predictions. Like “No Tumor” model performed well for the class “Pituitary”.

In the case of Resnet101, the model has the worst performance values for Meningioma, i.e., 50% for precision, recall, and F1 score. A 50% precision means only half of the positive predictions made by the model are correct. It shows that the model cannot distinguish between positive and negative instances.

Recall that 50% means the model is identifying half of the actual positive instances.

Table 8 shows that Proposed Model gained the best accuracy for all the classes among the models. It has learned the classes “Pituitary” and “NoTumor” with maximum metric values i.e., 99% and 100% respectively. Its outstanding learning and performance can also be seen in the model’s prediction images. It has a minimum loss value (0.01 for training and 0.08 for testing) that shows the robustness and certainty of the model.

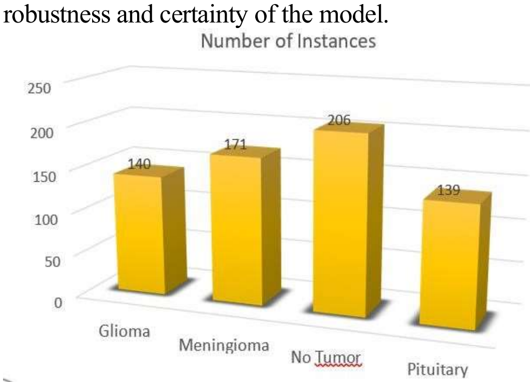


Figure 23: Number of instances of each class in the test dataset

4.7.1. Confusion matrices

The confusion matrices of the four models utilized in the study illustrate the classification instances, specifically the number of instances that have been accurately classified. This study addresses a multi-class classification problem, where the diagonal elements, represented in shades of blue, indicate the instances that have been correctly classified. The hue of blue intensifies with the rising number of instances. A maximum dark box along the diagonal indicates that all instances of that class have been accurately classified by the model. A lighter shade of

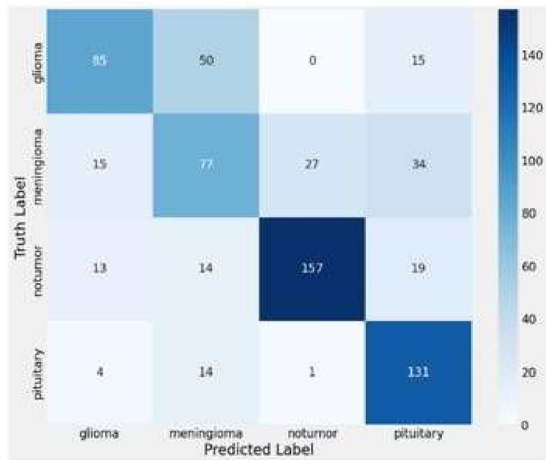
Journal of Theoretical and Applied Information Technology
30th June 2025. Vol.103. No.12

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

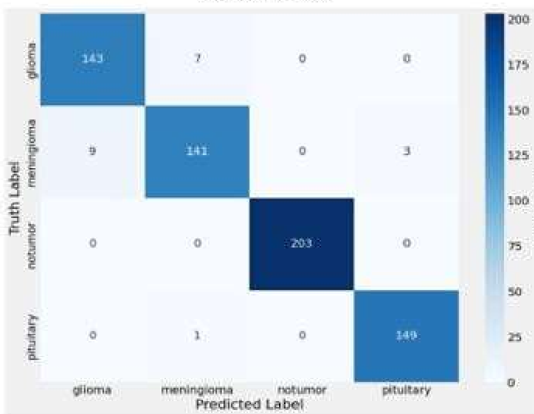
blue in the diagonal indicates that certain instances are not captured by the model. A very light shade of blue in the diagonal show model failed to classify, resulting in a majority of instances being misclassified.

Figure 24 illustrates that the Proposed Model exhibits the most pronounced dark shades of blue along the diagonal, indicating minimal missed instances. The off-diagonal elements indicate instances of misclassification. The confusion matrix illustrates the model's performance in terms of the classification of the instances.

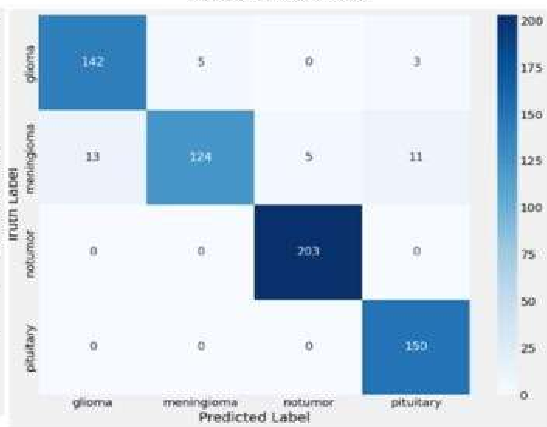
Model: CNN



Model: Resnet101



Model: Proposed Model



Model: XceptionNet

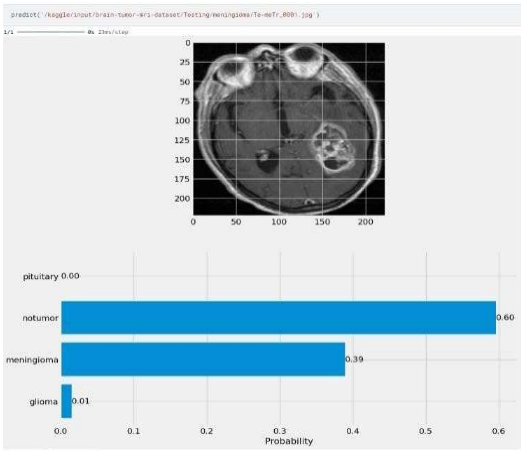
Figure 24: Confusion matrices of all the models used in this research study

4.7.2. Predictions

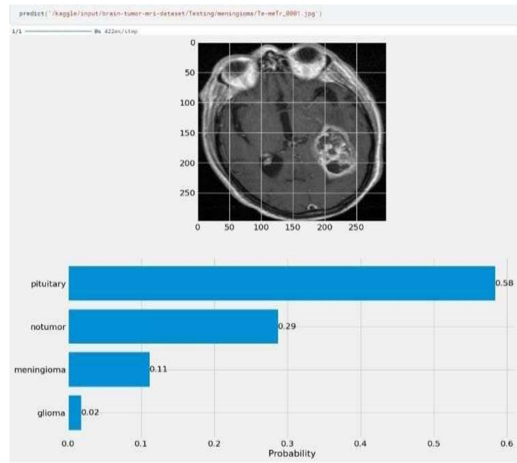
Figure 25 shows the impact of loss value on the model's predictions. To keep an analogy the same image is used for the four models. Loss value has a great impact on the prediction accuracy of the classifiers. If the two models have the same accuracy, but one model (say model A) has a greater loss value than the other (say model B). The model with greater loss (model A) will merely make predictions, with less certainty, whereas the model that has less loss value (model B) will be more certain in making predictions.

Journal of Theoretical and Applied Information Technology
30th June 2025. Vol.103. No.12

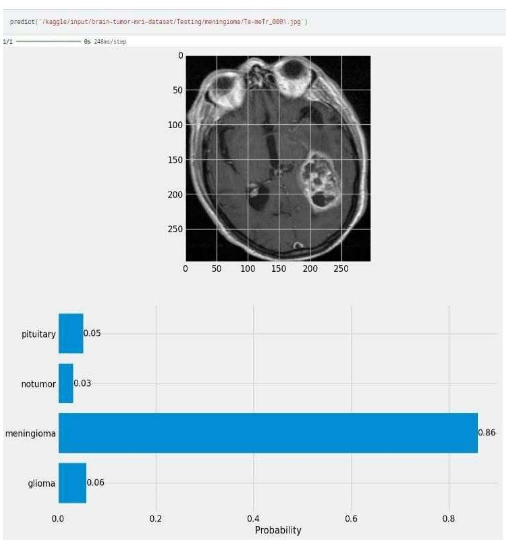
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195



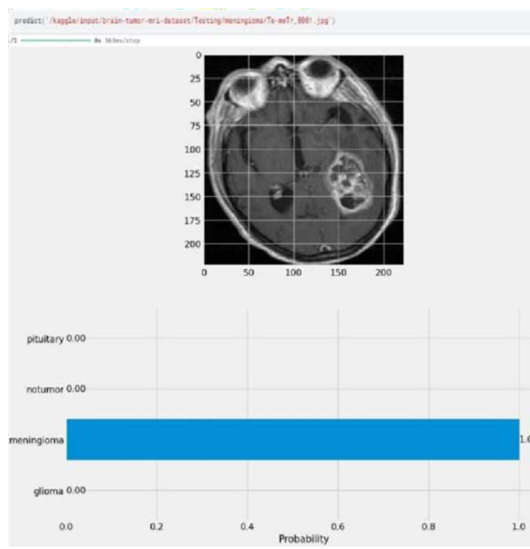
(a)
CNN, accuracy = 91.7 , loss = 0.35



(b)
Resnet101, accuracy = 68.60, loss =0.74



(c)
Proposed Model, accuracy = 96.95%, loss = 0.08



(d)
XceptionNet, accuracy = 94.36%, loss = 0.17

Figure 25: Predictions of four models with their accuracies and loss, a) CNN, b) Resnet101, c) Proposed Model, d) XceptionNet

Figure 22 indicates that the Proposed model achieves an accuracy of 96.95% and a minimal loss value of 0.08. This model is predicted with absolute certainty. XceptionNet has commendable accuracy at 94.36%; nevertheless, its loss value is much more than that of the proposed model, measuring at 0.17. The predictions generated by XceptionNet are influenced by the loss value, resulting in the model lacking high confidence and distributing some percentage of probabilities over other classes. CNN's accuracy is almost comparable to that of XceptionNet, at 91.7%, although it exhibits a much higher loss value of 0.35. The loss value adversely impacts the classifier's performance, as seen in part (a) of Figure 24. The CNN model predicts a 39% probability for the correct class and a 60% probability for a different class. The model exhibits just 39% confidence that the image is classified correctly. The model demonstrates inadequate learning despite a high classifier accuracy. Figure 14 of the CNN model indicates that the model has confidently predicted only one image out of four. Figure 24 demonstrates that a reduced loss value is associated with an increased confidence in the model. In constructing an autonomous prediction model, reliance just on the model's accuracy is

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

inadequate; the loss value must also be considered, and in fact, the loss value should be prioritized to achieve certainty.

In order to find the relation between loss and certainty score, we have tabulated the results in Table 9. The certainty score is the probability of the model for predicting the correct class. This probability can be seen in Figures 14, 16, 18, and 20. We have tabulated the results of certainty scores and loss values of four models, including the proposed model. Our proposed model achieved the best certainty score and lowest loss value.

Table 9: Certainty score and loss values of the models

Image class	Model	Certainty score	Loss
Glioma	CNN	0	0.3
	ResNet101	0.5	0.7
	Proposed Model	**0.99**	**0.08**
	XceptionNet	0.93	0.19
Meningioma	CNN	0.39	0.3
	ResNet101	0.11	0.7
	Proposed Model	**1**	**0.08**
	XceptionNet	0.86	0.19
No tumor	CNN	0.91	0.3
	ResNet101	0.22	0.7
	Proposed Model	**1**	**0.08**
	XceptionNet	0.92	0.19
Pituitary	CNN	0	0.3
	ResNet101	0.52	0.7
	Proposed Model	**0.98**	**0.08**
	XceptionNet	0.87	0.19

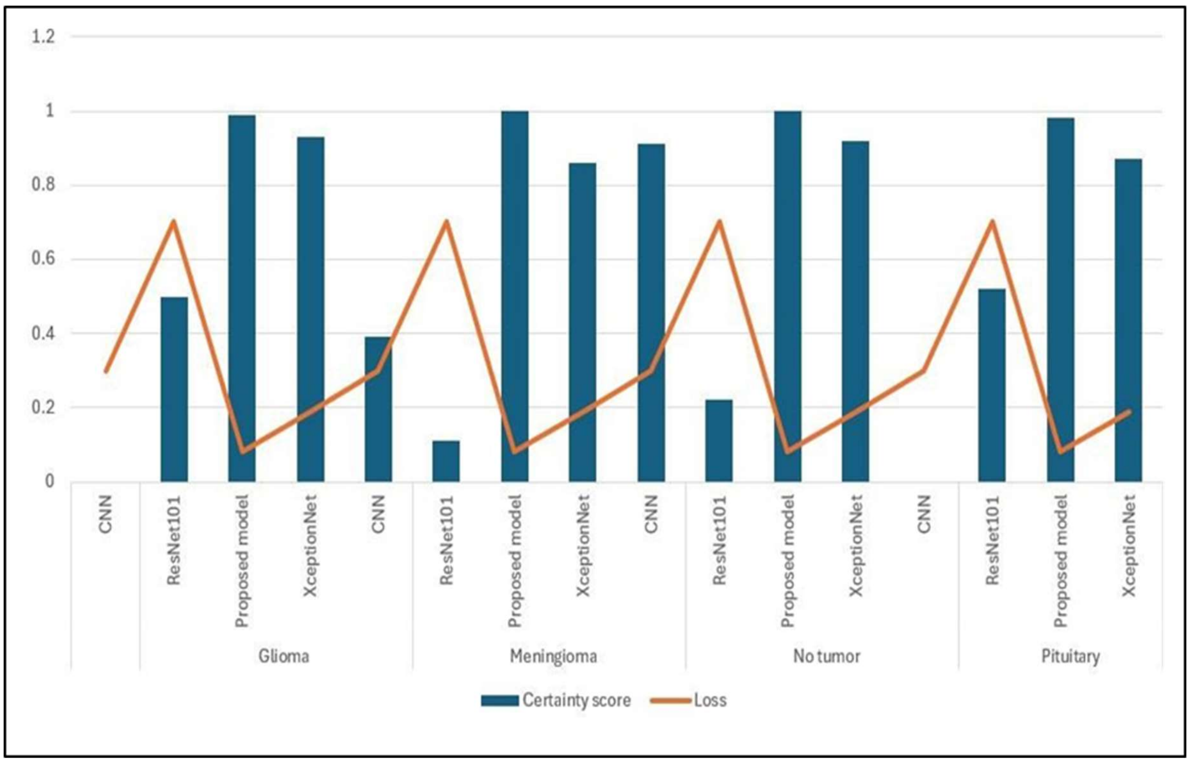


Figure 25: Certainty score and loss values of the models

The data presented in Table 9 is shown for enhanced comprehension in Figure 25. Analysis of the figure and table indicates that a rise in loss value corresponds to a decrease in the model's certainty score, and vice versa. To get a satisfactory certainty score, the model's loss value must be considered; otherwise, the model will not function well in real-time situations. Relying just on accuracy is inadequate since accuracy does not ascertain the model's certainty score; rather, the loss value serves this purpose.

5. CONCLUSION

This research paper concludes that while accuracy is important in AI models for brain tumor detection, the certainty in predictions plays a significant role in reliable tumor diagnosis. The study established a correlation between a model's loss value and greater certainty in predictions, emphasizing the critical role of loss value alongside conventional performance metrics like precision and recall. The AI models, including CNN, ResNet-50, XceptionNet, and a proposed model, were tested, prioritizing loss values for accurate detection of tumor cases and minimizing false negatives. The proposed model achieved 90% accuracy with superior recall and runtime, demonstrating the effectiveness of the approach. The conclusion also highlights the potential of the proposed approach to address the global shortage of specialized medical professionals by providing timely and accurate cancer detection tools. This contributes effectively to healthcare systems and medical education, enabling future AI applications to be used effectively in clinical practice. The objective of this research is to investigate the use of artificial intelligence in medical education, To address the limitations of conventional approaches in terms of satisfying the varied requirements of modern students and medical practices in Pakistan.

The Conclusion outlines future research directions, including integrating EEG with MRI for multimodal prediction, expanding datasets across hospitals to enhance generalizability, and developing explainable AI models that combine certainty scores with interpretability. These additions address current limitations and provide a clear roadmap for advancing brain tumor detection using AI.

In doing so, it shows the potential of artificial intelligence technologies, such as DL, to provide robustness of AI in terms of disease detection. It emphasizes the need for a balanced approach and advises combining the analytical capabilities of artificial intelligence with the knowledge of humans to guarantee that the plan is both fair and successful. The objective of this study is to provide educators and medical professionals, with a roadmap that will help them navigate the complexity of using AI in medical education and practice. The work will concentrate on preserving academic integrity and inclusion while also capitalizing on the revolutionary potential. EEG signals have great potential in diagnosing Brain tumors. *Future work*: In the future, we will use EEG signals for the detection of cancer detection along with MRI images that will help diagnose cancer at an early stage. We will use MRI images and EEG images not only for cancer but for other brain-related diseases as well.

DECLARATIONS:

Ethics approval: This material is the authors' own original work, which has not been previously published elsewhere. The paper reflects the authors' own research and analysis in a truthful and complete manner.

Authors' contributorship: The author confirms sole responsibility for the following: study conception and design, data collection, analysis and interpretation of results, and manuscript preparation.

Acknowledgements:
We declare that this manuscript is original, has not been published before and is not currently being considered for publication elsewhere.

Competing Interests: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Availability of data and material: Data will be available when requested.

Funding: No funding was received to assist with the preparation of this manuscript.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

REFERENCES

[1] H. Kaur and A. K. Gill, "Review of Brain Tumor Detection Using Various Techniques," *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 7, no. 5, pp. 867– 870, May 2017, doi: 10.23956/IJARCSSE/V7I4/0221.

[2] R. L. Siegel, A. N. Giaquinto, and A. Jemal, "Cancer statistics, 2024," *CA Cancer J Clin*, vol. 74, no. 1, pp. 12–49, Jan. 2024, doi: 10.3322/caac.21820.

[3] A. M. Sarhan, "Brain Tumor Classification in Magnetic Resonance Images Using Deep Learning and Wavelet Transform," *J Biomed Sci Eng*, vol. 13, no. 06, pp. 102–112, 2020, doi: 10.4236/jbise.2020.136010.

[4] J. Schüz, K. Pirie, G. K. Reeves, S. Floud, and V. Beral, "Cellular Telephone Use and the Risk of Brain Tumors: Update of the UK Million

Women Study," *J Natl Cancer Inst*, vol. 114, no. 5, pp. 704–711, May 2022, doi: 10.1093/JNCI/DJAC042.

[5] W. Jun and Z. Liyuan, "Brain Tumor Classification Based on Attention Guided Deep Learning Model," *International Journal of Computational Intelligence Systems*, vol.15, no.1, Dec. 2022, doi: 10.1007/s44196-022-00090-9.

[6] A. Rehman, S. Naz, M. I. Razzak, F. Akram, and M. Imran, "A Deep Learning-Based Framework for Automatic Brain Tumors Classification Using Transfer Learning," *Circuits Syst Signal Process*, vol. 39, no. 2, pp. 757–775, Feb. 2020, doi: 10.1007/s00034-019-01246-3.

[7] A. Bin Tufail, N. Anwar, M.T.B. Othman, I. Ullah, R.A. Khan, Y.K. Ma, D. Adhikari, A.U. Rehman, M. Shafiq and **H. Hamam** (2022) "Early-Stage Alzheimer's Disease Categorization Using PET Neuroimaging Modality and Convolutional Neural Networks in the 2D and 3D Domains", Sensors (Q1, H-Index:245), **22**, 18 pages; doi.org/10.3390/s22124609

[8] A. S. Lundervold and A. Lundervold, "An overview of deep learning in medical imaging focusing on MRI," May 01, 2019, *Elsevier GmbH*. doi: 10.1016/j.zemedi.2018.11.002.

[9] L. Rundo *et al.*, "Semi-automatic Brain Lesion Segmentation in Gamma Knife Treatments Using an Unsupervised Fuzzy C- Means Clustering Technique," *Smart Innovation, Systems and Technologies*, vol. 54, pp. 15–26, 2016, doi: 10.1007/978-3-319- 33747-0_2.

[10] A. Raza et al. "A Hybrid Deep Learning-Based Approach for Brain Tumor Classification", Electronics, doi.org/10.3390/electronics11071146.

[11] S. Bonte, I. Goethals, and R. Van Holen, "Machine learning based brain tumour segmentation on limited data using local texture and abnormality," *Comput Biol Med*, vol. 98, pp. 39–47, Jul. 2018, doi: 10.1016/j.compbiomed.2018.05.005.

[12] C. Militello *et al.*, "Gamma Knife treatment planning: MR brain tumor segmentation and volume measurement based on unsupervised Fuzzy C-Means clustering," *Int J Imaging Syst Technol*, vol. 25, no. 3, pp. 213–225, Sep. 2015, doi: 10.1002/ima.22139.

[13] J. Juan-Albarracín *et al.*, "Automated glioblastoma segmentation based on a multiparametric structured unsupervised classification," *PLoS One*, vol. 10, no. 5, May 2015, doi: 10.1371/journal.pone.0125143.

[14] "Do Cell Phones Cause Cancer? | Cellphones and Cancer | American Cancer Society." Accessed: Sep. 28, 2024. [Online]. Available: https://www.cancer.org/cancer/risk-prevention/radiationexposure/cellular-phones.html

[15] U. Raghavendra *et al.*, "Brain tumor detection and screening using artificial intelligence techniques: Current trends and future perspectives," *Comput Biol Med*, vol. 163, p. 107063, Sep. 2023, doi: 10.1016/J.COMPBIOMED.2023.107063.

[16] H. A. Shah, F. Saeed, S. Yun, J. H. Park, A. Paul, and J. M. Kang, "A Robust Approach for Brain Tumor Detection in Magnetic Resonance Images Using Finetuned EfficientNet," *IEEE Access*, vol. 10, pp. 65426–65438, 2022, doi: 10.1109/ACCESS.2022.3184113.

[17] M. S. I. Khan *et al.*, "Accurate brain tumor detection using deep convolutional neural network," *Comput Struct Biotechnol J*, vol. 20, pp. 4733–4745, Jan. 2022, doi: 10.1016/j.csbj.2022.08.039.

[18] S. Dhawan and C. C. Chen, "Comparison meta-analysis of intraoperative MRI-guided needle biopsy versus conventional stereotactic needle biopsies," Jan. 01, 2024, *Oxford University Press*. doi: 10.1093/noajnl/vdad129.

[19] O. Hansson, S. Lehmann, M. Otto, H. Zetterberg, and P. Lewczuk, "Advantages and disadvantages of the use of the CSF Amyloid β (Aβ) 42/40 ratio in the diagnosis of Alzheimer's Disease," Apr. 22, 2019, *BioMed Central Ltd.* doi: 10.1186/s13195-019-0485-0.

[20] A. Gudigar, U. Raghavendra, T. R. San, E. J. Ciaccio, and R. Acharya, "Application of Multiresolution Analysis for Automated Detection of Brain Abnormality Using MR Images: A Comparative Study."

[21] H. P. Chan, L. M. Hadjiiski, and R. K. Samala, "Computer-aided diagnosis in the era of deep learning," in *Medical Physics*, John Wiley and Sons Ltd, Jun. 2020, pp. e218–e227. doi: 10.1002/mp.13764.

[22] G. Mohi ud din dar, et al. "A Novel Framework

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

for Classification of Different Alzheimer's Disease Stages Using CNN Model", Electronics, 23, doi: org/10.3390/electronics12020469.

[23] M. S. Tahosin, M. A. Sheakh, T. Islam, R. J. Lima, and M. Begum, "Optimizing brain tumor classification through feature selection and hyperparameter tuning in machine learning models," *Inform Med Unlocked*,vol.43,Jan.2023,doi:10.1016/j.imu.2023.101414.

[24] T. Azad Abir, A. Siraji, and E. Ahmed, "Analysis of a novel MRI Based Brain Tumour Classification Using Probabilistic Neural Network (PNN)," *IJSRSET*, 2018, [Online]. Available: www.ijsrset.com

[25] P. Modiya and S. Vahora, "Brain Tumor Detection Using Transfer Learning with Dimensionality Reduction Method," *Original Research Paper International Journal of Intelligent Systems and Applications in Engineering IJISAE*, vol. 2022, no. 2, pp. 201–206, 2022, doi: 10.18201/ijisae.

[26] A. Vidyarthi, R. Agarwal, D. Gupta, R. Sharma, D. Draheim, and P. Tiwari, "Machine Learning Assisted Methodology for Multiclass Classification of Malignant Brain Tumors," *IEEE Access*, vol. 10, pp. 50624–50640, 2022, doi: 10.1109/ACCESS.2022.3172303.

[27] J. Kang, Z. Ullah, and J. Gwak, "Mri-based brain tumor classification using ensemble of deep features and machine learning classifiers," *Sensors*, vol. 21, no. 6, pp. 1–21, Mar. 2021, doi: 10.3390/s21062222.

[28] A. A. Asiri *et al.*, "Machine Learning-Based Models for Magnetic Resonance Imaging (MRI)-Based Brain Tumor Classification," *Intelligent Automation and Soft Computing*, vol. 36, no. 1, pp. 299–312, 2023, doi: 10.32604/iasc.2023.032426.

[29] T. Shahwar et al. "Automated Detection of Alzheimer's via Hybrid Classical Quantum Neural Networks", Electronics 2022, doi.org/10.3390/electronics11050721

[30] A. Echtioui et al. "Classification of BCI Multiclass Motor Imagery Task Based on Artificial Neural Network", Clinical EEG and Neuroscience, 55(4):455-464, 2024, doi: org/10.1177/15500594221148285.

[31] Das Sunanda, Aranya Riaz, and Labiba Nishat, "Brain Tumor Classification Using Convolutional Neural Network," IEEE, 2019.

[32] K. Danukusumom =Pranowo, and M. Maslim, "Indonesia ancient temple classification using convolutional neural network," IEEE, 2017.

[33] S. Hussain, S. M. Anwar, and M. Majid, "Segmentation of Glioma Tumors in Brain Using Deep Convolutional Neural Network," Aug. 2017, doi: 10.1016/j.neucom.2017.12.032.

[34] D. C. Febrianto, I. Soesanti, and H. A. Nugroho, "Convolutional Neural Network for Brain Tumor Detection," in *IOP Conference Series: Materials Science and Engineering*, Institute of Physics Publishing, Mar. 2020. doi: 10.1088/1757-899X/771/1/012031.

[35] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," CVPR, 2016, [Online]. Available: http://imagenet.org/challenges/LSVRC/2015/

[36] C. Szegedy, V. Vanhoucke, S. Ioffe, and J. Shlens, "Rethinking the Inception Architecture for Computer Vision," *CVPR*, 2016, Accessed: Sep. 26, 2024. [Online]. Available: https://www.cvfoundation.org/openaccess/content_cvpr_2016/papers/Szegedy_Rethinking_the_Inception_CVPR_ 2016_paper.pdf

[37] F. Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions," *CVPR*, 2017.

[38] "Review: Xception — With Depthwise Separable Convolution, Better Than Inception-v3 (Image Classification) | by Sik- Ho Tsang | Towards Data Science." Accessed: Sep. 26, 2024. [Online]. Available: https://towardsdatascience.com/review-xception-with-depthwise-separableconvolution-better-than-inception-v3-image-dc967dd42568.

[39] "Beyond Accuracy: Mastering the Confusion Matrix for Advanced Model Evaluation | by SSP | Medium." Accessed: Sep. 26, 2024. [Online]. Available: https://medium.com/@_SSP/confusionmatrix-f/ff01c5bbb6.