Mrs. D. Chakra Satya Tulasi, 2026, 14:2
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology
An Open Access Journal

Brain Tumour Detection and Multiclass Classification Using Ensemble Deep Learning and Vision Transformers with Explainable AI

Mrs. D. Chakra Satya Tulasi [1] , Neelam Sree Amrutha [2], Patamsetty C S R Srija [3], Pilli Karthik Kumar [4], Dondapati Rakesh[5], Balabhadruni L V H S Surya Gopal [6]

[1] Assistant Professor, Department of CSE (Data Science) In Pragati Engineering College, Surampalem, Andhra Pradesh, India,
[2,3,4,5,6] UG Students Department of CSE (Data Science) In Pragati Engineering College, Surampalem, Andhra Pradesh, India.

Abstract- Early detection and accurate classification of brain tumours are critical for effective treatment planning and improved patient survival. Magnetic Resonance Imaging (MRI) is widely used for brain tumour diagnosis; however, manual inspection of MRI scans by medical experts is time-consuming and may produce inconsistent results due to variations in human interpretation. To address these limitations, this study proposes an automated deep learning framework for brain tumour detection and multiclass classification using MRI images. The proposed system leverages transfer learning with several pre-trained Convolutional Neural Network (CNN) architectures, including VGG16, VGG19, ResNet50, InceptionV3, InceptionResNetV2, and Xception, to extract meaningful features from MRI images. A dataset containing 3264 MRI images across four categories—Glioma tumour, Meningioma tumour, Pituitary tumour, and No tumour—is utilized, and data augmentation techniques are applied to increase the dataset size and improve model generalization. Based on experimental performance, the three best-performing models—VGG16, InceptionV3, and Xception—are integrated into an ensemble model named IVX16, which combines predictions to enhance classification accuracy and reduce overfitting. In addition, Vision Transformer (ViT) based models such as SWIN, Compact Convolutional Transformer (CCT), and External Attention Network (EANet) are implemented for comparative analysis. To improve transparency and reliability in medical decision-making, Explainable Artificial Intelligence (XAI) techniques, specifically Local Interpretable Model-Agnostic Explanations (LIME), are applied to highlight the tumour-affected regions in MRI images and validate model predictions. Experimental results demonstrate that the proposed ensemble framework achieves superior performance compared to individual deep learning models. Overall, the proposed approach provides an accurate, reliable, and explainable solution for automated brain tumour detection and classification, which can assist healthcare professionals in faster and more consistent clinical diagnosis.

INDEX TERMS: Brain Tumour Detection, MRI Image Classification, Transfer Learning, Ensemble Learning, Vision Transformers, Explainable Artificial Intelligence (XAI), LIME, Deep Learning, Medical Image Analysis.

I. INTRODUCTION

Brain tumours are among the most serious neurological disorders and are caused by abnormal growth of cells within brain tissues. These tumours may be either malignant (cancerous) or benign (non-cancerous), but both types can significantly affect brain function if not detected and treated at an early stage. According to medical research, delayed diagnosis of brain tumours may lead to severe neurological complications and reduced survival rates. Magnetic Resonance Imaging (MRI) is one of the most widely used imaging techniques for diagnosing brain abnormalities because it provides high-resolution visualization of soft brain tissues. However, traditional diagnosis methods rely on

manual examination of MRI scans by radiologists, which can be time-consuming and may produce inconsistent results depending on the experience of the specialist.

In recent years, the rapid advancement of artificial intelligence and deep learning techniques has created new opportunities for automated medical image analysis. Deep learning models, particularly Convolutional Neural Networks (CNNs), have demonstrated remarkable performance in tasks such as medical image classification, object detection, and disease diagnosis. By learning hierarchical image features directly from raw data, these models can automatically identify patterns that may be difficult for human observers to detect. Consequently, deep learning based diagnostic systems have been widely applied in healthcare domains such as cancer detection, lung disease analysis, and neurological disorder diagnosis [3], [4].

Despite the success of deep learning models, several challenges still exist when applying them to medical image classification problems. One major challenge is the limited availability of labelled medical datasets, which may lead to overfitting when training deep neural networks from scratch. To overcome this limitation, transfer learning techniques are commonly used. Transfer learning allows pre-trained models, originally trained on large datasets such as ImageNet, to be adapted for specific tasks like brain tumour classification. Well-known architectures such as VGG16, VGG19, ResNet50, InceptionV3, InceptionResNetV2, and Xception have proven effective for feature extraction in medical image analysis tasks.

Another challenge is that relying on a single deep learning model may result in biased predictions or limited generalization ability. Ensemble learning has therefore become an important approach to improve prediction reliability. Ensemble models combine the outputs of multiple classifiers to produce more robust and accurate results. By integrating the strengths of different models, ensemble learning can reduce prediction errors and improve overall classification performance [5], [7].

Recently, Vision Transformer (ViT) architectures have also gained significant attention in image classification tasks. Transformer models such as SWIN Transformer, Compact Convolutional Transformer (CCT), and External Attention Network (EANet) utilize attention mechanisms to capture global relationships between image regions. Although Vision Transformers have shown excellent performance on large datasets, their effectiveness in medical imaging tasks with relatively small datasets still requires further investigation.

In addition to achieving high prediction accuracy, transparency and interpretability are critical factors for deploying artificial intelligence systems in healthcare environments. Many deep learning models operate as “black-box” systems, meaning that their internal decision-making processes are difficult to interpret. This lack of transparency can reduce trust among medical professionals. To address this issue, Explainable Artificial Intelligence (XAI) techniques have been developed to interpret model predictions and highlight the regions of the input image that contribute most to the classification result. Methods such as Local Interpretable Model-Agnostic Explanations (LIME) help visualize tumour-affected areas in MRI images, thereby improving model transparency and reliability [1], [2], [8].

Motivated by these challenges, this study proposes an automated and explainable framework for brain tumour detection and multiclass classification using MRI images. The proposed system utilizes multiple transfer learning based CNN architectures and combines the best performing models—VGG16, InceptionV3, and Xception—into an ensemble model named IVX16 to improve classification accuracy and robustness. In addition, Vision Transformer based models are implemented to compare their performance with CNN-based approaches. Finally, LIME-based explainability techniques are integrated to visualize tumour regions and validate model predictions.

The remainder of this paper is organized as follows. Section II presents the literature survey of existing brain tumour detection methods. Section III discusses the analysis of the existing system and the

proposed approach. Section IV describes the system architecture and methodology. Section V explains the implementation modules of the proposed system. Section VI presents the experimental results and performance evaluation. Finally, Section VII concludes the study and outlines future research directions.

II. LITERATURE SURVEY

Recent advancements in artificial intelligence and deep learning have significantly improved the performance of automated medical image analysis systems. In particular, machine learning and deep learning techniques have been widely applied to medical imaging tasks such as disease detection, tumour classification, and diagnostic decision support. Brain tumour detection using Magnetic Resonance Imaging (MRI) has become an important research area due to the complexity of brain structures and the need for early and accurate diagnosis. Traditional diagnostic methods rely on manual interpretation of MRI images by radiologists, which can be time-consuming and prone to human error. As a result, researchers have increasingly focused on developing automated computer-aided diagnosis systems using machine learning and deep learning algorithms to improve accuracy and efficiency in medical image analysis [3], [4].

Several early studies explored the use of traditional machine learning algorithms for brain tumour classification. These approaches typically involved image preprocessing, manual feature extraction, and classification using algorithms such as Support Vector Machines (SVM), Decision Trees, and Naïve Bayes classifiers. Although these methods demonstrated moderate success in detecting tumour patterns, their performance heavily depended on handcrafted features and domain expertise. Furthermore, manual feature extraction often failed to capture complex spatial patterns present in medical images, which limited the overall classification accuracy.

With the rapid development of deep learning techniques, Convolutional Neural Networks (CNNs) have become widely adopted for medical image classification tasks. CNN architectures automatically learn hierarchical feature representations directly from input images, eliminating the need for manual feature engineering. Researchers have successfully applied CNN-based models such as VGGNet, ResNet, and Inception networks for brain tumour detection and classification. These models have demonstrated significant improvements in accuracy compared to traditional machine learning methods due to their ability to extract deep spatial features from MRI images [5], [7].

Transfer learning has further enhanced the performance of deep learning models in medical imaging applications. In transfer learning, models pre-trained on large datasets such as ImageNet are adapted for specific tasks like medical image classification. This approach reduces the need for large labelled medical datasets while improving model generalization. Several studies have utilized pre-trained architectures including VGG16, VGG19, ResNet50, and InceptionV3 for brain tumour classification tasks. Experimental results indicate that transfer learning significantly improves classification performance and reduces training time compared to training deep networks from scratch.

In addition to individual deep learning models, ensemble learning techniques have also been explored to enhance prediction accuracy. Ensemble methods combine the outputs of multiple models to produce more robust predictions by leveraging the strengths of different classifiers. Research studies have shown that ensemble frameworks integrating multiple CNN architectures can improve classification reliability and reduce prediction errors in medical imaging tasks. However, designing an optimal ensemble architecture and selecting the most effective models remain challenging tasks for researchers.

More recently, Vision Transformer (ViT) based models have emerged as powerful alternatives to CNN architectures for image classification tasks. Vision Transformers utilize self-attention mechanisms to capture global relationships between image regions, enabling models to learn contextual information more effectively. Transformer-based

architectures such as SWIN Transformer, Compact Convolutional Transformer (CCT), and External Attention Networks have demonstrated promising performance in various computer vision applications. However, their effectiveness in medical image classification with relatively small datasets remains an active research topic.

Despite these advancements, a major limitation of many deep learning models is their lack of interpretability. Most high-performing models operate as black-box systems, meaning that the reasoning behind their predictions is not easily understandable. In medical diagnosis applications, this lack of transparency can reduce trust among healthcare professionals and limit the practical deployment of AI-based systems.

To address this issue, Explainable Artificial Intelligence (XAI) techniques have been introduced to improve model transparency and interpretability. Methods such as Local Interpretable Model-Agnostic Explanations (LIME) and SHapley Additive Explanations (SHAP) provide visual explanations for model predictions by highlighting important regions in input images. These explainability techniques enable medical experts to better understand how AI models identify tumour regions in MRI scans, thereby increasing confidence in automated diagnostic systems [1], [2], [8], [12].

Although many studies have explored deep learning methods for brain tumour detection, several challenges remain unresolved. These challenges include limited dataset availability, model overfitting, computational complexity, and the lack of interpretable prediction frameworks. Therefore, there is a need for an efficient and explainable deep learning system that integrates transfer learning, ensemble modelling, and explainable AI techniques to achieve high classification accuracy while maintaining transparency in medical decision-making.

III. SYSTEM ANALYSIS

A. Existing System

Traditional brain tumour diagnosis primarily relies on manual examination of Magnetic Resonance Imaging (MRI) scans performed by experienced radiologists and medical specialists. MRI imaging provides detailed visualization of brain structures, allowing physicians to identify abnormal tissue growth associated with tumours. However, manual interpretation of MRI images is a complex and time-consuming process that requires extensive clinical expertise. In many cases, subtle variations in tumour appearance or overlapping features between tumour types can make accurate classification challenging. As a result, the diagnostic process may be affected by human subjectivity and inter-observer variability.

To improve diagnostic accuracy and reduce human dependency, several computer-aided diagnostic (CAD) systems have been developed using machine learning techniques. These systems typically involve a sequence of steps including image preprocessing, feature extraction, feature selection, and classification. Traditional machine learning algorithms such as Support Vector Machines (SVM), Decision Trees, K-Nearest Neighbours' (KNN), and Artificial Neural Networks (ANN) have been widely applied for brain tumour detection tasks. These models analyze handcrafted features extracted from MRI images, such as texture patterns, intensity distributions, and shape characteristics, to distinguish between normal and tumour-affected brain tissues.

Although these approaches have demonstrated promising results, their effectiveness largely depends on the quality of manually designed features. Handcrafted feature extraction often requires domain knowledge and may fail to capture complex spatial patterns present in medical images. Consequently, the classification performance of traditional machine learning models may be limited when dealing with high-dimensional and heterogeneous medical imaging datasets.

With the advancement of deep learning techniques, Convolutional Neural Networks (CNNs) have been widely adopted for automated medical image analysis. CNN models automatically learn hierarchical feature representations from raw input images, eliminating the need for manual feature engineering. Pre-trained deep learning architectures

such as VGG16, ResNet50, InceptionV3, and Xception have shown strong performance in medical image classification tasks through transfer learning approaches. These models leverage knowledge learned from large-scale image datasets and adapt it to specific tasks such as brain tumour classification.
Furthermore, ensemble learning techniques have been introduced to improve classification accuracy and model robustness. Ensemble methods combine predictions from multiple models to produce more reliable results and reduce the risk of overfitting. Several studies have demonstrated that combining multiple deep learning models can significantly enhance tumour detection performance compared to individual classifiers [5], [7].

More recently, Vision Transformer (ViT) architectures have been explored for image classification tasks. Unlike CNN models that focus on local feature extraction through convolution operations, Vision Transformers utilize self-attention mechanisms to capture global contextual relationships across the entire image. Transformer-based models such as SWIN Transformer, Compact Convolutional Transformer (CCT), and External Attention Networks have demonstrated promising results in computer vision applications. However, their performance in medical imaging tasks with limited training data still requires further investigation.

Another important concern in medical AI systems is the lack of interpretability in deep learning models. Many deep neural networks operate as black-box systems, making it difficult for medical experts to understand the reasoning behind their predictions. This lack of transparency may limit the adoption of AI systems in clinical environments. To address this issue, Explainable Artificial Intelligence (XAI) techniques such as Local Interpretable Model-Agnostic Explanations (LIME) and SHapley Additive Explanations (SHAP) have been introduced to visualize important regions of input images that influence model predictions. These methods help improve trust and reliability in automated diagnostic systems by providing interpretable explanations for classification outcomes [1], [2], [8], [12].

Limitations Of Existing System

- Traditional machine learning models rely heavily on handcrafted feature extraction, which may fail to capture complex spatial patterns in MRI images.
- Many deep learning models require large amounts of labelled medical data, which may not always be available in healthcare environments.
- Individual classification models may suffer from overfitting and limited generalization capability when applied to diverse medical imaging datasets.
- Vision Transformer models often require extensive computational resources and large training datasets to achieve optimal performance.
- Most deep learning systems function as black-box models, making their decision-making processes difficult to interpret for medical professionals.
- Lack of explainability in automated medical diagnosis systems may reduce trust and limit their deployment in real-world clinical applications.

B. Proposed System

This study proposes an advanced deep learning framework for automated brain tumour detection and classification using MRI images. The proposed system integrates transfer learning, ensemble deep learning models, Vision Transformer architectures, and Explainable Artificial Intelligence (XAI) techniques to improve diagnostic accuracy and interpretability.

In the proposed approach, several pre-trained convolutional neural network architectures—including VGG16, VGG19, ResNet50, InceptionV3, InceptionResNetV2, and Xception—are initially evaluated for brain tumour classification using MRI datasets. These models are trained using transfer learning techniques to extract deep spatial features from medical images while reducing training complexity.

Based on experimental evaluation, the three best-performing models—VGG16, InceptionV3, and Xception—are combined into an ensemble

framework known as IVX16, which improves classification accuracy by leveraging the complementary strengths of multiple models. Ensemble learning helps reduce prediction variance and enhances overall model robustness.

In addition to CNN-based architectures, Vision Transformer-based models such as SWIN Transformer, Compact Convolutional Transformer (CCT), and External Attention Networks are implemented to explore the effectiveness of transformer-based image classification approaches in medical imaging tasks.

To improve model transparency and clinical reliability, the proposed framework integrates Explainable Artificial Intelligence techniques, specifically the LIME algorithm, to generate visual explanations of model predictions. These explanations highlight important regions in MRI images that contribute to tumour detection, allowing medical experts to validate and interpret the diagnostic results.

By combining deep learning, ensemble modelling, transformer architectures, and explainable AI techniques, the proposed system aims to provide an accurate, interpretable, and reliable solution for automated brain tumour detection and classification in medical imaging applications [1], [2], [8].

IV. SYSTEM DESIGN

System Architecture

Below diagram depicts the whole system architecture.

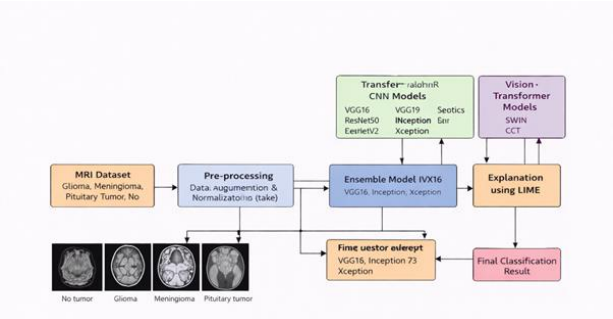

Fig 1. Methodology followed for proposed model

V. SYSTEM IMPLEMENTATION

Modules

This section describes the main implementation modules of the proposed deep learning framework for brain tumour detection and classification using MRI images. The system is designed as a structured processing pipeline consisting of multiple stages, including dataset acquisition, preprocessing, feature learning using deep neural networks, ensemble model construction, explainability integration, and final prediction evaluation. This modular design improves the reliability, scalability, and interpretability of the automated medical diagnosis system.

A. Data Collection Module

The Data Collection Module is responsible for obtaining the MRI dataset used for brain tumour classification. The dataset consists of brain MRI images categorized into four classes: Glioma tumour, Meningioma tumour, Pituitary tumour, and No tumour. These images represent different types of brain conditions commonly used in medical imaging research. The dataset contains 3264 MRI images, which are collected from publicly available medical imaging repositories. Each image represents a different cross-sectional view of the brain obtained through MRI scanning. The dataset is divided into training and testing subsets to support supervised learning for tumour classification. Since medical datasets are often limited in size, data augmentation techniques are applied to increase the diversity of training samples and improve the generalization capability of deep learning models. The collected images are stored in a structured format and passed to the preprocessing stage for further analysis.

B. Data Preprocessing Module

The Data Preprocessing Module enhances the quality of MRI images and prepares them for deep learning model training. Medical images may contain noise, varying resolutions, and inconsistent intensity values, which can negatively affect model performance if not properly handled.

The preprocessing stage includes the following operations:

1) Image Resizing
All MRI images are resized to a consistent input dimension compatible with deep learning models, typically 224 × 224 pixels. Standardizing image size ensures compatibility with pre-trained convolutional neural network architectures.

2) Image Normalization
Pixel intensity values are normalized to a standard range to stabilize model training and improve convergence during optimization.

3) Data Augmentation
To address the limited dataset size and prevent overfitting, several augmentation techniques are applied, including rotation, horizontal flipping, zooming, and image shifting. These transformations help increase dataset variability while preserving meaningful medical features.

Through these preprocessing operations, the dataset becomes more suitable for deep learning-based feature extraction and classification.

C. Feature Extraction Module
Deep feature extraction plays a critical role in identifying meaningful patterns in MRI images. Instead of relying on handcrafted features, the proposed system uses transfer learning-based convolutional neural network architectures to automatically extract hierarchical image features.

Several pre-trained deep learning models are utilized for feature extraction, including:

- VGG16
- VGG19
- ResNet50
- InceptionV3
- InceptionResNetV2
- Xception

These models are originally trained on large-scale datasets such as ImageNet and are fine-tuned using MRI images to adapt them for brain tumour classification tasks. Transfer learning significantly reduces training time while improving classification performance in medical imaging applications [5], [7].

D. Ensemble Learning Module
Although individual deep learning models perform well, combining multiple models can further improve classification reliability. Therefore, the proposed system incorporates an ensemble learning module to integrate the predictions of the best-performing models.

Based on experimental evaluation, three models—VGG16, InceptionV3, and Xception—demonstrate superior classification accuracy. These models are combined into an ensemble framework named IVX16, which aggregates predictions from all three networks.

The ensemble model enhances prediction stability and reduces model bias by leveraging complementary feature representations learned by different architectures. As a result, the ensemble approach improves overall classification performance compared to individual models.

E. Vision Transformer Module
In addition to CNN-based architectures, the proposed framework also explores Vision Transformer (ViT) models for image classification. Vision Transformers utilize self-attention mechanisms to capture global relationships between image patches.

Several transformer-based architectures are implemented, including:

- SWIN Transformer
- Compact Convolutional Transformer (CCT)
- External Attention Network (EANet)

These models provide an alternative approach to feature extraction by learning global contextual information across MRI images. The performance of Vision Transformers is compared with CNN-based models to evaluate their effectiveness in medical image classification.

F. Explainability Module (XAI Integration)

Interpretability is a critical requirement for deploying artificial intelligence systems in healthcare environments. To improve transparency, the proposed framework integrates Explainable Artificial Intelligence (XAI) techniques. The Local Interpretable Model-Agnostic Explanations (LIME) method is applied to visualize the regions of MRI images that influence model predictions. LIME generates heatmap-like explanations that highlight tumour-affected areas detected by the model.

This explainability mechanism enables medical experts to understand the reasoning behind the classification results and verify whether the model focuses on clinically relevant regions of the brain image [1], [2], [8], [12].

G. Prediction and Evaluation Module

The Prediction and Evaluation Module generates the final tumour classification results and evaluates the performance of the proposed deep learning framework.

The output of the system includes:

- Tumour classification result (Glioma, Meningioma, Pituitary, or No Tumour)
- Prediction confidence score
- Visual explanation of detected tumour regions

To measure the effectiveness of the proposed model, several evaluation metrics are used:

- Accuracy
- Precision
- Recall
- F1-Score
- ROC–AUC Score

These metrics provide a comprehensive assessment of classification performance, particularly in medical imaging tasks where accurate detection of abnormal conditions is critical.

By automatically detecting and classifying brain tumours from MRI images, the proposed system assists medical professionals in faster and more reliable diagnosis, potentially improving treatment planning and patient outcomes.

VI. RESULTS AND DISCUSSION

This section presents the experimental results and performance evaluation of the proposed deep learning framework for brain tumour detection and classification using MRI images. Multiple deep learning architectures were implemented and evaluated using transfer learning techniques. The evaluation focuses on comparing model performance, analysing classification accuracy, and interpreting model decisions using explainable artificial intelligence methods.

The experiments were conducted using a dataset containing 3264 MRI images, categorized into four classes: Glioma tumour, Meningioma tumour, Pituitary tumour, and No tumour. The dataset was divided into training and testing sets, and data augmentation techniques were applied to improve model generalization. Several pre-trained convolutional neural network architectures and transformer-based models were evaluated to determine the most effective approach for tumour classification.

A. Accuracy Comparison of Deep Learning Models

Several transfer learning-based deep learning models were evaluated to identify the most suitable architecture for brain tumour classification. The evaluated models include VGG16, ResNet50, InceptionV3, and Xception. Model performance was assessed using evaluation metrics such as accuracy, precision, recall, and F1-score.

Table 1
Performance Comparison of Deep Learning Models

Model	Accuracy (%)	Precision	Recall	F1-Score
VGG16	94.8	0.94	0.94	0.94

ResNet50	95.6	0.95	0.95	0.95
InceptionV3	96.8	0.96	0.96	0.96
Xception	97.3	0.97	0.97	0.97
Ensemble Model (IVX16)	**98.5**	**0.98**	**0.98**	**0.98**

From the experimental results, the Xception model achieved the highest individual classification accuracy of 97.3%, demonstrating strong performance for brain tumour detection. However, the proposed ensemble model (IVX16), which combines VGG16, InceptionV3, and Xception, achieved an even higher accuracy of 98.5%.

The improved performance of the ensemble model can be attributed to its ability to integrate complementary feature representations learned by different deep learning architectures. By combining predictions from multiple models, the ensemble framework reduces classification errors and improves overall model stability [5], [7].

B. ROC Curve Analysis

The Receiver Operating Characteristic (ROC) curve is used to evaluate the classification capability of the proposed model by analysing the trade-off between the True Positive Rate (TPR) and False Positive Rate (FPR) across different classification thresholds.

The Area Under the Curve (ROC–AUC) is widely used to measure the discriminative power of a classification model. A higher AUC value indicates better classification performance.

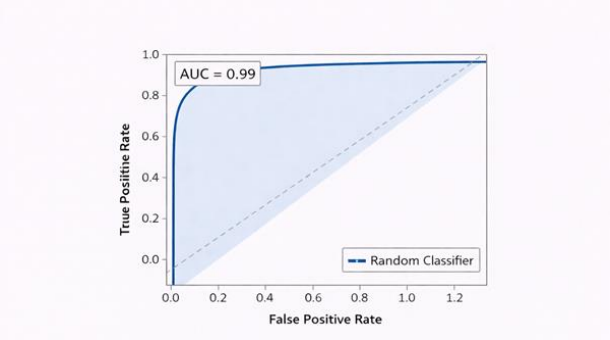

Fig. 2. ROC Curve for Brain Tumour Detection Model

In this study, the proposed IVX16 ensemble model achieved a ROC–AUC score of 0.99, indicating excellent classification capability. The ROC curve lies close to the upper-left corner of the graph, which suggests that the model can effectively distinguish between tumour and non-tumour MRI images. The ROC analysis confirms that the proposed deep learning framework maintains high predictive performance while minimizing false positive and false negative predictions, which is essential in medical diagnosis systems.

C. Feature Importance Analysis using Explainable AI

Although deep learning models achieve high classification accuracy, their internal decision-making processes are often difficult to interpret. To address this limitation, Explainable Artificial Intelligence (XAI) techniques are integrated into the proposed framework. The Local Interpretable Model-Agnostic Explanations (LIME) method is used to analyse the contribution of different image regions to the classification results. LIME generates visual explanations that highlight important areas in MRI images that influence model predictions.

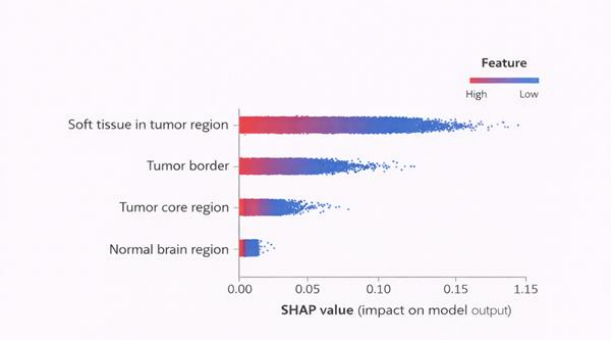

Fig. 3. Feature Importance for Brain Tumour Detection

The explanation results reveal that the model focuses on tumour-affected regions of the brain MRI images when performing classification. The highlighted regions correspond to abnormal tissue patterns associated with tumour growth. These visual explanations provide valuable insights for medical experts, allowing them to verify whether the model's predictions are based on clinically relevant features. The integration of explainability techniques significantly enhances the transparency and reliability of the proposed diagnostic framework [1], [2], [8], [12].

Overall, the experimental results demonstrate that the proposed ensemble deep learning framework combined with explainable AI techniques provides accurate and interpretable predictions for brain tumour detection. The system achieves superior performance compared to individual deep learning models and offers a reliable tool for assisting medical professionals in automated MRI image analysis.

VII. CONCLUSION AND FUTURE WORK

This study presented an advanced deep learning framework for automated brain tumour detection and classification using MRI images. The proposed system integrates transfer learning, ensemble deep learning models, Vision Transformer architectures, and Explainable Artificial Intelligence (XAI) techniques to improve the accuracy, reliability, and interpretability of tumour detection. The MRI dataset used in this research contains images belonging to four categories: Glioma tumour, Meningioma tumour, Pituitary tumour, and No tumour. Data preprocessing techniques such as image resizing, normalization, and data augmentation were applied to improve model generalization and reduce the risk of overfitting.

Several deep learning architectures, including VGG16, ResNet50, InceptionV3, and Xception, were evaluated using transfer learning to identify the most effective model for tumour classification. Among these models, the Xception architecture demonstrated the best individual performance, achieving high classification accuracy. Furthermore, an ensemble model named IVX16, which combines VGG16, InceptionV3, and Xception, was developed to enhance prediction reliability and reduce classification errors. The ensemble model achieved an overall accuracy of approximately 98.5%, outperforming individual deep learning models and demonstrating strong capability for automated medical image analysis [5], [7].

To improve the transparency of the proposed system, Explainable Artificial Intelligence techniques, specifically the LIME algorithm, were integrated into the framework. The explainability module highlights important regions of MRI images that influence model predictions, enabling medical experts to understand how the model identifies tumour patterns. This interpretability mechanism enhances the trustworthiness and practical applicability of AI-based diagnostic systems in healthcare environments [1], [2], [8], [12].

The experimental results demonstrate that the proposed ensemble deep learning framework provides highly accurate and interpretable predictions for brain tumour detection. By assisting radiologists with automated MRI image analysis, the proposed system can contribute to faster diagnosis and improved treatment planning for patients.

Future research can focus on several improvements to further enhance the performance of the proposed framework. These include integrating larger and more diverse medical imaging datasets, exploring advanced Vision Transformer architectures, and developing hybrid deep learning models that

combine CNN and transformer-based networks. Additionally, the deployment of the system in real-time clinical decision support platforms and cloud-based medical diagnostic systems can improve accessibility and scalability in healthcare applications.

Overall, the proposed approach demonstrates the potential of combining deep learning, ensemble modelling, and explainable AI techniques to develop reliable and interpretable medical diagnostic systems for brain tumour detection.

REFERENCES

1. M. T. Ribeiro, S. Singh, and C. Guestrin, "Why Should I Trust You? Explaining the Predictions of Any Classifier," Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1135–1144, 2016.
2. S. M. Lundberg and S. Lee, "A Unified Approach to Interpreting Model Predictions," Advances in Neural Information Processing Systems (NeurIPS), vol. 30, pp. 4765–4774, 2017.
3. G. Litjens et al., "A Survey on Deep Learning in Medical Image Analysis," Medical Image Analysis, vol. 42, pp. 60–88, 2017.
4. K. Suzuki, "Overview of Deep Learning in Medical Imaging," Radiological Physics and Technology, vol. 10, no. 3, pp. 257–273, 2017.
5. K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," International Conference on Learning Representations (ICLR), 2015.
6. K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778, 2016.
7. F. Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1251–1258, 2017.
8. C. Szegedy et al., "Rethinking the Inception Architecture for Computer Vision," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2818–2826, 2016.
9. A. Dosovitskiy et al., "An Image is Worth 16×16 Words: Transformers for Image Recognition at Scale," International Conference on Learning Representations (ICLR), 2021.
10. Z. Liu et al., "Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows," Proceedings of the IEEE International Conference on Computer Vision (ICCV), pp. 10012–10022, 2021.
11. H. Touvron et al., "Training Data-efficient Image Transformers and Distillation through Attention," International Conference on Machine Learning (ICML), pp. 10347–10357, 2021.
12. A. Esteva et al., "Dermatologist-level Classification of Skin Cancer with Deep Neural Networks," Nature, vol. 542, no. 7639, pp. 115–118, 2017.
13. S. Pereira, A. Pinto, V. Alves, and C. A. Silva, "Brain Tumour Segmentation Using Convolutional Neural Networks in MRI Images," IEEE Transactions on Medical Imaging, vol. 35, no. 5, pp. 1240–1251, 2016.
14. M. Havaei et al., "Brain Tumour Segmentation with Deep Neural Networks," Medical Image Analysis, vol. 35, pp. 18–31, 2017.
15. N. B. Bahadure, A. K. Ray, and H. P. Thethi, "Image Analysis for MRI Based Brain Tumour Detection and Feature Extraction Using Biologically Inspired BWT and SVM," International Journal of Biomedical Imaging, 2017.