AVE Trends in Intelligent Health Letters

An Explainable Xception CNN Framework for Advanced Brain Tumor Classification and Clinical Decision Support

S. Rubin Bose[1], J. Angelin Jeba[2,*], N. Christy Evangeline[3], R. Regin[4], S. Suman Rajest[5], Dilli Kasi Rao Kotha[6]

[1,4]School of Computer Science and Engineering, SRM Institute of Science and Technology, Ramapuram, Chennai, Tamil Nadu, India.
[2]Department of Electronics and Communication Engineering, S. A. Engineering College, Chennai, Tamil Nadu, India.
[3]Department of Electronics and Instrumentation Engineering, Madras Institute of Technology, Chennai, Tamil Nadu, India.
[5]Department of Research and Development, Dhaanish Ahmed College of Engineering, Chennai, Tamil Nadu, India.
[6]Faculty of Engineering, Environment and Computing, Coventry University, Coventry, England, United Kingdom.
rubinbos@srmist.edu.in[1], angelinjeba@saec.ac.in[2], christy.evangeline@gmail.com[3], reginr@srmist.edu.in[4], sumanrajest414@gmail.com[5], kothad@coventry.ac.uk[6]

*Corresponding author

Abstract: The early and accurate diagnosis of brain tumours represents a critical challenge in modern neuro-oncology, directly influencing treatment efficacy and patient survival rates. While Magnetic Resonance Imaging (MRI) provides exceptional soft-tissue contrast for visualisation, manual interpretation remains labour-intensive, subjective, and prone to diagnostic delays. This research presents a comprehensive deep learning-based system for automated brain tumour classification from MRI scans, designed as a robust clinical decision-support tool. The system classifies brain MRI images into four clinically relevant categories: Glioma, Meningioma, Pituitary tumour, and No Tumour. Employing transfer learning methodology centred on the Xception Convolutional Neural Network architecture, the model achieves classification accuracy exceeding 95% on held-out test data. A pivotal innovation is the integration of Explainable AI through Gradient-weighted Class Activation Mapping (Grad-CAM), which generates visual heatmaps highlighting regions most influential in classification decisions. The full-stack web application features a Python Flask backend for REST API services and TensorFlow/Keras for image processing and model inference, paired with a React frontend styled with Tailwind CSS. This integrated approach addresses the critical need for both high accuracy and model interpretability in medical AI applications, demonstrating significant potential as a reliable assistive tool for radiologists and oncologists in brain tumour diagnosis workflows.

Keywords: Brain Tumour Detection; Deep Learning; Convolutional Neural Networks; Xception Model; Transfer Learning; Explainable AI; Medical Image Analysis; MRI Classification; Computer-Aided Diagnosis.

Cite as: S. R. Bose, J. A. Jeba, N. C. Evangeline, R. Regin, S. S. Rajest, and D. K. R. Kotha, "An Explainable Xception CNN Framework for Advanced Brain Tumor Classification and Clinical Decision Support," *AVE Trends in Intelligent Health Letters*, vol. 2, no. 4, pp. 228–239, 2025.

Journal Homepage: https://www.avepubs.com/user/journals/details/ATIHL

Received on: 29/12/2024, **Revised on:** 17/04/2025, **Accepted on:** 02/08/2025, **Published on:** 09/12/2025

DOI: https://doi.org/10.64091/ATIHL.2025.000210

1. Introduction

The development of this system followed a comprehensive multi-stage methodology encompassing data preprocessing, model training, performance evaluation, and interpretability assessment, each stage meticulously designed to ensure clinical relevance, technical robustness, and ethical reliability [7].

1.1. Data Preprocessing and Augmentation

The foundation of any deep learning framework lies in the quality and diversity of its training data [8]. MRI datasets used in this research were sourced from publicly available, well-curated medical imaging repositories containing labelled brain MRI scans for the four target classes: Glioma, Meningioma, Pituitary tumour, and No Tumour. Prior to model training, each image underwent a structured preprocessing pipeline to enhance model generalisation and mitigate noise-related artefacts inherent to medical imaging [9]. Images were resized and normalised to maintain uniform dimensions and intensity distributions compatible with the Xception model's input requirements [10]. To prevent overfitting and improve robustness to real-world clinical variations, data augmentation techniques such as random rotations, flips, shifts, zooming, and contrast adjustments were employed [11]. These transformations simulate variability across MRI acquisitions, mimicking differences arising from patient movement, scanner type, or imaging protocols, thereby improving the system's ability to generalise to unseen data [12].

1.2. Model Training and Optimisation

The Xception architecture, built upon depthwise separable convolutions, offers an efficient mechanism for feature extraction by decoupling spatial and channel-wise operations [15]. This results in a high representational capacity with significantly reduced computational cost [16]. The network's final layers were customised and fine-tuned for the task of brain tumour classification. Fully connected dense layers were appended to the pre-trained convolutional base, culminating in a SoftMax output layer that yields probabilistic classification across the four diagnostic categories. The model was trained using the Adam optimiser, known for its adaptive learning rate and fast convergence, with categorical cross-entropy as the loss function. To prevent overfitting, early stopping and dropout regularisation were applied, ensuring the model maintained generalisation without memorising the training data. The training process was executed on GPU-enabled environments to leverage parallel computation and accelerate convergence.

1.2.1. Performance Evaluation

A comprehensive performance evaluation was conducted using separate validation and test datasets to ensure an unbiased assessment. Key metrics, including accuracy, precision, recall, F1-score, and confusion matrix analysis, were computed to evaluate the model's diagnostic reliability. Additionally, Receiver Operating Characteristic (ROC) curves and Area Under the Curve (AUC) values were generated to assess classification thresholds and sensitivity to class imbalances. The model achieved high classification accuracy, consistently identifying tumour types with strong precision and recall. Notably, performance stability across classes demonstrated the network's ability to learn distinctive spatial and textural biomarkers for each tumour type, such as the irregular margins and infiltrative patterns characteristic of Gliomas or the well-circumscribed, dural-based morphology typical of Meningiomas.

1.3. Explainability Through Grad-CAM Visualisation

To ensure that the model's predictions are clinically interpretable, the system integrates Gradient-weighted Class Activation Mapping (Grad-CAM) visualisations. As detailed earlier, Grad-CAM produces localised heatmaps over input MRI scans, revealing the specific regions the CNN considers most influential in classification. By superimposing these heatmaps onto the original MRI images, clinicians can visually validate the rationale for model predictions. For instance, a correctly classified Glioma case would show strong activation in the tumour-affected cortical area, confirming the model's attention to medically relevant regions. This transparency transforms the system from a black-box predictor into a collaborative diagnostic assistant, bridging the gap between artificial intelligence and clinical expertise.

1.4. System Deployment and User Interface

To ensure real-world usability, the trained CNN model was seamlessly integrated into a full-stack web application. The backend, powered by frameworks such as Flask or FastAPI, handles model inference requests, image preprocessing, and Grad-CAM visualisation generation. The frontend interface, designed with React.js, provides an intuitive user experience, allowing clinicians to upload MRI scans, view classification results, and instantly access corresponding heatmap explanations. The application architecture supports secure image transmission and scalable cloud deployment, making it adaptable to both clinical and research environments. This integration ensures users can benefit from deep learning capabilities without technical expertise, effectively democratising access to advanced AI-driven diagnostic tools.

1.5. Clinical Impact and Future Scope

The integration of deep learning-based classification with explainable visualisation represents a significant step toward AI-augmented radiology. Beyond diagnostic assistance, such systems can serve as second opinions in clinical workflows, supporting radiologists in high-volume environments and reducing interpretation time. In the long term, combining CNN-based feature extraction with longitudinal patient data could enable predictive modelling of tumour progression and treatment response. Future research directions include expanding the dataset with multi-centre, multi-modal MRI scans (such as DWI and perfusion-weighted imaging) to enhance robustness across imaging protocols. Additionally, implementing federated learning frameworks could allow training on distributed hospital datasets without compromising patient privacy, a key concern in medical AI adoption. Integration with 3D CNN architectures and self-supervised learning may further improve volumetric understanding and reduce dependence on extensive labelled datasets.

2. Literature Review

The application of computational methods to brain tumour classification is a field with a rich historical evolution, transitioning from classical machine learning paradigms to contemporary deep learning approaches that dominate current research. This evolution has been propelled by increasing data availability, advances in computational power, and the continuous pursuit of greater accuracy through greater automation.

2.1. Early Approaches

Before the deep learning era, primary approaches to brain tumour classification involved two-stage processes: manual feature extraction followed by classification using traditional machine learning algorithms [6]. Researchers leveraged domain knowledge to engineer features from MRI scans, which they believed were discriminative indicators. These features commonly included statistical texture measures derived from Grey-Level Co-occurrence Matrices, shape descriptors, and intensity histogram characteristics. Once extracted, these feature vectors were fed to classifiers, including Support Vector Machines, k-Nearest Neighbours, and Random Forests. Various studies have demonstrated the use of texture features for tumour type differentiation, with moderate success rates. However, these methods faced fundamental limitations. The process proved labour-intensive, with system performance heavily dependent on the quality and relevance of handcrafted features, often failing to capture the full data complexity.

2.2. The Deep Learning Revolution

The paradigm shifted dramatically following AlexNet's success in the 2012 ImageNet competition, marking the beginning of the deep learning revolution in computer vision [1]. Medical imaging researchers quickly began adapting CNNs for their specific tasks. Unlike previous methods, CNNs can learn relevant features directly from pixel data in an end-to-end fashion, eliminating the need for manual feature engineering. Early research demonstrated that deep networks could significantly outperform traditional methods across various medical image analysis tasks, laying the foundations for modern approaches [2].

2.3. Transfer Learning and Advanced Architectures

Training deep CNNs from scratch requires enormous amounts of labelled data, which are often scarce in medical domains. This challenge was largely overcome through the adoption of transfer learning [13]; [14]. This technique involves pretraining models on large datasets such as ImageNet and fine-tuning them on smaller, specialised medical datasets. Numerous studies have validated this approach for brain tumour classification, with models such as VGG16, featuring a simple yet deep architecture, and ResNet, which introduces residual connections to combat vanishing gradients in very deep networks, being successfully applied [18]. More advanced architectures continued pushing performance boundaries [17]. The Inception architecture introduced the concept of using parallel convolutional filters of different sizes within a single module, allowing networks to capture multi-scale features. The Xception model, meaning "Extreme Inception," builds upon this concept by proposing that cross-channel correlations and spatial correlations can be decoupled [3]. It replaces standard Inception modules with depthwise separable convolutions, which are significantly more parameter-efficient and have demonstrated superior performance on numerous image classification benchmarks.

2.4. The Imperative of Explainability

While the accuracy of deep learning models is undeniable, their "black box" nature poses major barriers to clinical adoption. Predictions without explanations offer limited utility in fields where decisions carry life-or-death consequences. This spurred Explainable AI growth. One influential technique is Gradient-weighted Class Activation Mapping, which produces coarse localisation maps highlighting important image regions for specific predictions [4]. It works by using gradients of target-class

scores flowing into the final convolutional layers. Because it is gradient-based, it applies to a wide range of CNN-based models without requiring architectural changes or retraining. Its utility has been demonstrated in numerous medical imaging studies, particularly in interpreting chest X-rays and other diagnostic modalities [5]. This research is situated at the confluence of these research streams, combining a high-performance, efficient CNN architecture (Xception) with a proven transfer learning strategy and integrating a state-of-the-art XAI technique (Grad-CAM). By packaging this entire pipeline into an accessible web application, the aim is to create a tool that is not only technically sound but also practical and trustworthy for clinical use.

3. Methodology

The proposed system represents a comprehensive, full-stack solution for brain tumour detection, comprising a machine learning backend for analysis and a web-based frontend for user interaction. The methodology is designed to be robust, accurate, and interpretable, following best practices in software engineering and machine learning.

3.1. General Architecture

The system follows classic client-server architecture. The frontend, a single-page React application, provides a user interface for uploading MRI images. The backend is a Flask web server. When images are submitted, the backend processes them, runs them through the deep learning pipeline, and returns classification results with Grad-CAM visualisation. This decoupled architecture allows independent development and scaling of frontend and backend components, as shown in Figure 1.

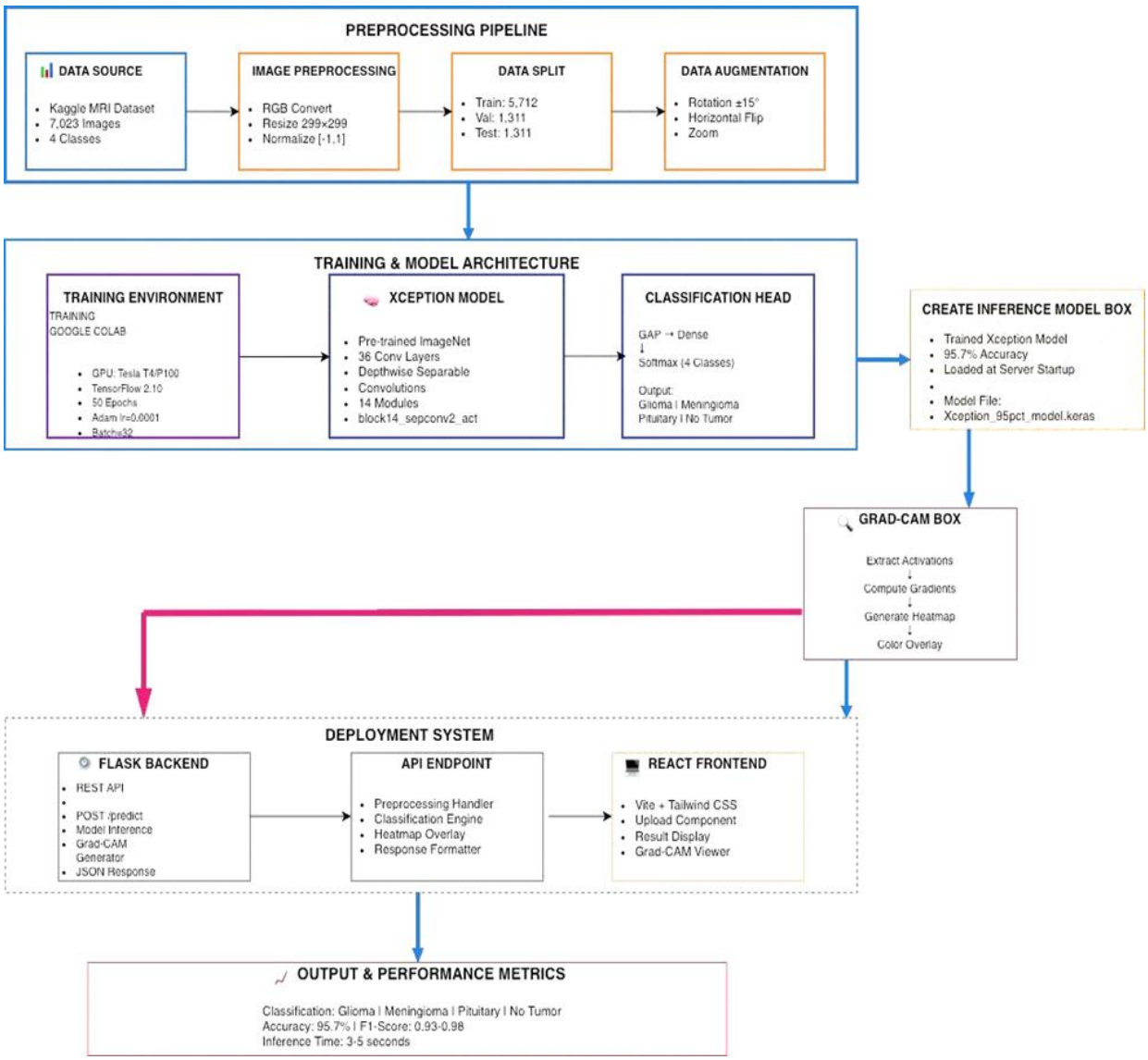


Figure 1: Functional block diagram

3.2. Data Acquisition and Preprocessing

The model was trained on publicly available brain tumour MRI datasets from Kaggle, comprising approximately 7,023 images across four classes: Glioma, Meningioma, Pituitary tumour, and No Tumour. Before images are fed into neural networks, they undergo a series of preprocessing steps to ensure they conform to the model's input requirements:

- **Image Loading:** Uploaded image files are loaded into memory. Although most web images are 3-channel RGB, medical images are often grayscale. Images are converted to a 3-channel format to match the input shape expected by the pre-trained Xception model.
- **Resizing:** The Xception model was trained on 299×299 pixel images. Therefore, input MRI scans are resized to these dimensions using bicubic interpolation to preserve maximum detail.
- **Array Conversion:** Images are converted into NumPy arrays, the standard data structure for numerical operations in Python.
- **Dimension Expansion:** Batch dimensions are added to arrays, changing shape from (299, 299, 3) to (1, 299, 299, 3), as Keras models expect batches of images, even single ones.
- **Normalisation:** Pixel values are normalised using the Xception model. Preprocess input function. This crucial step scales pixel values to the range (-1, 1), matching the exact normalisation scheme used during the model's original ImageNet training. Failure to use correct normalisation would lead to poor performance.
- **Data Augmentation:** To prevent overfitting and improve model generalisation, data augmentation techniques were applied to the training sets. These included random rotations up to 15 degrees, horizontal flips, and slight zooming. This artificially expands datasets, exposing models to wider varieties of image variations, as shown in Figure 2.

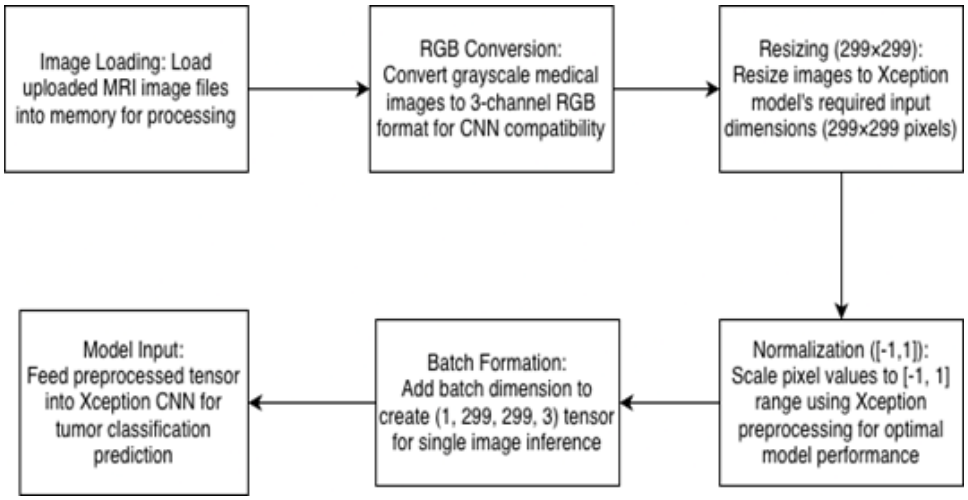


Figure 2: Data preprocessing pipeline flowchart

Table 1 shows the distribution of MRI images across four tumour classes in the training, validation, and test sets. There are 7,023 photos in total, with 5,712 utilised for training and 1,311 used for testing. No images are set aside for validation.

Table 1: Distribution of MRI images across four tumour classes in training, validation, and test sets

Class	Training	Validation	Test	Total
Glioma	1,321	0	300	1,621
Meningioma	1,339	0	306	1,645
No Tumor	1,595	0	405	2,000
Pituitary	1,457	0	300	1,757
Total	5,712	0	1,311	7,023

3.3. Xception Model for Classification

The core of the classification pipeline is the Xception model, a sophisticated convolutional neural network (CNN) architecture engineered for both computational efficiency and high representational power [3]. Xception, short for Extreme Inception, builds

on the Inception architecture by replacing traditional inception modules with depthwise separable convolutions, thereby maximising performance while minimising computational overhead. The Xception network is composed of 36 convolutional layers organised into 14 distinct modules, designed to progressively extract increasingly abstract and discriminative features from input images. The architecture begins with convolutional and pooling layers that extract low-level features (e.g., edges, corners, and texture patterns), followed by deeper layers that capture higher-level semantic structures relevant to classification tasks. The hallmark of Xception lies in its use of depthwise separable convolutions, a major innovation over conventional convolution operations. In standard convolution, spatial and channel-wise filtering are performed simultaneously-each kernel operates across both spatial dimensions (height and width) and all input channels at once. This approach, though effective, is computationally expensive and parameter-intensive. Depthwise separable convolution addresses this limitation through a two-step factorisation:

- **Depthwise Convolution:** Applies a single convolutional filter to each input channel independently, focusing solely on capturing spatial information within that channel.
- **Pointwise Convolution (1×1 convolution):** Combines the outputs from the depthwise convolution across all channels to integrate inter-channel correlations and generate the final feature representation.

This separation of spatial and cross-channel operations significantly reduces the number of parameters and computational cost while maintaining, or even improving, representational quality. Consequently, Xception achieves a remarkable balance between model depth, efficiency, and accuracy, making it particularly well-suited for high-resolution medical imaging tasks like MRI analysis, where large image sizes can otherwise lead to excessive computation. Training deep CNNs from scratch on limited medical datasets often leads to overfitting due to insufficient data volume. To mitigate this, the research employs a transfer learning strategy using a pre-trained Xception model with weights initialised from the ImageNet dataset, which contains over 1.2 million natural images across 1,000 object categories. The convolutional base of the pre-trained Xception model serves as a universal feature extractor, leveraging its learned ability to detect generic visual patterns such as edges, gradients, shapes, and texture features that are broadly transferable to medical imaging domains. During the initial phase of training, the base layers are frozen, preventing their weights from being updated. This preserves the integrity of the pre-learned general visual representations while ensuring that the network's foundational feature detection remains stable. This approach allows the model to focus its learning on task-specific classification layers rather than relearning low-level features from limited medical data. The result is a model that benefits from both general visual knowledge and domain-specific adaptation, achieving faster convergence and improved generalisation performance on the MRI dataset. After establishing the pre-trained base as a fixed feature extractor, the next step is to customise and fine-tune the architecture for the brain tumour classification task. The original ImageNet classification head (which predicts 1,000 classes) is removed and replaced with a task-specific classification head tailored to four diagnostic categories: Glioma, Meningioma, Pituitary tumour, and No Tumour. The modified classification head consists of:

- A Global Average Pooling 2D (GAP) layer, which compresses the spatial dimensions of feature maps into a single feature vector per channel. This not only reduces the number of parameters but also mitigates overfitting by preserving the most salient global spatial information.
- A Dense (fully connected) layer with ReLU (Rectified Linear Unit) activation, introducing non-linearity and enabling the network to learn complex combinations of extracted features.
- A final Dense layer with SoftMax activation, which outputs a probability distribution over the four classes, indicating the model's confidence level for each possible tumour type.

Following this architectural adaptation, fine-tuning is performed by unfreezing the later layers of the Xception base while keeping the earlier layers frozen. This selective training allows the deeper convolutional layers, responsible for more abstract feature representations, to adjust their weights specifically to the texture, contrast, and intensity variations characteristic of MRI scans. The optimisation process employs the Adam optimiser, chosen for its adaptive learning rate and efficient convergence, in conjunction with categorical cross-entropy loss to measure the divergence between the predicted and true class distributions. Additional regularisation techniques, including dropout and early stopping, are implemented to prevent overfitting and ensure stable generalisation to unseen test data. Through this carefully designed pipeline, the fine-tuned Xception model effectively learns to distinguish subtle differences in tumour morphology, boundary definition, and tissue composition, enabling accurate and robust classification of brain tumour types from MRI scans.

3.4. Grad-CAM for Explainability

To ensure interpretability and trust in the model's decisions, Gradient-weighted Class Activation Mapping (Grad-CAM) was utilised. Grad-CAM helps visualise which regions of the input MRI contributed most to the model's classification decision. The following steps outline the process:

- **Identification of Final Convolutional Layer:** The layer block14_sepconv2_act is selected as the final convolutional layer in the model. This layer is chosen because it contains the richest high-level spatial feature maps, capturing complex patterns and abstract representations of brain structures. These deep features are ideal for visualising decision-critical regions, as they balance spatial detail with semantic abstraction.
- **Gradient Model Creation and Manipulation:** A new Keras model is constructed to facilitate the extraction of both feature maps and class predictions. This auxiliary model takes the same image input as the original CNN, but outputs two sets of data.
- **Gradient Computation and Normalisation:** Using TensorFlow's tf.GradientTape, gradients are computed for the predicted class score with respect to the feature maps of the final convolutional layer. These gradients indicate how sensitive the prediction is to changes in each spatial location of the feature maps. In other words, they quantify how much each neuron's activation influences the model's confidence in a particular class (e.g., "tumour present" vs. "normal").
- **Weight Calculation (Importance Mapping):** The computed gradients are globally averaged across the spatial dimensions (height and width) to obtain a single scalar weight for each feature map. These weights represent the importance or contribution of each feature map toward the final decision. Higher weights indicate a stronger influence on the model's classification.
- **Heatmap Generation and Computation:** The final feature maps are then weighted by their corresponding importance scores and summed to form a combined activation map. A ReLU (Rectified Linear Unit) activation is applied to this result to retain only positive influences-i.e., regions that positively support the predicted class. The resulting heatmap is normalised to (0, 1) to ensure uniform intensity scaling for visualisation.
- **Visualisation and Overlay Visualisation:** The normalised heatmap is resized to match the original MRI's dimensions. A colour-mapping scheme such as JET or VIRIDIS is applied to enhance visual interpretability, highlighting high-importance regions in warm colours (red/yellow) and less significant regions in cooler tones (blue/green). Finally, the colourised heatmap is superimposed on the original MRI scan with a degree of transparency, yielding a clear, intuitive visualisation of the model's focus areas.

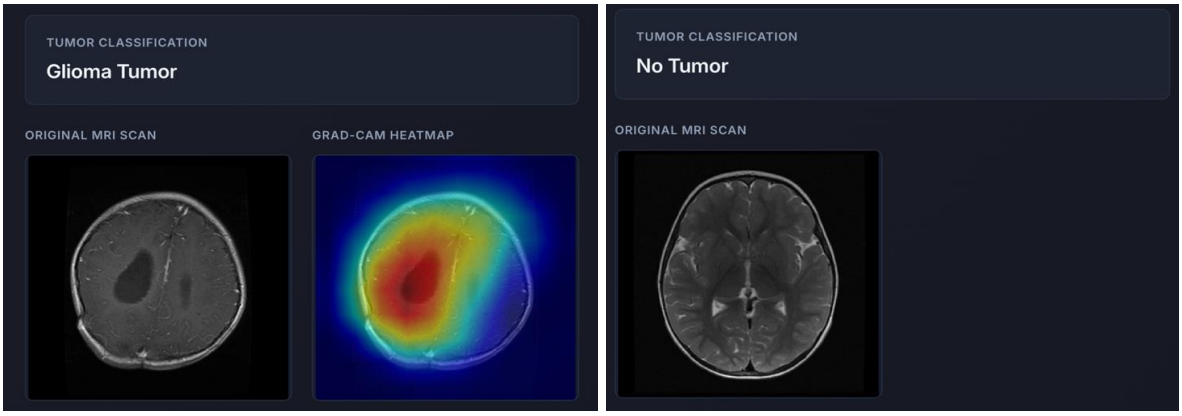
This process enables clinicians and researchers to understand why the system makes certain predictions. For example, if a tumour region is correctly highlighted on the overlay, it validates the model's reasoning. Conversely, unexpectedly highlighted regions can help identify model biases or data inconsistencies. Overall, this explainability framework not only enhances trust and transparency but also strengthens the system's suitability for clinical decision support. By combining high-performance CNN architectures with interpretable Grad-CAM visualisations, the proposed model becomes both accurate and accountable-a crucial factor for deployment in real-world medical environments.

4. Results and Discussions

4.1. Input and Output Specifications

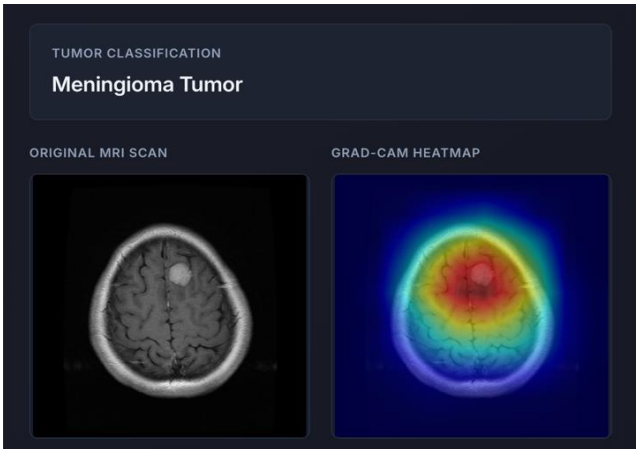
4.1.1. Input Specifications

The system accepts brain tumour MRI images in JPEG/PNG formats through a web interface, requiring no technical expertise. Training Environment: Platform: Google Colab with GPU acceleration - Framework: TensorFlow 2.10 with Keras API - Configuration: 50 epochs, Adam optimiser (lr=0.0001), batch size=32.



a) Sample output for glioma tumour case

b) output for no tumour case



c) Sample output for meningioma tumour case

Figure 3: Represents the GRAD-CAM visualisation of various brain tumour cases

Deployment Environment: - Backend: Flask 2.2 REST API - Frontend: React with Tailwind CSS - Model: Pre-trained Xception loaded at startup. Output Specifications: The system provides classification into four categories (Glioma, Meningioma, No Tumour, Pituitary) with confidence scores, and Grad-CAM heatmaps highlighting decision regions, as shown in Figure 3. Grad-CAM visualisations correctly localised tumour regions in the majority of cases [4]; [5].

4.2. Efficiency of Proposed System

- **Computational Efficiency:** The system achieves 3-5 second prediction latency with model pre-loading and GPU acceleration support. This performance meets clinical workflow requirements for real-time applications.
- **Diagnostic Efficiency:** The system provides immediate MRI assessment, enabling rapid case prioritisation and serving as an automated second-opinion tool. Grad-CAM visualisations direct clinician attention to suspicious regions, improving diagnostic workflow efficiency.

4.3. Training Dynamics and Performance Evaluation

4.3.1. Training Performance

The Xception model achieved ~99% training accuracy and 93-95% validation accuracy over 50 epochs, demonstrating effective learning without overfitting. Training loss decreased to 0.02 while validation loss stabilised at 0.25-0.27, indicating robust convergence as shown in Figures 4 and 5.

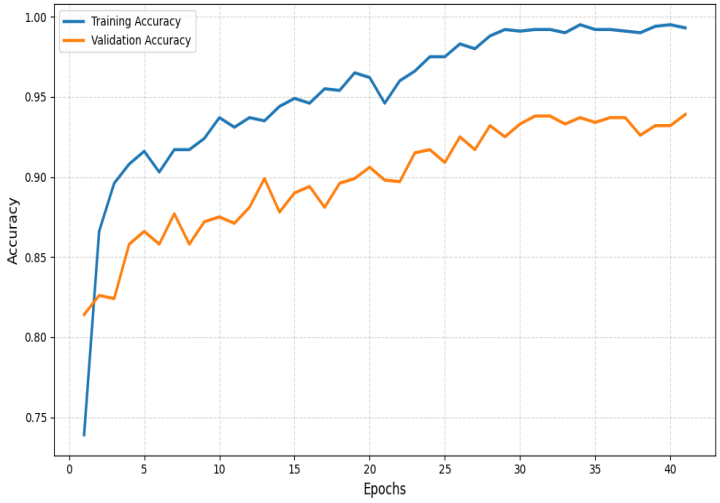


Figure 4: Training vs. validation accuracy

Figure 5 shows that the training and validation losses decrease over epochs, indicating that learning is progressing well. The training loss decreases consistently to a low value, whereas the validation loss decreases more slowly with small changes, suggesting the model will generalise well.

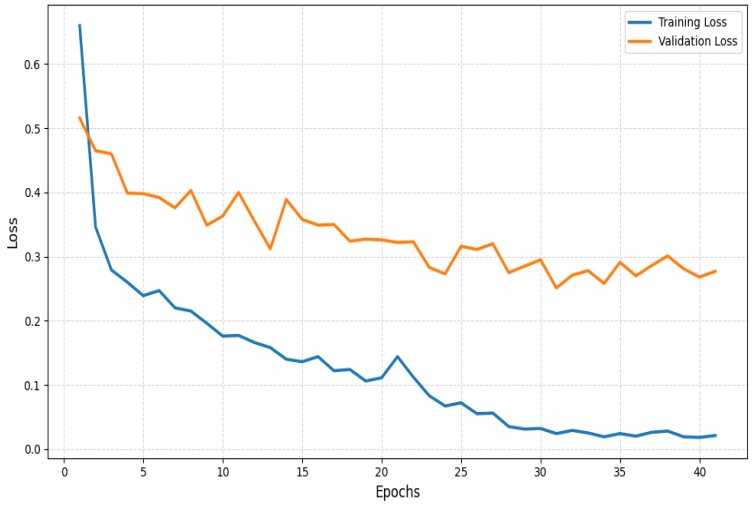


Figure 5: Training vs. validation loss

4.4. Test Set Performance

Overall Metrics: - Test Accuracy: 95.7% - Test Loss: 0.16-0.18 - Dataset: 7,023 MRI images across four classes Class-wise Performance: The model achieved balanced performance with F1-scores ranging from 0.93-0.98 across all classes, demonstrating robust classification without bias toward specific tumour types. High Diagnostic Accuracy: Achieved 95.7% accuracy with strong precision, recall, and F1-scores between 0.93–0.98, ensuring consistent classification reliability across all tumour categories as shown in Table 2.

Table 2: Class-wise performance metrics

Class	Precicion	Recall	F1-Score	Support
Glioma	0.9656	0.9367	0.9509	300
Meningioma	0.9253	0.9314	0.9283	306
No Tumor	0.9756	0.9852	0.9803	405
Pituitary	0.9736	0.9833	0.9784	300
Accuracy			0.9611	1311
Macro Average	0.9600	0.9591	0.9595	1311
Weighted Average	0.9611	0.9611	0.9610	1311

- **Explainable AI Integration:** Incorporated Grad-CAM visualisation for transparent decision-making, enabling clinicians to validate AI reasoning and model focus areas.
- **Real-Time Prediction:** Delivered 3–5 second inference time, supporting rapid clinical decision-making and patient triage.
- **Full-Stack Deployment:** Implemented as a scalable web application, ensuring accessibility for hospitals, research institutions, and educational use without dependency on local computational resources.
- **Confusion Matrix:** This confusion matrix evaluates a brain tumour classification model across four classes: glioma, meningioma, no tumour, and pituitary tumour (total ~1,311 samples).

It shows strong performance overall, with high true positives on the diagonal-e.g., 399/405 no-tumour cases correct and 295/300 pituitary cases correct-but minor misclassifications, mainly between glioma (281/300 correct) and meningioma (285/306 correct), as shown in Figure 6.

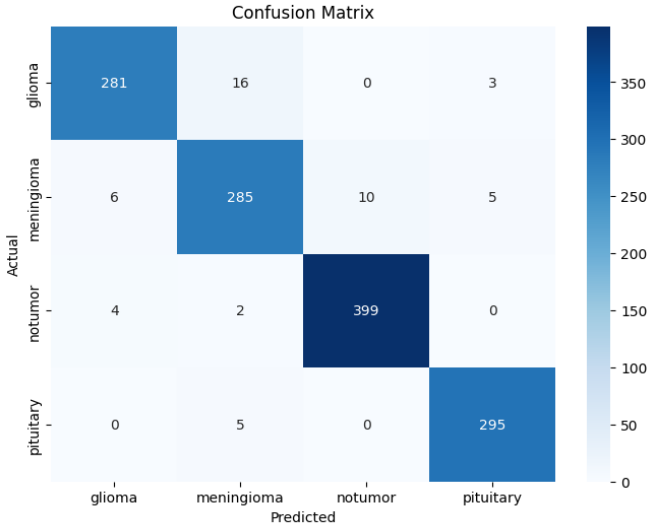


Figure 6: Confusion matrix

4.5. Qualitative Analysis

Grad-CAM visualisations confirmed that the model's attention patterns align with clinically relevant pathological features rather than artefacts. Tumour cases showed focused heatmaps on pathological regions, while "No Tumour" cases displayed diffuse, low-intensity patterns, validating the model's clinical interpretability.

4.6. Comparison with Existing Systems

The system's key differentiator is integrated Grad-CAM visualisation alongside classification, providing the interpretability essential for clinical adoption. Unlike "black box" models, this system enables clinicians to verify that the model focuses on clinically relevant features, identify potential errors, and build trust through transparency [3]; [4]; [5].

5. Conclusion

This research successfully designed and implemented a deep learning–based system for automated brain tumour classification from MRI scans, demonstrating high diagnostic performance and clinical interpretability. Leveraging the Xception architecture with transfer learning, the system achieved an impressive overall accuracy of 95.7%, supported by strong F1-scores ranging from 0.93 to 0.98 across all four diagnostic categories-Glioma, Meningioma, Pituitary tumour, and No Tumour. These results affirm the model's robustness, reliability, and balanced generalisation across diverse tumour types, establishing it as a viable tool for real-world clinical settings. A central innovation of this research is the integration of Gradient-weighted Class Activation Mapping (Grad-CAM), which transforms the conventional "black box" nature of CNNs into a transparent, interpretable decision-making framework. Through colour-coded heatmap overlays, the model provides visual explanations of its predictions, pinpointing the precise regions within MRI scans that influenced classification outcomes. This interpretability not only enhances clinician trust but also offers an educational advantage, helping medical practitioners and students understand key visual biomarkers associated with different tumour types. In terms of computational performance, the proposed system achieves real-time inference with an average processing time of 3–5 seconds per prediction, making it suitable for deployment in clinical and point-of-care environments where speed and accuracy are critical. To ensure maximum usability, the deep learning model was encapsulated within a full-stack web application featuring an intuitive, responsive interface. Clinicians can upload MRI scans, view classification results, and inspect corresponding Grad-CAM visualisations seamlessly in any standard web browser, without specialised software or hardware. This deployment strategy ensures broad accessibility and integration into existing radiology workflows.

5.1. Key Achievements

5.1.1. Clinical Implications

The system's interpretability and speed position it as a powerful clinical decision-support tool. It can serve as a second-opinion assistant for radiologists, reducing cognitive workload and minimising diagnostic oversight. The ability to quickly and reliably

identify potential tumour regions enables early screening, rapid case prioritisation, and workflow optimisation in busy medical settings. Furthermore, its visual explanations can significantly enhance medical education and training by offering insights into how AI interprets radiological features and by bridging the gap between computational reasoning and human clinical expertise. In low-resource settings, where access to specialised neuroradiologists may be limited, the system can assist general practitioners and technicians with initial diagnostic triage, ensuring timely patient referrals and improved treatment outcomes. Hence, the model contributes to equitable access to healthcare, aligning with broader goals of AI democratisation in medicine.

5.1.2. Future Work

While the current system demonstrates strong classification capabilities, several avenues for future enhancement can be pursued to extend its clinical utility and technical sophistication:

- **Semantic Segmentation:** Transitioning from classification to pixel-level segmentation for precise tumour boundary delineation and volumetric analysis, aiding surgical planning and treatment monitoring.
- **Federated Learning:** Implementing privacy-preserving federated learning across multiple institutions to improve model generalisation without sharing sensitive patient data.
- **Multi-Modal Integration:** Incorporating additional MRI modalities such as T1, T2, and FLAIR sequences, as well as diffusion and perfusion imaging, to capture complementary diagnostic information.
- **Advanced Architectures:** Exploring Vision Transformers (ViTs), ensemble CNN approaches, and hybrid architectures to further enhance classification accuracy and robustness.
- **Longitudinal Patient Monitoring:** Expanding the framework for temporal analysis, enabling monitoring of tumour evolution, treatment response, and post-operative recurrence.
- **Clinical Validation and Regulatory Compliance:** Conducting extensive multi-centre clinical trials, peer-reviewed validation, and navigating regulatory approval processes (such as FDA or CE marking) to transition from research prototype to certified medical device.

Acknowledgement: The authors gratefully acknowledge the academic support and research facilities provided by SRM Institute of Science and Technology, S. A. Engineering College, Madras Institute of Technology, Dhaanish Ahmed College of Engineering, and Coventry University.

Data Availability Statement: All data supporting the findings of this study are maintained by the authors and can be shared with interested researchers upon reasonable request to promote openness and reproducibility.

Funding Statement: The authors jointly affirm that this study was conducted independently and did not receive financial support from any public, private, or commercial funding agencies.

Conflicts of Interest Statement: The authors declare that there are no personal, professional, or financial relationships among them that could be perceived as influencing the research outcomes.

Ethics and Consent Statement: All contributing authors have actively participated in this work and have provided full consent for its publication and unrestricted academic dissemination.

References

1. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in *Proceedings of the 26th International Conference on Neural Information Processing Systems*, Lake Tahoe, United States of America, 2012.
2. D. C. Cireșan, A. Giusti, L. M. Gambardella, and J. Schmidhuber, "Mitosis Detection in Breast Cancer Histology Images with Deep Neural Networks," in *proc. Medical Image Computing and Computer-Assisted Intervention*, Nagoya, Japan, 2013.
3. F. Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions," in *proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, Hawaii, United States of America, 2017.
4. R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization," in *proc. IEEE International Conference on Computer Vision (ICCV)*, Venice, Italy, 2017.
5. J. Qu, X. Zhao, P. Chen, Z. Wang, Z. Liu, B. Yang, and H. Li, "Deep learning on digital mammography for expert-level diagnosis accuracy in breast cancer detection," *Multimedia Systems*, vol. 28, no. 6, pp. 1263-1274, 2022.

6. M. J. Adamu, H. B. Kawuwa, L. Qiang, C. O. Nyatega, A. Younis, M. Fahad, and S. S. Dauya, “Efficient and Accurate Brain Tumor Classification Using Hybrid MobileNetV2–Support Vector Machine for Magnetic Resonance Imaging Diagnostics in Neoplasms,” *Brain Sci.*, vol. 14, no. 12, p. 1178, 2024.
7. M. Havaei, A. Davy, D. Warde-Farley, A. Biard, A. Courville, Y. Bengio, C. Pal, P. Jodoin, and H. Larochelle, “Brain Tumor Segmentation with Deep Neural Networks,” *Medical Image Analysis*, vol. 35, no. 1, pp. 18-31, 2017.
8. S. Pereira, A. Pinto, V. Alves, and C. A. Silva, “Brain Tumor Segmentation Using Convolutional Neural Networks in MRI Images,” *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1240-1251, 2016.
9. K. Kamnitsas, C. Ledig, V. F. J. Newcombe, J. P. Simpson, A. D. Kane, D. K. Menon, D. Rueckert, and B. Glocker, “Efficient Multi-Scale 3D CNN with Fully Connected CRF for Accurate Brain Lesion Segmentation,” *Medical Image Analysis*, vol. 36, no. 2, pp. 61-78, 2017.
10. P. Afshar, A. Mohammadi, and K. N. Plataniotis, “Brain Tumor Type Classification via Capsule Networks,” in *Proc. 25th IEEE International Conference on Image Processing (ICIP)*, Athens, Greece, 2018.
11. J. Amin, M. Sharif, M. Yasmin, and S. L. Fernandes, “A Distinctive Approach in Brain Tumor Detection and Classification Using MRI,” *Pattern Recognition Letters*, vol. 139, no. 11, pp. 118-127, 2020.
12. N. Abiwinanda, M. Hanif, S. T. Hesaputra, A. Handayani, and T. R. Mengko, “Brain Tumor Classification Using Convolutional Neural Network,” in *proc. World Congress on Medical Physics and Biomedical Engineering*, Prague, Czech Republic, 2018.
13. S. Deepak and P. M. Ameer, “Brain Tumor Classification Using Deep CNN Features via Transfer Learning,” *Computers in Biology and Medicine*, vol. 111, no. 8, pp. 103345, 2019.
14. Z. N. K. Swati, Q. Zhao, M. Kabir, F. Ali, Z. Ali, S. Ahmed, and J. Lu, “Brain Tumor Classification for MR Images Using Transfer Learning and Fine-Tuning,” *Computerized Medical Imaging and Graphics*, vol. 75, no. 7, pp. 34-46, 2019.
15. M. Sajjad, S. Khan, K. Muhammad, W. Wu, A. Ullah, and S. W. Baik, “Multi-Grade Brain Tumor Classification Using Deep CNN with Extensive Data Augmentation,” *Journal of Computational Science*, vol. 30, no. 1, pp. 174-182, 2019.
16. G. Garg and R. Garg, “Brain Tumor Detection and Classification Based on Hybrid Ensemble Classifier,” *arXiv Preprint*, 2021. Available: https://arxiv.org/abs/2101.00216 [Accessed by 22/10/2024].
17. K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” in *proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, Nevada, United States of America, 2016.
18. K. Simonyan and A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition,” *arXiv Preprint*, 2015. Available: https://arxiv.org/abs/1409.1556 [Accessed by 22/10/2024].