



CerebralNet meets Explainable AI: Brain tumor detection and classification with probabilistic augmentation and a deep learning approach

Arth Agrawal , Jyotismita Chaki ^{*}

School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, India



ARTICLE INFO

Keywords:

Brain Tumor
Deep Learning
Atrous Convolution and Spatial Pyramid Pooling
MobileNetV2
Local Interpretable Model-agnostic Explanations
Image Augmentation

ABSTRACT

Brain tumors pose a significant and often life-threatening challenge to human health. Accurate and timely brain tumor detection is crucial for early intervention, improved treatment outcomes, and ultimately, enhanced patient survival and quality of life. Automation is crucial for overcoming the limitations of manual analysis, enabling faster, more objective, and potentially more accurate diagnoses of brain tumors. This study introduces the CerebralNet architecture, a new approach for brain tumor detection and classification that uses a pre-trained MobileNetV2 backbone. Key innovations include the integration of Atrous Spatial Pyramid Pooling and Atrous Convolution blocks, enhancing feature extraction, and capturing multi-scale contextual information which is crucial for accurate brain tumor detection and classification because brain tumors exhibit significant variability in size, shape, and appearance. Furthermore, a new probabilistic image augmentation selection strategy is employed, incorporating 10 augmentation techniques such as Gaussian noise, Gaussian blur, random rotations, and intensity and color variations to simulate real-world imaging artifacts. This approach, combined with the augmented Brain MRI Dataset, significantly improves model robustness and generalizability. Rigorous evaluation on both the imbalanced original (BM) and augmented (ABM) datasets demonstrates exceptional performance, exceeding 91 % accuracy on the BM dataset and achieving 96 % on the ABM dataset. This study also incorporates LIME (Local Interpretable Model-agnostic Explanations) for model interpretability, providing valuable insights into the model's decision-making process. These findings demonstrate the potential of the new augmentation strategy and Atrous MobileNetV2 architecture to significantly improve the accuracy and reliability of brain tumor detection and classification, paving the way for improved clinical outcomes.

1. Introduction

Brain tumors, characterized by their diverse nature and significant malignancy rate, pose a serious threat to human health. This insidious disease, characterized by an alarmingly high incidence rate of approximately 7 per 100,000 individuals annually, casts a long shadow on public health. With over 100 distinct types and nearly one-third classified as malignant, brain tumors present a complex and multifaceted diagnostic and therapeutic landscape [1].

The human brain, a marvel of complexity housing with billions of neurons, serves as the central command center of the entire nervous system, orchestrating every aspect of human existence. However, the emergence of brain tumors, abnormal cell growths within or surrounding this vital organ, can have devastating consequences. These growths disrupt the delicate balance of neural activity, presenting a considerable health risk to cognitive function, motor skills, and overall well-being.

These growths can be classified as primary, originating within the brain itself or its surrounding tissues, or secondary, arising from the spread of cancer from other parts of the body. Brain tumors encompass a diverse spectrum of conditions, each with unique characteristics and implications [2]. For example, Astrocytomas, arising from star-shaped glial cells, exhibit varying degrees of malignancy. Glioblastoma, a highly aggressive form of astrocytoma, is particularly challenging to treat. Ependymomas originate from cells lining the spinal cord and brain ventricles. Gangliogliomas are rare tumors composed of both nerve cells and glial cells. Germ cell tumors, such as germinomas, primarily affect the pineal gland and the brain's midline structures. Meningiomas, arising from the protective layers enveloping the brain, are typically slow-growing and benign. Neurocytomas are rare tumors that originate from nerve cells. Oligodendrogiomas are slow-growing tumors that arise from oligodendrocytes, a type of glial cell. Papillomas are benign, wart-like growths that can occur in various parts of the body, including

* Corresponding author.

E-mail address: jyotismita.c@gmail.com (J. Chaki).

the brain. Schwannomas are benign tumors that arise from Schwann cells, which produce the myelin sheath that insulates nerve fibers. Tuberculomas are not true tumors but rather granulomatous lesions caused by tuberculosis infection. This diverse range of brain tumor types underscores the complexity of this disease and the importance of accurate diagnosis and tailored treatment approaches. The World Health Organization (WHO) employs a grading system to categorize brain tumors based on their growth characteristics and severity, providing a crucial framework for guiding clinical management. Grade I and II tumors are considered low-grade, exhibiting generally slower growth rates and often demonstrating a more favorable prognosis. Conversely, Grade III and IV tumors are classified as high-grade, characterized by more aggressive growth patterns and a poorer prognosis [3].

Despite advancements in treatment, the 5-year survival rate of this disease for adults remains a sobering 36 %, underscoring the critical need for improved diagnostic and therapeutic strategies. [4]. Accurate and timely diagnosis is paramount for successful treatment outcomes, as treatment choices, including radiation, chemotherapy, and surgery, are intricately linked to the tumor's type, size, and location. Early diagnosis through advanced imaging techniques is crucial for optimal treatment outcomes, as timely intervention significantly improves the chances of successful management and minimizes the potential for severe complications. While neuroimaging techniques like Positron Emission Tomography (PET) and Computed Tomography (CT) offer valuable insights into the internal physiological state, Magnetic Resonance Imaging (MRI) emerges as the gold standard because of its non-penetrating nature and exceptional capacity for generating detailed brain images [5]. However, manual analysis of MRI scans is a time-intensive and inherently subjective procedure, likely to human error. To address these limitations and improve diagnostic efficiency, the development of automated methods for brain tumor detection and classification has become a crucial area of research [6].

Machine Learning has revolutionized numerous fields, and its impact on medical image analysis, particularly brain tumor detection and classification, has been profound. Traditional approaches to brain tumor detection and classification significantly dependent on handcrafted features, requiring substantial human knowledge and often limiting their adaptability to the complex and diverse brain tumor natures [7]. However, the advent of deep learning has ushered in a new era of possibilities. Unlike their predecessors, deep learning algorithms possess the remarkable ability to automatically learn discriminative features directly from the rich information embedded within MRI data. This paradigm shift empowers these models to characterize tumors with greater accuracy and robustness, surpassing the limitations of traditional methods. While deep learning offers immense potential, training these models from scratch can be computationally demanding, requiring significant time and resources [8]. This necessitates careful consideration of computational efficiency and the availability of substantial datasets for optimal model performance. By employing transfer learning, we can significantly enhance the performance of brain tumor detection models [9]. Pre-trained models provide a strong foundation of learned features, facilitating the acquisition of task-specific knowledge and mitigating the risk of overfitting, particularly when dealing with limited medical image datasets.

Recognizing the challenges associated with brain tumor detection and classification from MR images, this study proposes an automated three-fold approach to address these complexities. The primary motivation behind this proposed approach stemmed from the desire to enhance the accuracy and robustness of brain tumor detection and classification. To achieve this, we integrated Atrous Spatial Pyramid Pooling (ASPP) [10] and Atrous Convolution blocks into the transfer learning architecture and created a new architecture i.e., CerebralNet. This proposed approach uses the strengths of the transfer learning model while incorporating ASPP to effectively capture multi-scale contextual information within the brain tumor images. By incorporating Atrous convolutions, we effectively enlarge the receptive fields of the network

without increasing the number of parameters, enabling the model to capture finer details and broader contextual information within the tumor region. Furthermore, the inclusion of 10 distinct augmentations, selected using a new probabilistic augmentation selection strategy, significantly enhances the model's generalization ability and robustness to variations in image quality, thereby improving its overall performance and clinical applicability. The CerebralNet model is assessed on two datasets, including 15 different classes: 14 distinct brain tumor classes (Tuberculoma, Astrocytoma, Schwannoma, Carcinoma, Papilloma, Ependymoma, Oligodendrogioma, Ganglioglioma, Neurocytoma, Germinoma, Meningioma, Granuloma, Glioblastoma, and Medulloblastoma) and a healthy (no tumor) class. A comprehensive evaluation framework is employed, encompassing a suite of evaluation measures including precision, accuracy, F1-score, recall, and Matthews Correlation Coefficient. Along with this, in this study, we employ LIME (Local Interpretable Model-agnostic Explanations), an Explainable AI (XAI) technique, to enhance model interpretability. Importantly, LIME can reveal subtle patterns and relationships within the MRI data that may not be readily apparent to human experts, potentially leading to new understandings of the distinctive features of various tumor subtypes and their impact on the model's predictions. By integrating the power of extensive data augmentation, a transfer learning architecture with advanced techniques like ASPP, and an XAI technique aims to achieve an efficient brain tumor detection and classification model.

The key contributions of this study lie in a 3-fold method that significantly boosts the effectiveness of brain tumor detection and classification.

A. Enrichment of the Dataset with a new Probabilistic Augmentation Technique (First Fold): This study employs an automated and new approach for selecting augmentation techniques to effectively introduce realistic variations in brain MRI images. A diverse pool of augmentation techniques, including Contrast Limited Adaptive Histogram Equalization (CLAHE), Defocus, Downscale, Gaussian Blur, Horizontal Flip, Hue Saturation Value adjustments, Random Gamma, Transpose, Zoom Blur, and Pixel Dropout, is considered. These techniques simulate common imaging artifacts and variations encountered in clinical settings, such as noise, blurring, intensity, and color variations. A new algorithm is proposed to randomly select a single or subset of these techniques based on a probability distribution. This probabilistic approach explores a wide range of augmentation combinations, leading to a more diverse and representative set of augmented images. By dynamically selecting augmentation techniques for each image of the original dataset (BM), this approach introduces greater variability and realism into the augmented dataset (ABM), ultimately enhancing the model's robustness and generalization capabilities.

B. Proposed CerebralNet for Brain Tumor Detection and Classification (Second Fold): This study introduces a new approach to brain tumor detection and classification by using and extending the strengths of the MobileNetV2 architecture. Recognizing the critical need for accurate and efficient models, we propose a refined architecture termed "CerebralNet" for improved brain tumor classification. This innovation builds upon the efficient MobileNetV2 backbone, incorporating key architectural enhancements to enhance feature extraction and capture multi-scale contextual information within brain tumor images. Specifically, we integrate Atrous Spatial Pyramid Pooling (ASPP) and Atrous Convolution blocks into the MobileNetV2 framework. ASPP, by employing atrous convolutions with varying dilation rates, effectively captures multi-scale contextual information, enabling the model to discern objects of varying sizes and accurately localize tumor boundaries. Furthermore, the integration of Atrous convolutions within the MobileNetV2 architecture significantly expands the receptive fields of the network without increasing the number of parameters, allowing the model to capture broader contextual information within the tumor region

[11]. This refined architecture aims to significantly improve the accuracy and robustness of brain tumor classification while maintaining computational efficiency, making it a promising approach for clinical applications.

C. Understanding Model Decisions with LIME-based Explainability (Third Fold): To improve the transparency and understandability of our CerebralNet model for brain tumor detection and classification, we integrated LIME, an XAI technique. By employing LIME, we can better comprehend the factors inducing CerebralNet's predictions. Specifically, LIME enables us to [12]: (1) understand and identify the factors and the most crucial image features, such as tumor margins, texture patterns, and intensity variations, that significantly influence CerebralNet's detection and classification accuracy, and (2) build trust in the model's predictions by providing human-understandable explanations for its classifications. This increased transparency not only improves the model's credibility but also facilitates the identification of potential biases or limitations, ultimately leading to a more robust and reliable brain tumor detection and classification system.

This article is organized as follows: [Section 2](#) delves into existing research on automated brain tumor detection and classification, exploring various approaches and their associated limitations. At the core of this research, [Section 3](#) presents a new approach that utilizes the probabilistic augmentation selection strategy, Atrous convolution, XAI, and performance measures for accurate and robust brain tumor detection and classification. Subsequently, [Section 4](#) presents the findings obtained from the application of the proposed approach, such as detailed performance analysis, comparisons with existing methods, and insightful visualizations. [Section 5](#) discusses the significance of these results within the broader context of brain tumor detection and classification and their implications for clinical practice. Finally, [Section 6](#) concludes by summarizing the key findings and contributions of this research, highlighting the potential of the proposed approach for improving the accuracy and reliability of brain tumor detection and classification, and outlining promising avenues for future research and development in this field.

2. Related works

Research in brain tumor detection and classification has witnessed a surge in the application of deep learning techniques [13–15]. Convolutional Neural Networks (CNNs) have emerged as a dominant force, with researchers actively investigating various architectures and using strategies such as transfer learning, data augmentation, and ensemble methods to improve model robustness and achieve higher levels of accuracy.

The HDL2BT (Hierarchical Deep Learning-Based Brain Tumor) system in [16] employs a CNN-based approach for brain tumor classification. Evaluated on a dataset of 3264 MRI images with 3 tumor types, it achieved 92.13 % precision and a 7.87 % miss rate. CNNs within HDL2BT excel at automatically extracting intricate features from MRI images, surpassing human radiologists in identifying subtle patterns. The study in [17] proposes two deep learning models for brain tumor classification: a 23-layer CNN trained on a large dataset (3064 images) and a hybrid model combining VGG16 with a mirrored 23-layer CNN. The hybrid model achieved 100 % accuracy on a smaller dataset (152 images). The 23-layer CNN excels at learning intricate features from MRI images, while the hybrid model, despite achieving high accuracy, presents increased computational complexity due to the integration of VGG16. The study in [18] presents a novel hybrid deep learning model for brain tumor classification, combining CNNs with LSTM (Long Short-Term Memory). Evaluated on the Kaggle dataset (3264 MRI images with 3 tumor types), the model achieved an accuracy of 92 %. This hybrid approach effectively uses the strengths of both CNNs and LSTMs, enabling the extraction of both spatial and temporal features for

improved classification performance. The study in [19] proposes a CNN architecture for brain tumor identification. Starting with an initial convolutional layer (16 filters, 3x3 kernel), the model employs successive convolutional and max-pooling layers to extract increasingly complex features. Dropout (0.5) is applied to prevent overfitting. Evaluated on 3264 MR images with 3 tumor types, the model achieves 93.3 % accuracy, 98.43 % AUC, and 91.19 % recall. While effective, the fixed architecture may not be optimal for all tumor types or imaging modalities. The study in [20] proposes LeU-Net (Less Layered and less complex U-Net), a simplified U-Net architecture, for brain tumor detection. Evaluated on a dataset of 253 MR (155 tumor and 98 no tumor) images, LeU-Net achieved 98 % accuracy on cropped images and 94 % on uncropped images. LeU-Net's simplified design reduces computational complexity, making it suitable for resource-constrained environments. However, the model's sensitivity to background information, as evidenced by the performance difference between cropped and uncropped images, highlights the need for robust preprocessing in clinical applications. The study in [21] presents a brain tumor detection and classification system by integrating a fusion-based contrast enhancement technique, saliency map-based segmentation, and deep feature optimization with EfficientNetB0. The system achieves accuracies of 95.14 %, 94.89 %, and 95.94 % on BraTS2018, BraTS2019, and BraTS2020 datasets, respectively. While using EfficientNetB0 for powerful feature extraction, the accuracy of saliency maps can be influenced by factors such as image contrast, noise levels, and the presence of artifacts, which can lead to unreliable segmentation results. These inaccuracies can significantly impact the subsequent steps of feature extraction and classification. The order in which features are fused can significantly impact the outcome. An inappropriate fusion order can lead to a sub-optimal combination and potentially degrade the overall performance. Serial fusion may lead to the loss of valuable information from earlier stages, as subsequent fusion steps may not fully utilize all the available information. The study in [22] proposes an EDN-SVM (Ensemble Deep Neural-Support Vector Machine) framework for brain tumor detection. This framework integrates image preprocessing (Adaptive Contrast Enhancement Algorithm, median filtering), Fuzzy c-means clustering, GLCM (Gray Level Co-occurrence Matrix) feature extraction, and an ensemble of NNs (Neural Networks) and SVMs. Evaluated on 255T1-weighted (98 healthy and 155 tumor) MRI images, it achieved 97.93 % accuracy, 92 % sensitivity, and 98 % specificity. While EDN-SVM uses the strengths of both NNs and SVMs, the performance is sensitive to the accuracy of Fuzzy c-means segmentation, which may be limited for tumors with irregular shapes. GLCM primarily captures textural information within an image. While it can provide insights into the spatial distribution of pixel intensities, it may not effectively capture the intricate shapes, borders, and morphological characteristics of brain tumors. Tumors can exhibit diverse shapes (spherical, irregular, infiltrative) and margins (well-defined, indistinct), which are crucial for accurate diagnosis. Finding the optimal combination of hyperparameters for both the SVM and NN components can be a time-consuming and challenging process. The study in [23] evaluated four transfer learning architectures (ResNet152, VGG19, DenseNet169, MobileNetV3) on a Kaggle dataset of 7,023 MRI brain images with 3 tumor types. After data augmentation, MobileNetV3 achieved the highest accuracy (99.75 %). MobileNetV3's efficient architecture offers advantages in terms of computational cost and memory usage, making it suitable for resource-constrained deployments. The study in [24] proposes a deep learning-based approach for brain tumor classification using EfficientNetV2. Evaluated on a dataset comprising 1621 Glioma, 1645 Meningioma, 1757 Pituitary, and 2000 No-tumor images, the model achieved 99.16 % accuracy. Data augmentation techniques, including rotation, brightness adjustment, and flipping, were employed to enhance model robustness. EfficientNetV2's efficient architecture and strong feature extraction capabilities contributed to this high accuracy. The study in [25] proposes a CNN architecture inspired by GoogLeNet, incorporating an Attention-based Inception Module (ABIM) for brain

tumor classification. Evaluated on a Figshare dataset of 3064 MRI images with 3 tumor types, the model achieved an average accuracy of 97.97 % for glioma, 94.76 % for meningioma, and 99.22 % for pituitary tumors using five-fold cross-validation. The ABIM effectively captures multi-scale features, improving tumor characterization. The study in [26] proposes an ensemble model combining SCNN (Shallow CNN) and VGG16 for brain tumor classification. Trained on a dataset of 3064 MRI images with 3 tumor types, the ensemble achieved 97.77 % accuracy. By combining features from SCNN and VGG16, the model uses the strengths of both architectures: SCNN for local features and VGG16 for complex feature extraction. The study in [27] investigates five deep learning architectures (Xception, DenseNet201, DenseNet121, ResNet152V2, InceptionResNetV2) for brain tumor diagnosis. Xception achieved the highest accuracy (99.67 % for three-class i.e., glioma, meningioma, pituitary and 95.87 % for four-class i.e., glioma, meningioma, pituitary, and healthy due to its efficient depthwise separable convolutions. The study in [28] proposes a modified ResNet50 model for brain tumor detection. The original classification layer is replaced with four additional layers, including dropout and smoothing layers. Using the Adam optimizer, the model achieved 92 % accuracy on the Kaggle brain MRI dataset.

Some studies have incorporated XAI techniques to enhance the interpretability of their models, providing valuable insights into the decision-making process. For example, the study in [29] proposes a five-step approach for brain tumor classification, including image enhancement, custom CNN segmentation, feature extraction using a modified MobileNetV2, entropy-based feature selection, and M-SVM (Multiclass-SVM) classification. Evaluated on BraTS 2018 and Figshare datasets, the method achieved 97.47 % and 98.92 % accuracy, respectively. While MobileNetV2 offers computational efficiency, the overall proposed method is sensitive to image quality and may not accurately segment complex tumors. Additionally, entropy-based feature selection can be sensitive to parameter choices. Grad-CAM (Gradient-weighted Class Activation Mapping), used for interpretability, may produce coarse-grained heatmaps and struggle with multiple object localization. The study in [30] investigates brain tumor detection using POD (Proper Orthogonal Decomposition) with transfer learning models (MobileNetV2, Inception-v3, ResNet101, VGG-19) on the BraTS 2016 dataset. MobileNetV2 achieved the highest accuracy (99.21 %). While pre-trained models offer advantages, POD's performance can be sensitive to image quality variations. SHAP (Shapley Additive explanations), values, used for interpretability, may have approximation errors due to sampling-based methods. The study in [31] introduces ViT-CB, a novel architecture combining Vision Transformer (ViT), PCA, and CatBoost for brain tumor classification. Evaluated on two datasets, ViT-CB achieved high accuracy: 99.316 % on the first dataset (binary classification) and 90.004 % on the second (3 tumor and 1 healthy class classification). While ViT effectively captures long-range dependencies, PCA can be computationally expensive for large datasets, and SHAP values used for interpretability may have approximation errors due to sampling.

Research Gaps and Some Common Challenges of the Existing Models: One of the significant challenges of most of the above studies is the reliance on relatively small and potentially limited datasets. Many studies utilize datasets with a few number of images and limited variations in terms of tumor types, imaging modalities, and patient demographics. This can result in overfitting, resulting in poor generalization performance on unseen data. Consequently, the model's performance may degrade significantly when confronted with clinical data, which exhibits considerable variability in terms of image quality, acquisition parameters, and patient characteristics. Furthermore, the presence of noise or artifacts, which are ubiquitous in clinical MRI scans, can significantly impact the model's accuracy.

The second critical limitation of the above studies is the use of standard convolutional neural networks with a fixed receptive field. This strategy may struggle to capture the spatial context of brain tumors, which can vary significantly in size and shape. Brain tumors often

exhibit subtle visual cues, such as subtle intensity variations, subtle mass effects on surrounding structures, and subtle changes in tissue texture. Brain tumors exhibit significant variability in size and shape. Without the ability to capture information at multiple scales, the model may miss these subtle cues, leading to reduced sensitivity and potentially higher rates of false negatives. Also, without the ability to capture information at different scales, the model may not be able to effectively handle tumors of varying sizes and shapes, potentially leading to reduced performance. The lack of multi-scale context information can make the model more susceptible to variations in image quality, noise, and other artifacts, potentially leading to decreased robustness and generalization performance.

Another critical limitation in most of the previous studies is the inherent "black box" nature of deep learning models. While these models can achieve high accuracy, understanding the specific features and decision-making processes that underlie their predictions remains challenging. This lack of interpretability can hinder clinical trust and adoption. Radiologists and clinicians require insights into the model's reasoning to understand its strengths and limitations, identify potential biases, and ultimately integrate the model effectively into their clinical workflows. These limitations underscore the need for ongoing research to address the challenges of data variability, model generalization, and interpretability to enhance the clinical applicability and reliability of deep learning-based brain tumor detection and classification systems.

Solution Using the Developed Model: This study distinguishes itself by incorporating a comprehensive data augmentation strategy, employing 10 distinct augmentations selected by a new probabilistic augmentation selection algorithm. This approach significantly differentiates our work from previous studies, where data augmentation was often limited or less extensively explored. By introducing this level of diversity into the training data, we effectively expose the model to a wider range of image variations, including rotations, flips, brightness adjustments, and other transformations. This data augmentation strategy plays a crucial role in enhancing the model's generalization ability, mitigating overfitting, and improving its robustness to variations in image quality that may be faced in clinical practice.

This study introduces a novel approach to brain tumor classification by incorporating Atrous Convolution into the transfer learning model architecture. This constitutes a significant contribution as previous studies primarily relied on conventional convolutional neural networks. Atrous convolutions, unlike traditional convolutions, introduce gaps between filter parameters, effectively enlarging the receptive field without increasing the number of parameters. This enhanced receptive field allows the model to capture broader contextual information within the image, potentially improving feature extraction and classification accuracy. Atrous convolutions, particularly within the ASPP module, allow the model to effectively capture contextual information at multiple scales, which is crucial for distinguishing between subtle tumor boundaries and surrounding normal brain tissue.

Furthermore, the study utilizes LIME for model interpretability. Compared to SHAP and Grad-CAM, LIME offers several advantages. Firstly, LIME is more model-agnostic and can be applied to a wider range of machine learning models, not just deep neural networks. Secondly, LIME provides local explanations, focusing on the specific predictions made for individual instances, offering more granular insights into the model's decision-making process for each input image.

3. Proposed methodology for brain tumor detection and classification

The proposed approach for brain tumor detection and classification utilizes a multi-faceted strategy, encompassing a new probabilistic image augmentation strategy, a new model architecture, and XAI.

- Data Acquisition and a New Augmentation Selection Strategy: The foundation of this study is a base dataset of brain MRI images. This

study incorporates a new data augmentation strategy for brain MRI images by probabilistically selecting a subset of augmentation techniques from a diverse pool, including geometric transformations, noise injection, and intensity variations, improving the model's resilience to clinical image diversity.

- New CerebralNet Model Architecture: The base model of MobileNetV2 architecture is enhanced with ASPP and Atrous Convolutions to effectively capture multi-scale contextual information within brain tumor images, improving feature extraction and localization.
- Model Evaluation and Performance Analysis: CerebralNet was trained on both original and augmented datasets, exposing it to a wide range of image variations and improving its ability to generalize to unseen data. Hyperparameter tuning was performed to optimize model performance, considering factors such as the number of layers in the fully connected layers, the number of atrous convolution layers in the ASPP module, and their corresponding dilation rates. The model's performance was rigorously evaluated on two datasets. Key performance metrics such as Matthews Correlation Coefficient (MCC), precision, accuracy, recall and F1-score were used to assess CerebralNet's effectiveness.
- Model Interpretability: To gain insights into the model's decision-making process, LIME was employed. LIME provides localized explanations by approximating CerebralNet's predictions for a given

input sample. LIME visualizations helped to identify the most influential image regions that contributed to CerebralNet's predictions, shedding light on the model's decision-making process and highlighting the key features considered by the model for detection and classification.

The succeeding subsections provide a comprehensive description of the key steps involved in the proposed approach, with Fig. 1 graphically illustrates the entire workflow.

3.1. Dataset description

3.1.1. Public dataset

Among the available brain tumor datasets, the "Brain Tumor MRI Images 44 Classes" dataset stands out due to its comprehensive nature and public accessibility [32]. This dataset, a private collection of T1, contrast-enhanced T1, and T2 MR images, is meticulously curated by radiologists and made available for research purposes, ensuring anonymity by removing patient identifiers. In this study, we used the T1 MRI image category. The primary motivation for utilizing T1 MRI images in this study stems from their inherent ability to provide valuable anatomical information crucial for accurate brain tumor characterization [33]. T1 images exhibit excellent contrast between gray matter,

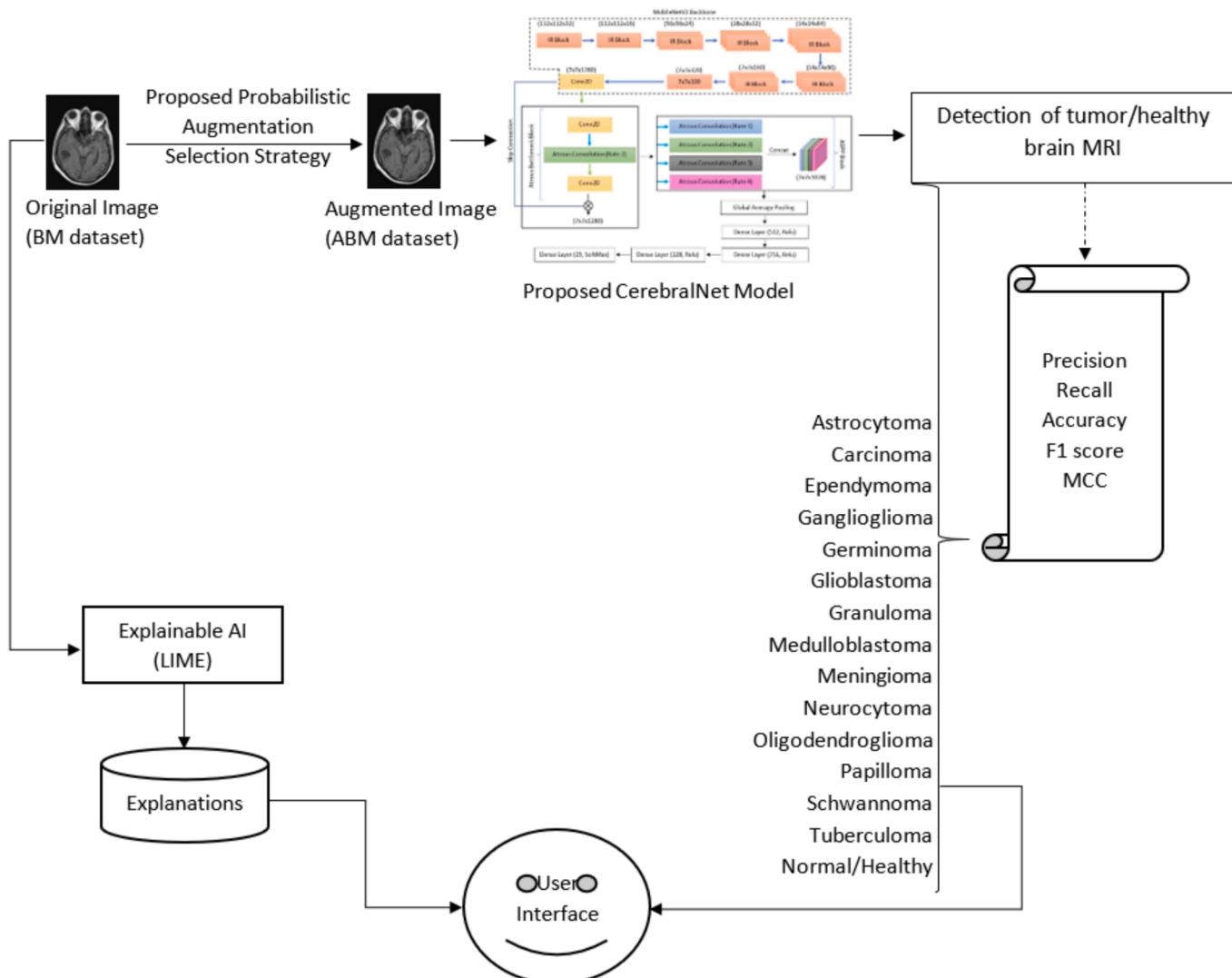


Fig. 1. The visual workflow representation of the proposed methodology.

white matter, and cerebrospinal fluid, enabling clear visualization of brain structures. Furthermore, T1 images effectively delineate tumor margins and demonstrate variations in tissue density, providing valuable cues for tumor identification and characterization. While contrast-enhanced T1 and T2 images offer valuable information, such as tumor vascularity and edema, respectively, T1 images provide a robust foundation for initial tumor detection and classification. Focusing on T1 images allows for a focused analysis of key anatomical features and enables the development of a flexible and robust model for brain tumor detection and classification [34]. It encompasses a diverse range of 15 classes, including 14 distinct tumor types: tuberculoma, astrocytoma, schwannoma, carcinoma, papilloma, ependymoma, oligodendrogloma, ganglioglioma, neurocytoma, germinoma, meningioma, granuloma, glioblastoma, and medulloblastoma, alongside a healthy class. The number of images within each class for the T1 MRI modality, utilized in this study, is detailed in [Table 1](#). This dataset, referred to as BM throughout this study, serves as a valuable resource for developing and evaluating the proposed brain tumor detection and classification model.

In the field of brain tumor detection and classification, clinical datasets often exhibit class imbalance which is reflected in [Table 1](#), with a disproportionate representation of certain tumor types or a limited number of images for rare tumor subtypes. This class imbalance can significantly influence the model performance, potentially leading to biased predictions and poor generalization. To handle this challenge effectively, we employ the proposed probabilistic augmentation selection strategy (discussed in the next section).

3.1.2. Generation of an augmented dataset (ABM) from the original BM dataset based on a new probabilistic augmentation selection algorithm

The employed data augmentation technique involves applying a combination of different image augmentation techniques based on a proposed probabilistic augmentation selection algorithm. The working principle of the algorithm is as follows.

Step 1: Define Augmentation Pool.

An augmentation pool is created, encompassing a diverse set of image transformation techniques. This pool includes Contrast Limited Adaptive Histogram Equalization (CLAHE) for enhancing image contrast, Defocus and Zoom Blur to simulate blurring effects, Downscale to mimic lower resolution images, Gaussian Blur to introduce noise, Horizontal Flip for geometric transformations, Hue Saturation Value adjustments to alter color characteristics, Random Gamma to adjust image brightness, Transpose to introduce rotational variations, and Pixel Dropout to simulate pixel dropout artifacts.

assigned moderate probabilities to enhance contrast and simulate varying intensity conditions across scans, which are common in real-world MRI datasets. Gaussian Blur, Pixel Dropout, Hue Saturation Value and Downscale were assigned lower moderate probabilities. These techniques help simulate differences in scanner settings or acquisition protocols, enhancing the model's robustness to real-world variability. Defocus and Zoom Blur were kept low to avoid excessive distortion while still introducing occasional soft-focus variations. These base probabilities were selected to balance diversity and realism, ensuring that augmentations enrich the dataset without introducing unrealistic or excessive distortions. Augmentations are applied stochastically based on these probabilities, allowing the model to experience a varied yet controlled set of transformed images during training.

In this work, we adopted a heuristic approach to set the augmentation base probabilities rather than relying on automated hyperparameter tuning algorithms. This decision was driven by the need for clinical interpretability, domain relevance, and computational efficiency. In medical imaging, particularly with brain MRI, it is critical to preserve anatomical fidelity and avoid introducing unrealistic distortions that could compromise model learning or diagnostic utility. Using domain knowledge allowed us to assign probabilities that reflect real-world variations—such as common intensity fluctuations, spatial transformations, and scanner-related artifacts—while ensuring that sensitive structures in the brain are not degraded by excessive or inappropriate augmentation. Moreover, hyperparameter tuning methods, while powerful, can be infeasible or unnecessary in cases where domain expertise already provides a strong basis for decision-making. By using a heuristic approach, we maintained full control over the augmentation process and ensured that all transformations remained biologically and clinically plausible, which is essential in healthcare applications. Thus, this method takes a list of base probabilities for each augmentation technique as input as mentioned in the section 4.1.2.

MRI scans are highly susceptible to variations in noise, intensity, and geometric distortions, and the sequence in which these augmentations are applied can significantly impact the final image and, consequently, the model's ability to accurately detect tumors. If we don't use any ordered factors and instead apply augmentations in a random order, several problems can arise. Applying certain augmentations in the wrong order can lead to an irreversible loss of diagnostically relevant information. A random order might also introduce unrealistic combinations of transformations, generating augmented images that do not reflect real-world variability in MRI scans. This can confuse the model and hinder its ability to generalize; for instance, applying a strong geometric distortion followed by an intensity adjustment might create

$$\text{Augmentation}_{\text{Pool}} = [\text{CLAHE}, \text{Defocus}, \text{Downscale}, \text{GaussianBlur}, \text{HorizontalFlip}, \text{HueSaturationValue},$$

$$\text{RandomGamma}, \text{Transpose}, \text{ZoomBlur}, \text{PixelDropout}] \quad (1)$$

Step 2: Assign Probabilities for each Augmentation and Order the Augmentations

An initial set of base probabilities was defined for each augmentation technique. Not all augmentation techniques are equally likely to occur or equally beneficial in simulating clinical scenarios. These base probabilities were determined through a combination of empirical observation and domain expertise, aiming to reflect the likelihood and relevance of each augmentation type in simulating real-world variations in brain MRI images. Horizontal Flip and Transpose were assigned relatively high probabilities as they offer significant spatial variability without compromising anatomical integrity—particularly useful in brain MRI, where lateral symmetry is common. CLAHE and Random Gamma were

an image with severely warped anatomy and unusual contrast, which is rarely seen in practice. Ultimately, training a model on data augmented with a suboptimal or random order of transformations can lead to

Table 1

The number of images within each class used in this study.

Class	No. of Images	Class	No. of Images
Astrocytoma	176	Meningioma	272
Carcinoma	66	Neurocytoma	130
Ependymoma	45	Oligodendrogloma	86
Ganglioglioma	20	Papilloma	66
Germinoma	27	Schwannoma	148
Glioblastoma	55	Tuberculoma	28
Granuloma	30	Normal	251
Medulloblastoma	23		

reduced accuracy, sensitivity, and specificity in tumor detection. The model might become overly sensitive to certain types of artifacts or fail to recognize tumors in images with specific combinations of distortions. A model trained with randomly ordered augmentations might also be less robust to the variability of real-world MRI scans, struggling to generalize to images acquired with different scanners or protocols, or images with unexpected combinations of noise and artifacts. To mitigate these issues ordering factors was employed which allows for a dynamic and adaptable augmentation strategy. By tailoring the sequence of transformations to each individual image based on the sampled factor values, we enable the model to learn from a wider range of augmentation sequences. This ultimately enhances its ability to generalize and accurately detect tumors in diverse and unseen MRI data. Therefore, we define four ordered factors: α (for blur: Defocus, Gaussian Blur and Zoom Blur), β (for intensity/color: CLAHE, Hue Saturation Value and Random Gamma), γ (for geometric: Downscale, Transpose, Horizontal Flip), and δ (for noise: Pixel Dropout) to ensure that the augmentations are applied in a controlled manner, preserving the integrity of the images while still introducing sufficient diversity for robust training.

Geometric augmentations such as downscaling, transposition, and horizontal flipping are applied at the beginning of the pipeline. This prioritization is critical because geometric operations affect the spatial structure and orientation of the image. By applying these transformations first, we ensure that all subsequent augmentations are executed on the altered spatial domain, which better simulates realistic anatomical variations that might occur across different patient scans or acquisition settings. Applying geometry-altering steps later in the pipeline risks distorting the effects of other augmentations (like blur or intensity) by shifting or warping the already-augmented pixel distribution. For example, flipping or transposing an already-blurred image may produce unnatural edge patterns, while flipping a noised image might spatially displace the noise, making it appear unrealistic.

Once the spatial structure has been adjusted, blur augmentations such as Defocus, Gaussian Blur, and Zoom Blur are applied. The purpose of introducing blur is to simulate different acquisition artifacts and focus-related inconsistencies that might naturally occur in clinical settings. Blurring after geometry ensures that the blur aligns with the final spatial layout of the image and mimics how focus loss would occur in real-world imaging—on already-oriented structures. Importantly, applying blur before geometric changes could distort the realistic distribution of blur, as spatial transformations might alter the directionality or intensity distribution of the blurring effect. Hence, positioning blur as second in the pipeline preserves realism and anatomical coherence.

Intensity and color transformations, such as CLAHE, HSV modifications, and Random Gamma correction, are designed to simulate variability in scanning conditions, such as lighting, tissue contrast, or equipment calibration. These augmentations should follow spatial and blur changes to ensure that the contrast or color manipulations are applied on top of a spatially consistent and realistically blurred image. By applying these adjustments third, the model is exposed to a wider range of tissue intensities and contrasts, enhancing its robustness to changes in brightness or modality-specific characteristics. Applying intensity augmentation too early could result in intensity variations that are later smeared or spatially redistributed by geometric or blur transformations, reducing the effectiveness and realism of the augmentation.

Finally, noise augmentation—specifically pixel dropout—is applied last. This decision is deliberate and essential to simulate sensor or environmental noise in its most natural form: as a final degradation layer on already-processed images. Introducing noise at the end ensures that the random dropout of pixels affects the image as a whole, including all the previous geometric, blur, and intensity modifications, thereby maximizing variability without compromising spatial structure or contrast logic. Moreover, if noise were applied earlier in the pipeline, subsequent transformations could modify or even diminish the noise effect, leading to inconsistency or unrealistic visual artifacts. Keeping noise as the last augmentation step retains the fidelity of the intended

noise simulation and allows for better control over the final image quality.

The carefully chosen order—geometric (γ) → blur (α) → intensity/color (β) → noise (δ)—is grounded in both visual realism and biomedical relevance. It mimics the natural progression of distortions that can occur during MRI acquisition and post-processing. This ordered strategy not only improves the robustness and generalizability of deep learning models but also preserves the anatomical correctness required for high-stakes medical imaging tasks such as diagnosis, segmentation, or disease classification.

The optimal values for the ordered factors were determined through a hyperparameter tuning algorithm [40] as discussed in section 4.1.2. The chosen values represent a balance between data augmentation and image realism, resulting in improved model performance and robustness.

Step 3: Select and Apply Augmentations for an Image.

For each image within the BM dataset, an empty list is initialized to store the selected augmentations. Subsequently, for each augmentation technique in the predefined pool, a random number between 0 and 1 is generated. If this randomly generated value falls below the corresponding probability threshold, the respective augmentation technique is appended to the list of selected augmentations. Finally, the identified augmentations are sequentially applied to the image based on the ordering factors ($\alpha, \beta, \gamma, \delta$), resulting in an augmented version of the original image.

The algorithm of the proposed probabilistic augmentation strategy is presented in the Algorithm1.

Algorithm1: Probabilistic Augmentation Strategy

```

Input:
• image: Input MRI image
• base_probs: Dictionary of base probabilities for each
augmentation
• augmentation_groups: A mapping of ordered factors to
corresponding augmentation types:
o  $\alpha \rightarrow$  {Defocus, Gaussian Blur, Zoom Blur}
o  $\beta \rightarrow$  {CLAHE, HueSaturationValue, RandomGamma}
o  $\gamma \rightarrow$  {Downscale, Transpose, HorizontalFlip}
o  $\delta \rightarrow$  {PixelDropout}
Output: Augmented Image

1. Initialize:
   selected_augmentations  $\leftarrow$  empty dictionary
   factor_values  $\leftarrow$  empty dictionary
2. For each factor f in  $\{\alpha, \beta, \gamma, \delta\}$ :
   a. For each augmentation a in augmentation_groups[f]:
      i. Retrieve base probability p  $\leftarrow$  base_probs[a]
      ii. Generate random number r  $\in [0, 1]$ 
      iii. If r < p:
           Add a to selected_augmentations[f]
   b. If selected_augmentations[f] is not empty:
      i. Sample random float value v  $\in [0.0, 1.0]$ 
      ii. Assign factor_values[f]  $\leftarrow$  v
3. Sort selected_augmentations by factor_values in ascending
order:
   ordered_factors  $\leftarrow$  sort(factor_values.keys(),
by=factor_values)
4. For each factor f in ordered_factors:
   For each augmentation a in selected_augmentations[f]:
      - Sample required parameters for a
      - Apply to an image
5. Return augmented image

```

This proposed approach not only mitigates the impact of class imbalance and introduces more variability in the dataset but also enhances CerebralNet's ability to generalize and accurately classify brain tumors in diverse clinical scenarios. Through the introduction of these variations, the class imbalance within the BM dataset is mitigated where each of the 15 classes is represented by an equal number of images (400), ensuring equitable representation of all classes in the training and evaluation process. The workflow of the image augmentation process is visually represented in Fig. 2.



Fig. 2. The workflow of the image augmentation process.

3.2. Proposed CerebralNet architecture

Accurate and robust brain tumor detection and classification demands a model that can effectively capture the intricate spatial and contextual information inherent in MRI images for several critical reasons. Firstly, brain tumors exhibit a wide range of morphologies, from well-defined masses to diffuse infiltrations, necessitating the model to accurately discern subtle variations in tumor shape, size, and boundaries. Secondly, the surrounding brain anatomy plays a crucial role in tumor characterization. For instance, the proximity of a tumor to critical brain structures, such as the brainstem or ventricles, can significantly impact its behavior and prognosis. Therefore, the model must be able to effectively analyze the spatial relationships between the tumor and surrounding brain structures, such as the cortex, white matter tracts, and cerebrospinal fluid. Furthermore, the model must be able to effectively capture the subtle variations in tissue intensity and texture within and around the tumor, which can provide valuable clues regarding tumor types. To achieve this, our approach incorporates Atrous Spatial Pyramid Pooling (ASPP) and Atrous Convolution blocks into the MobileNetV2 architecture.

The choice of MobileNetV2 as the foundation for our brain tumor detection and classification model is driven by several key considerations [35]. Firstly, MobileNetV2's lightweight architecture, characterized by its depthwise separable convolutions and inverted residual blocks, significantly reduces computational cost and memory usage compared to traditional deep convolutional neural networks. This efficiency is paramount for deployment in limited resource environments, like point-of-care settings or mobile devices, enabling rapid and efficient analysis of brain tumor images. Secondly, despite its lightweight architecture, MobileNetV2 has demonstrated remarkable performance on a wide range of image recognition tasks, achieving high accuracy while maintaining a small model size. This combination of high accuracy and computational efficiency makes MobileNetV2 an ideal choice for developing practical and deployable solutions for brain tumor detection and classification in clinical settings. Furthermore, MobileNetV2's pre-trained weights on ImageNet provide a strong initial representation of visual features, accelerating the training process and improving the model's ability to extract relevant features from brain tumor images.

While MobileNetV2 excels in efficiency, its standard convolutions have limitations in capturing the diverse range of sizes and shapes exhibited by brain tumors. The fixed receptive fields of standard convolutions can hinder the model's ability to effectively capture both fine-grained details, such as the precise location of small tumors, and broader contextual information, such as the relationship between the tumor and surrounding brain structures. To address this limitation, we integrated Atrous Spatial Pyramid Pooling (ASPP) and Atrous Convolution blocks into the MobileNetV2 architecture to create a new Atrous MobileNetV2 model. The block details of the proposed Atrous MobileNetV2 as well as the motivation of the same are as follows.

3.2.1. MobileNetV2 backbone

The model utilizes the pre-trained MobileNetV2 architecture as its

foundation, excluding the top classification layers. This uses the pre-trained weights learned on the ImageNet dataset, providing a strong initial representation of visual features.

3.2.2. Integration of the Atrous bottleneck block

An atrous bottleneck block is introduced after the MobileNetV2 backbone. The atrous bottleneck block is a key component of the proposed architecture. It employs a “bottleneck” structure, where the number of channels is initially reduced before being expanded. This block consists of three layers. (i) 1x1 Depthwise Separable Convolution: The block begins with a 1x1 depthwise separable convolution, acting as a bottleneck. This operation significantly reduces the number of input channels, leading to a substantial decrease in computational cost. Depthwise separable convolutions, unlike traditional convolutions, factorize the convolution operation into two steps, i.e., depthwise convolution which applies a single filter to each input channel independently, and pointwise convolution which applies 1x1 convolution to combine the outputs of the depthwise convolutions. This factorization significantly reduces the number of parameters and computations compared to traditional convolutions, making the model more efficient in terms of both memory and processing time [36]. (ii) Atrous Convolution: Atrous convolutions [37], also known as dilated convolutions, are a powerful technique for capturing multi-scale information within an image. Unlike traditional convolutions, which have a fixed receptive field size, atrous convolutions “dilate” the convolution kernel by inserting spaces between the kernel weights. This effectively increases the receptive field of the convolution without increasing the filter size or the number of parameters. The “dilation rate” parameter controls the spacing between the weights of the convolution kernel. By increasing the dilation rate, the receptive field of the convolution expands, allowing the model to capture larger regions of the input image. In the context of brain tumor detection and classification, atrous convolutions are particularly beneficial. Brain tumors can exhibit varying sizes and shapes, and their appearance can be influenced by surrounding anatomical structures. Atrous convolutions with different dilation rates enable the model to capture contextual information at multiple scales, effectively capturing both fine-grained details and larger spatial relationships within the tumor and its surrounding regions. Also, Atrous convolutions achieve this expanded receptive field without increasing the spatial dimensions of the feature maps. This is crucial for accurately localizing the tumor within the image, as preserving spatial information is essential for accurate diagnosis and treatment planning. By incorporating atrous convolutions within the atrous bottleneck block, the proposed model can effectively capture multi-scale contextual information, leading to improved performance in detecting and classifying brain tumors of varying sizes and shapes. (iii) 1x1 Depthwise Separable Convolution: Finally, another 1x1 depthwise separable convolution is applied to adjust the number of channels. A residual connection is employed, where the outcome of the block is added to the original input. This technique mitigates the issue of vanishing gradients, empowering CerebralNet to learn deeper representations more effectively. Residual connections also facilitate the flow of information through the network,

improving training stability and accelerating convergence.

3.2.3. Integration of the Atrous Spatial Pyramid Pooling (ASPP)

In the proposed architecture, the ASPP module [10] comprises four parallel branches, each employing a 3×3 depthwise separable convolution with different dilation rates: 1, 2, 3, and 4. Dilation rate 1 branch captures fine-grained details and local information within the image. Dilation rate 2 branch captures information at a slightly larger scale, incorporating a wider context around each pixel. Dilation rate 3 branch captures information at a larger scale, effectively increasing the receptive field of the convolution. Finally, the dilation rate 4 branch captures the most global information, providing a broad context for the classification task. By employing these parallel branches with varying dilation rates, the ASPP module effectively captures multi-scale contextual information within the image. This is crucial for accurately detecting and classifying brain tumors of varying sizes and shapes, as the appearance of a tumor can vary significantly depending on its size and location within the brain. The outputs of these four branches are then concatenated along the channel dimension. This concatenation effectively combines the information captured at different scales, providing the model with a richer and more comprehensive representation of the input image.

3.2.4. Global average pooling

After the ASPP module, global average pooling is applied to the

feature maps. This layer reduces the spatial dimensionality of feature maps to a single vector, preserving global information [39]. This operation reduces parameters in subsequent fully connected layers, improving efficiency and mitigating overfitting.

3.2.5. Fully connected layers

The model includes three fully connected layers with 512, 256, and 128 neurons, respectively (finalized by using the hyperparameter tuning algorithm discussed in Section 4.4), followed by batch normalization. These layers extract high-level features from the pooled features.

3.2.6. Output layer

The final layer is a fully connected layer with 15 output neurons, representing the 15 different classes used in this study, using SoftMax activation for probability distribution.

The visual representation of the proposed CerebralNet architecture is shown in Fig. 3. Here the inverted residual block of MobileNetV2 architecture is represented as IR Block.

3.3. Performance measures

To evaluate CerebralNet's performance, a set of metrics was employed. By evaluating CerebralNet's performance on both the original (BM) and augmented (ABM) datasets, we can acquire valuable insights of its capabilities. These metrics, including recall, precision,

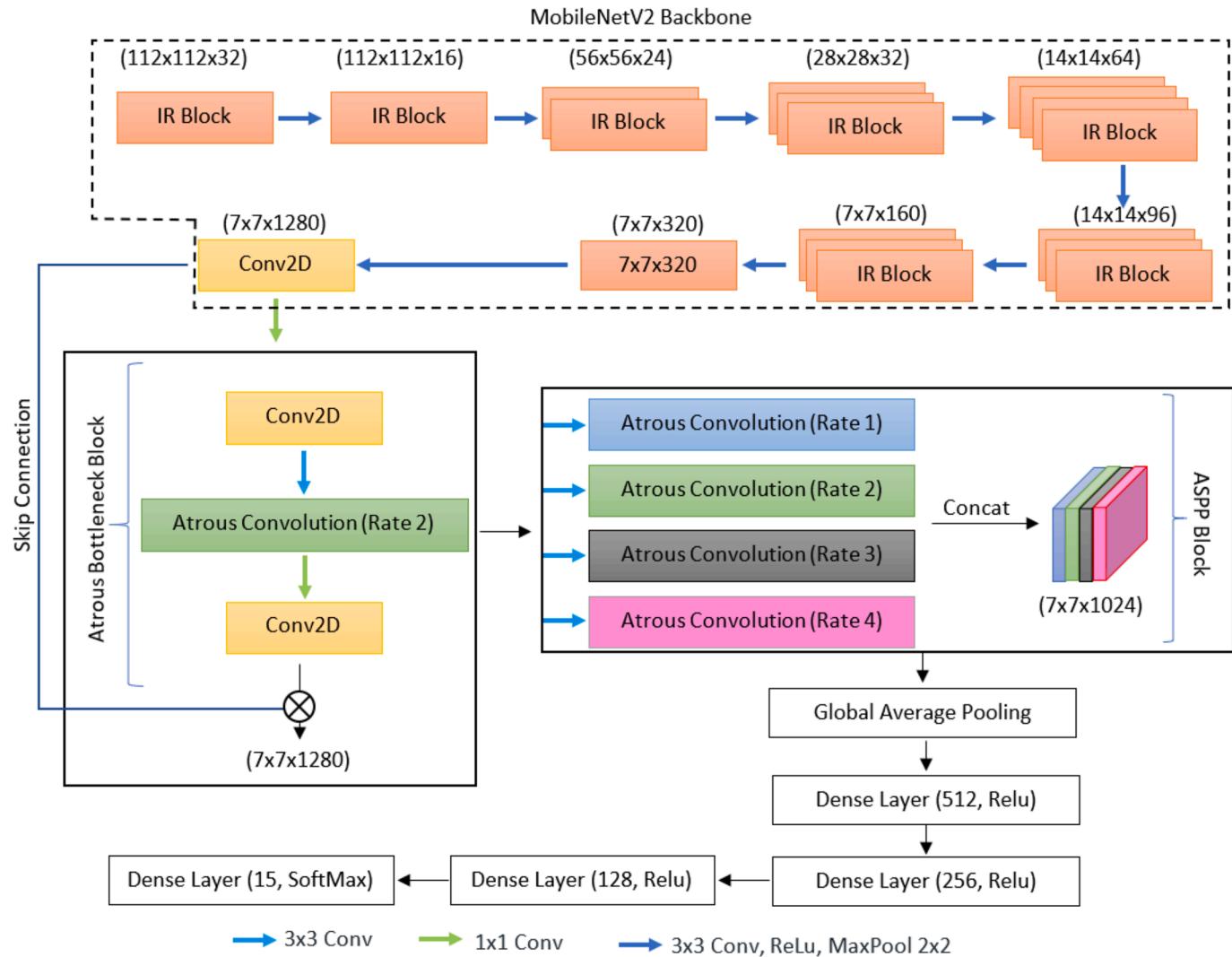


Fig. 3. The visual representation of the proposed CerebralNet architecture.

accuracy, F1-score, and MCC are derived from the confusion matrix, providing a detailed assessment of the model's ability to correctly classify brain tumor images across various classes, including healthy controls, while effectively differentiating between different tumor types. Here are the motivations for using these metrics in this study and the respective equations [38].

Accuracy: This overall metric delivers an overall measure of CerebralNet's ability to correctly classify both healthy and tumor-affected brain images. It is crucial to understand the model's overall performance in the clinical setting.

$$\text{Accuracy} = \frac{TN + TP}{TN + FN + TP + FP} \quad (2)$$

Precision: This metric is particularly important in the context of brain tumor detection and classification, as it emphasizes CerebralNet's capability to minimize false positives. High precision is crucial to avoid unnecessary interventions and reduce patient anxiety caused by false alarms.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (3)$$

Recall (Sensitivity): This metric is critical in the context of brain tumor detection and classification as it highlights CerebralNet's capability to correctly identify all instances of brain tumors. High recall is crucial to ensure that no tumors are missed, as missed diagnoses can have severe consequences for patient health.

$$\text{Recall} = \frac{TP}{TP + FN} \quad (4)$$

F1-Score: This metric offers a balanced measure of recall and precision, presenting a comprehensive assessment of CerebralNet's performance across both aspects. It is particularly valuable in situations where both minimizing false positives and maximizing true positives are equally important.

$$\text{F1 Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (5)$$

MCC: This metric offers a more reliable assessment of CerebralNet's performance, particularly in cases of class imbalance, which is common in medical image datasets. It considers all four combinations of true and false positives and negatives, providing a balanced measure of CerebralNet's performance.

$$\text{MCC} = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP) \times (TP + FN) \times (TN + FP) \times (TN + FN)}} \quad (6)$$

By evaluating the model's performance across these metrics on BM and ABM, we can gain a comprehensive understanding of its strengths and weaknesses, identify areas for improvement, and ultimately contribute to the development of more reliable and effective brain tumor detection and classification systems.

3.4. LIME: Understanding the decisions of the proposed model

LIME [12] is an effective method that provides crucial insights into the decision-making process of complex machine learning models, particularly deep neural networks, often referred to as "black boxes." By generating a locally interpretable linear model that captures the local behavior of the complex model around a specific input, this method offers human-understandable descriptions. This process involves introducing small variations to the input data and fitting a simple linear model to the resulting instances and their predictions. The resulting explanation, typically represented as a collection of features and their associated weights, highlights the relative significance of individual features in driving the model's prediction. Furthermore, this method

provides valuable insights by differentiating between positive and negative explanations, identifying the features that positively influence the prediction and those that negatively influence it.

LIME begins by perturbing the input image. This involves generating a set of modified versions of the original image, such as blurring, cropping, or removing portions. These perturbations simulate variations that might occur in clinical scenarios. Each perturbed image is then fed into the black-box model to obtain a corresponding prediction. LIME assigns weights to these perturbed images based on their proximity to the original image. Images that are more similar to the original image receive higher weights. This ensures that the explanations are focused on the local neighborhood around the original image. The proximity weighting of LIME is calculated by using the equation (7).

$$\pi_x(z) = \exp \left[\frac{-D(x, z)^2}{\sigma^2} \right] \quad (7)$$

where $\pi_x(z)$ is the proximity weight for a perturbed image z , x is the original image, $D(x, z)$ is the distance between x and z , and σ is a kernel width parameter that controls the locality of the explanation.

LIME then fits a simple, interpretable model (often a linear model or a decision tree) to the predictions of the black-box model on these perturbed images. This local model approximates the behavior of the complex model in the vicinity of the original image. The coefficients of the fitted linear model represent the importance of different features in the prediction. These coefficients are then used to generate an explanation, highlighting the most influential image features (e.g., specific regions, textures, or edges) that play a role to CerebralNet's prediction. The explanation model of LIME uses the equation (8).

$$g(z') = w_0 + \sum_i w_i z'_i \quad (8)$$

where $g(z')$ is the prediction of the local linear model for a perturbed image z' , w_i are the weights of the features in the linear model, and z'_i are the features of the slightly changed image.

LIME aims to find the weights (w_i) that minimize the following objective function:

$$L(f, g, \pi_x) + \Omega(g) \quad (9)$$

where $L(f, g, \pi_x)$ measures the unfaithfulness of the explanation model g in approximating the predictions of the original model f in the local neighborhood, and $\Omega(g)$ is a penalty term that encourages the simplicity of the explanation model.

By minimizing this objective function, LIME generates a locally interpretable explanation that faithfully approximates the behavior of the complex model around the specific input image.

By perturbing the input brain MR images and observing the model's predictions on these modified versions, LIME constructs a simple, locally linear model that approximates the complex behavior of the proposed model around a specific MR input image. This process allows us to identify the most influential image regions, such as areas with specific textures, intensities, or boundaries, that significantly contribute to the model's classification decision. By visualizing these influential regions, LIME offers a window into CerebralNet's internal reasoning, helping us understand which image features the proposed model deems most important for distinguishing between healthy brain tissue and different tumor types. This enhanced interpretability not only builds trust in the proposed model's predictions but also facilitates the identification of potential biases and limitations, ultimately improving the reliability and clinical applicability of the developed system.

4. Experimental results and performance analysis

To ensure efficient and timely training of CerebralNet, a high-performance computing environment was utilized. Specifically, a T4 GPU with substantial memory was employed. The T4 GPU, renowned for

its powerful processing capabilities and high memory bandwidth, significantly accelerated the training process by enabling parallel processing of numerous computations involved in deep learning. This accelerated training facilitated rapid experimentation with different hyperparameter configurations, leading to more efficient model optimization and reduced overall development time. Furthermore, the high RAM capacity of the T4 GPU was crucial for handling the large number of parameters within the proposed architecture and the processing of high-resolution brain MRI images, ensuring smooth and uninterrupted model training. The training process was implemented using TensorFlow version 2.17.1, which offers a powerful and flexible platform for executing experiments with transfer learning models.

4.1. Datasets

Two datasets were incorporated in this study to enhance the generalization and robustness of the CerebralNet model.

4.1.1. Baseline dataset (BM)

The BM dataset, as mentioned in [Section 3.1.1](#), is derived from a subdivision of the renowned “Brain Tumor MRI Images 44 Classes” dataset [32], serving as the foundation for this study. [Fig. 4](#) visually demonstrates representative instances from the BM dataset, providing a visual understanding of the data utilized in this research.

4.1.2. Augmented dataset (ABM) using proposed probabilistic augmentation selection strategy

The ABM dataset was generated by selecting and applying a diverse set of augmentations using the proposed probabilistic strategy described in [section 3.1.2](#). The $\text{Base}_{\text{Prob}}$ of the augmentations are selected as follows: Horizontal Flip: 0.8, CLAHE: 0.5, Random Gamma: 0.5, Gaussian Blur: 0.3, Pixel Dropout: 0.2, Hue Saturation Value: 0.2, Defocus: 0.05, Zoom Blur: 0.05, Downscale: 0.2 and Transpose: 0.3. The optimal combination of ordered factors is selected using a hyperparameter tuning algorithm [40] as shown below.

Symbol	Group	Intended Order	Suggested Value Range
Γ	Geometric	1st	[0.0 – 0.4]
A	Blur	2nd	[0.3 – 0.6]
B	Intensity/Color	3rd	[0.5 – 0.8]
Δ	Noise	4th	[0.7 – 1.0]

The suggested value ranges for these factors are intentionally overlapping to introduce controlled randomness. This ensures that while the intended order is preserved on average, the overlapping ranges allow augmentations to switch places occasionally, but only with reasonable neighbors (e.g., blur might precede geometric once in a while). This stochasticity helps prevent the model from overfitting to a strict, artificial augmentation pipeline.

Furthermore, for each selected augmentation technique, the corresponding hyperparameter values were sampled by using the hyperparameter tuning algorithm. These ranges were carefully chosen to introduce realistic variations while preserving the essential features of

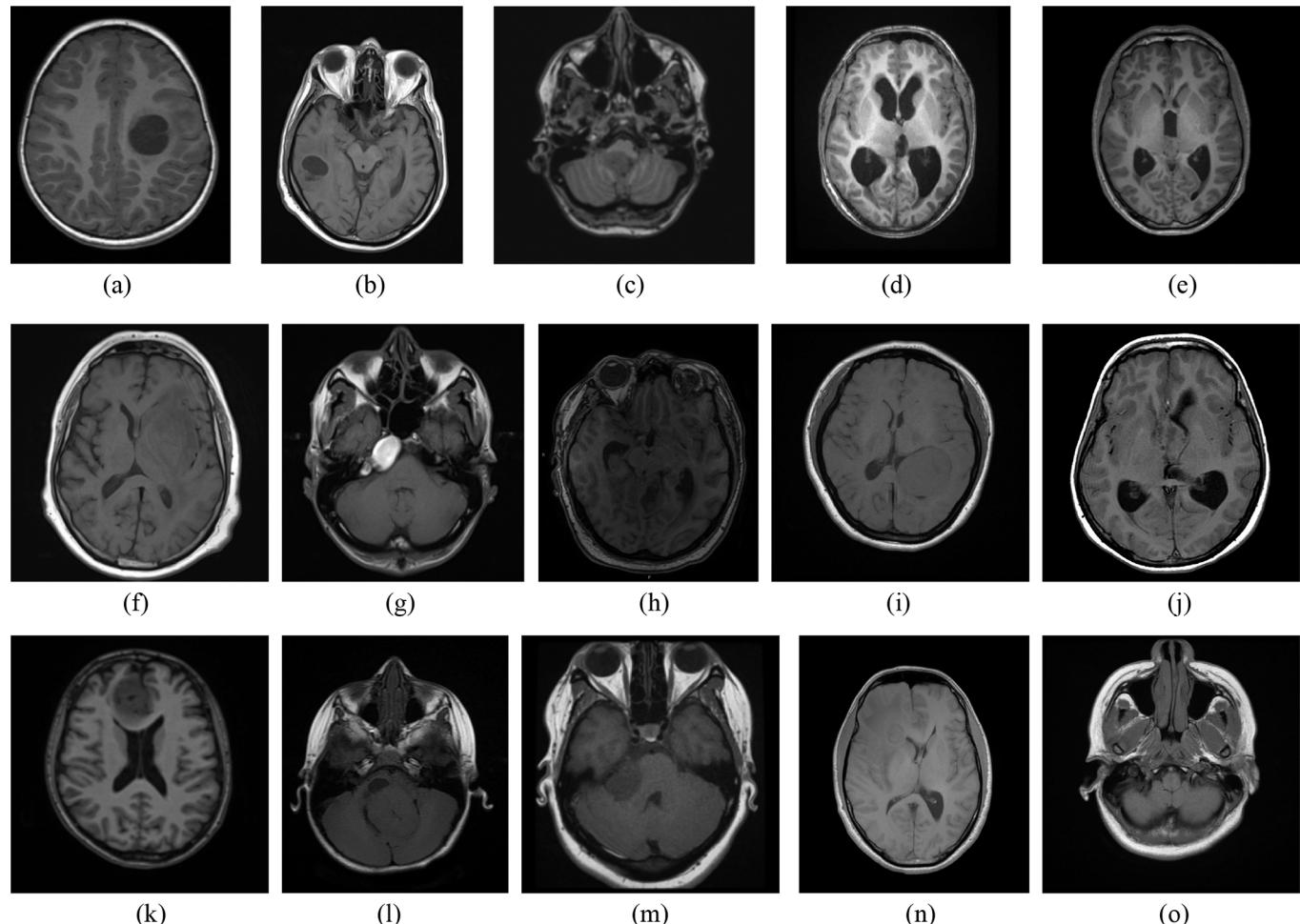


Fig. 4. The representative examples from the BM dataset: (a) Astrocytoma, (b) Carcinoma, (c) Ependymoma, (d) Ganglioglioma, (e) Germinoma, (f) Glioblastoma, (g) Granuloma, (h) Medulloblastoma, (i) Meningioma, (j) Neurocytoma, (k) Oligodendrogioma, (l) Papilloma, (m) Schwannoma, (n) Tuberculoma, (o) Normal.

the brain tumor within the image. For instance, the CLAHE limit was set within a range of [1, 1.5], the Defocus range was set to [1, 1.5], the Downscale range was set to [0.8, 1.0], the Gaussian Blur kernel size was varied within a range of [1,3], Hue Saturation Value adjustments were applied within a range of [-5, 5], Random Gamma was applied within a range of [0.8, 1.2], the Transpose operation was randomly applied, the Zoom Blur maximum factor and step factor parameters were set within a range of [1.0, 1.03] and [0.01, 0.01] respectively, and Pixel Dropout was applied with a maximum dropout rate of 210 pixels. This randomized approach generated a diverse set of augmented images, effectively simulating the inherent variability and potential artifacts encountered in clinical MRI scans, thereby enhancing the generalizability and robustness of the trained model. Fig. 5 (Class name [Augmentations applied]) visually demonstrates the impact of these augmentations (one or multiple) on the instances within the ABM dataset.

4.2. Training criteria

To rigorously evaluate the performance of the proposed model, a 5-fold approach was employed. The brain tumor dataset was divided into training and testing sets, with an 80/20 split ensuring a balanced depiction of all tumor classes and the healthy class within each subset. To further enhance CerebralNet's generalization and robustness, a 5-fold cross-validation strategy was implemented on the training data. This involved partitioning the training data into five equal folds, iteratively training the model on four-folds, and validating its performance

Table 2

The data distribution across the training, validation, and testing sets used in this study.

Dataset	Class Name	Training Set	Validation Set	Test Set
BM	Astrocytoma	113	28	35
	Carcinoma	42	11	13
	Ependymoma	29	7	9
	Ganglioglioma	13	3	4
	Germinoma	18	4	5
	Glioblastoma	35	9	11
	Granuloma	19	5	6
	Medulloblastoma	14	4	5
	Meningioma	174	43	55
	Neurocytoma	83	21	26
	Oligodendrogioma	55	14	17
	Papilloma	42	11	13
	Schwannoma	94	24	30
	Tuberculoma	18	4	6
	Normal (Healthy)	161	40	50
ABM	All Classes	256	64	80

on the remaining fold. This process was repeated five times, ensuring that each fold served as the validation set once, thereby maximizing the utilization of the training data. Table 2 provides a detailed breakdown of the data distribution across the training, validation, and testing sets.

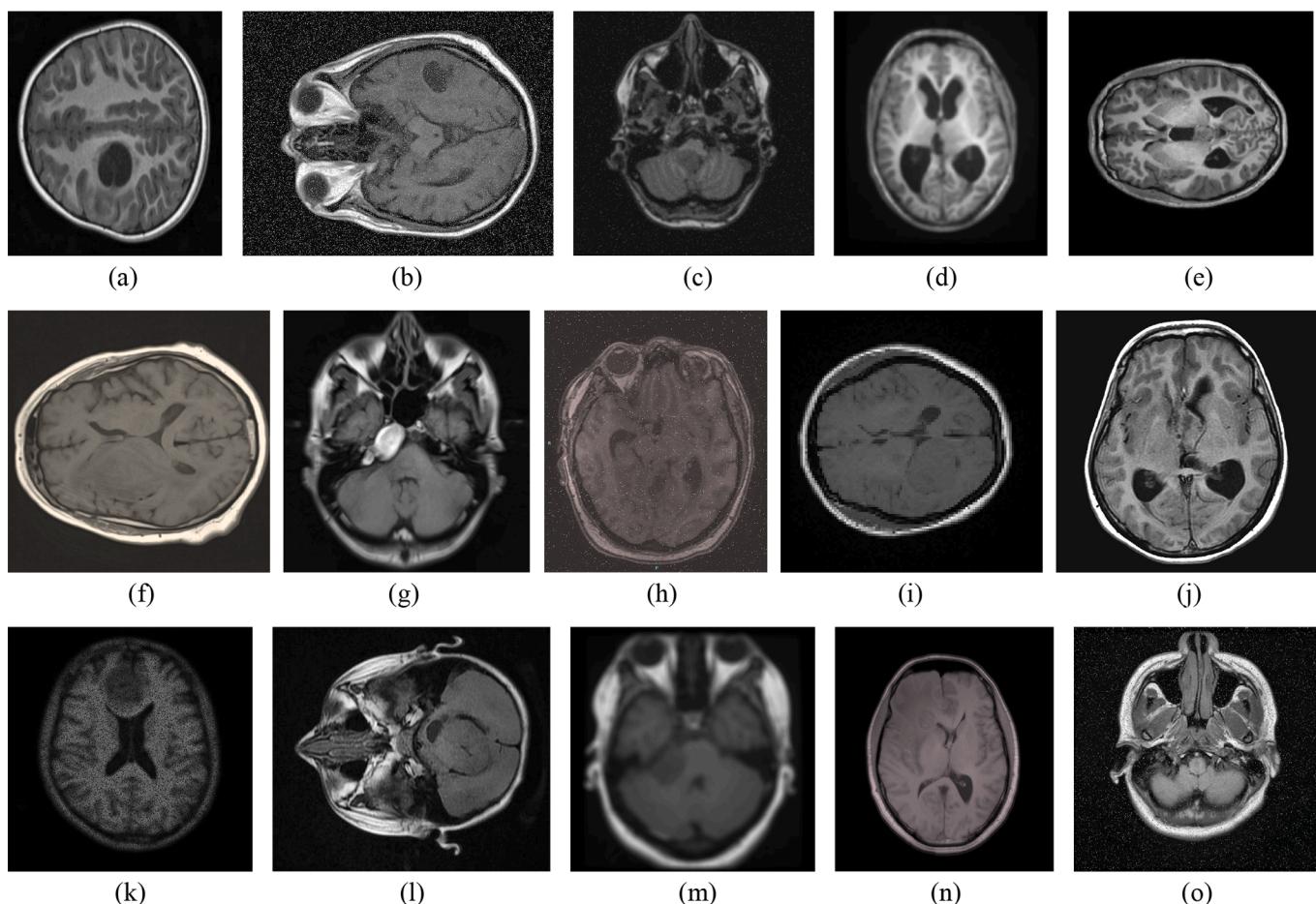


Fig. 5. The representative examples from the ABM dataset: (a) Astrocytoma [Horizontal Flip, Random Gamma], (b) Carcinoma [Pixel Dropout, Transpose], (c) Ependymoma [Defocus, Pixel Dropout], (d) Ganglioglioma [Gaussian Blur], (e) Germinoma [Random Gamma, Horizontal Flip], (f) Glioblastoma [Hue Saturation Value, Transpose], (g) Granuloma [Zoom Blur, CLAHE], (h) Medulloblastoma [Pixel Dropout, Hue Saturation Value, Random Gamma], (i) Meningioma [Transpose, Downscale], (j) Neurocytoma [CLAHE, Hue Saturation Value, Random Gamma], (k) Oligodendrogioma [Zoom Blur, Pixel Dropout], (l) Papilloma [CLAHE, Transpose, Defocus], (m) Schwannoma [Downscale, Defocus], (n) Tuberculoma [Hue Saturation Value], (o) Normal [Pixel Dropout, CLAHE, Downscale].

4.3. Finding optimal hyperparameters for transfer learning models

While transfer learning models demonstrate remarkable capabilities in automatically learning complex relationships within brain tumor images, the risk of overfitting remains a concern. Random sampling during training can inadvertently lead the model to establish connections between features that may not be relevant in clinical scenarios, hindering its ability to generalize to unseen data. To mitigate this risk and optimize model performance, we employed hyperparameter tuning. This process involved systematically evaluating different combinations of batch sizes (128, 64, and 32) and learning rates (0.1, 0.01, 0.001, and 0.0001) to identify the optimal hyperparameter settings for each of the 12 transfer learning models considered in this study. The ABM dataset, characterized by its balanced class distribution and diverse image variations, served as the foundation for this hyperparameter tuning process. Furthermore, to prevent overfitting and enhance generalization, we incorporated EarlyStopping [41], a technique that monitors the model's performance on a validation set and halts training if the validation loss fails to improve for a specified number of epochs. This approach effectively prevents the model from memorizing noise within the training data, leading to improved generalization and enhanced performance on unseen brain tumor images. The optimal hyperparameter settings identified through this process, including a learning rate of 0.001 and a batch size of 64, significantly reduced the gap between training and validation performance, demonstrating the effectiveness of this optimization strategy. The optimal hyperparameter settings for each model are summarized in [Table 3](#).

4.4. Construction of the proposed CerebralNet model

The architectural details of the 12 transfer learning models utilized in this study are presented in [Table 4](#).

To select the most suitable transfer learning model for our brain tumor detection and classification system, an in-depth evaluation of the abovementioned 12 distinct transfer learning architectures was conducted. This analysis used the augmented brain tumor dataset (ABM), chosen for its balanced class distribution and diverse image variations, ensuring a robust evaluation environment. The performance of each transfer learning model concerning accuracy and training time was assessed, and the results are summarized in [Table 5](#). This evaluation process guided the selection of the most promising model architecture for subsequent refinement and integration into the final system.

From [Table 5](#) it is clear that while models such as DenseNet121, and InceptionV3 demonstrated high accuracy (94.67 %, and 94.42 %, respectively), MobileNetV2 achieved a competitive accuracy of 95.42 % while exhibiting significantly faster training times (3573.86 s) compared to these deeper and more complex models. DenseNet169 and DenseNet201, despite achieving high accuracy, required substantially longer

Table 3

The optimal hyperparameter settings for each model used in this study.

Transfer learning Model	Optimized Hyperparameter Train Approach	Optimizer	Loss function	Batch size	Learning rate	Epochs
V16 [3,14]	5-fold cross-validation	Adam	Categorical cross-entropy	64	0.001	40
V19 [9,10]						40
R50 [16]						40
R101 [9]						35
R152 [10]						30
IV3 [9]						35
MV2 [2,9]						40
MV3S [10]						40
MV3L [10]						40
D121 [15]						30
D169 [10]					0.01	30
D201 [15]						30

*V16: VGG16, V19: VGG19, R50: ResNet50, R101: RestNet101, R152: RestNet152, IV3: InceptionV3, MV2: MobileNetV2, MV3S: MobileNetV3_Small, MV3L: MobileNetV3_Large, D121: DenseNet121, D169: DenseNet169, D201: DenseNet201.

Table 4

The architectural details of the 12 transfer learning models utilized in this study.

Transfer learning model	Layer depth (with input layer)	Input image size	Trainable parameters
VGG16 [3,14]	19	224 x 224 x 3	14,714,688
VGG19 [9,10]	22	23,534,592	20,024,384
ResNet50 [16]	175	42,552,832	58,219,520
ResNet101 [9]	345	21,768,352	551
ResNet152 [10]	515	2,223,872	154
InceptionV3 [9]	311	927,008	MobileNetV3_Small [10]
MobileNetV2 [2,9]	154	2,971,952	MobileNetV3_Large [10]
MobileNetV3_Small [10]	157	6,953,856	DenseNet121 [15]
MobileNetV3_Large [10]	187	12,484,480	DenseNet169 [10]
DenseNet121 [15]	427	18,092,928	DenseNet201 [15]
DenseNet169 [10]	595		
DenseNet201 [15]	707		

Table 5

The performance of each transfer learning model used in this study.

Transfer learning model	Test Accuracy (%), Train Time (Sec)	Transfer learning model	Test Accuracy (%), Train Time (Sec)
VGG16 [3,14]	91.83, 2487.25	MobileNetV2 [2,9]	95.42, 3573.86
VGG19 [9,10]	90.75, 3570.8	MobileNetV3_Small [10]	88.75, 3888.33
ResNet50 [16]	88.67, 1974.47	MobileNetV3_Large [10]	89.42, 2835.64
ResNet101 [9]	89.5, 4466.63	DenseNet121 [15]	94.67, 3641.07
ResNet152 [10]	88.92, 4663.11	DenseNet169 [10]	95.25, 3846.49
InceptionV3 [9]	94.42, 3730.57	DenseNet201 [15]	95.17, 4169.13

training times (3846.49 s and 4169.13 s, respectively). This exceptional balance of high accuracy and computational efficiency makes MobileNetV2 particularly well-suited for clinical applications, where rapid and efficient analysis is crucial. That's why we selected MobileNetV2 as the base model of this study.

To further enhance the capabilities of MobileNetV2, we proposed the integration of Atrous Spatial Pyramid Pooling (ASPP) and Atrous Convolution blocks, as detailed in [Section 3.2](#). The optimal hyperparameters for the proposed model were determined through a rigorous hyperparameter tuning process. This involved systematically exploring a wide range of configurations. For the atrous bottleneck block, the dilation rate was systematically varied within a range of 1 to 6. In the ASPP module, the number of atrous convolution layers was explored from 1 to 6, with corresponding dilation rates ranging from 1 to 6 for each layer. Additionally, the hyperparameter tuning process investigated different configurations for the fully connected layers, including

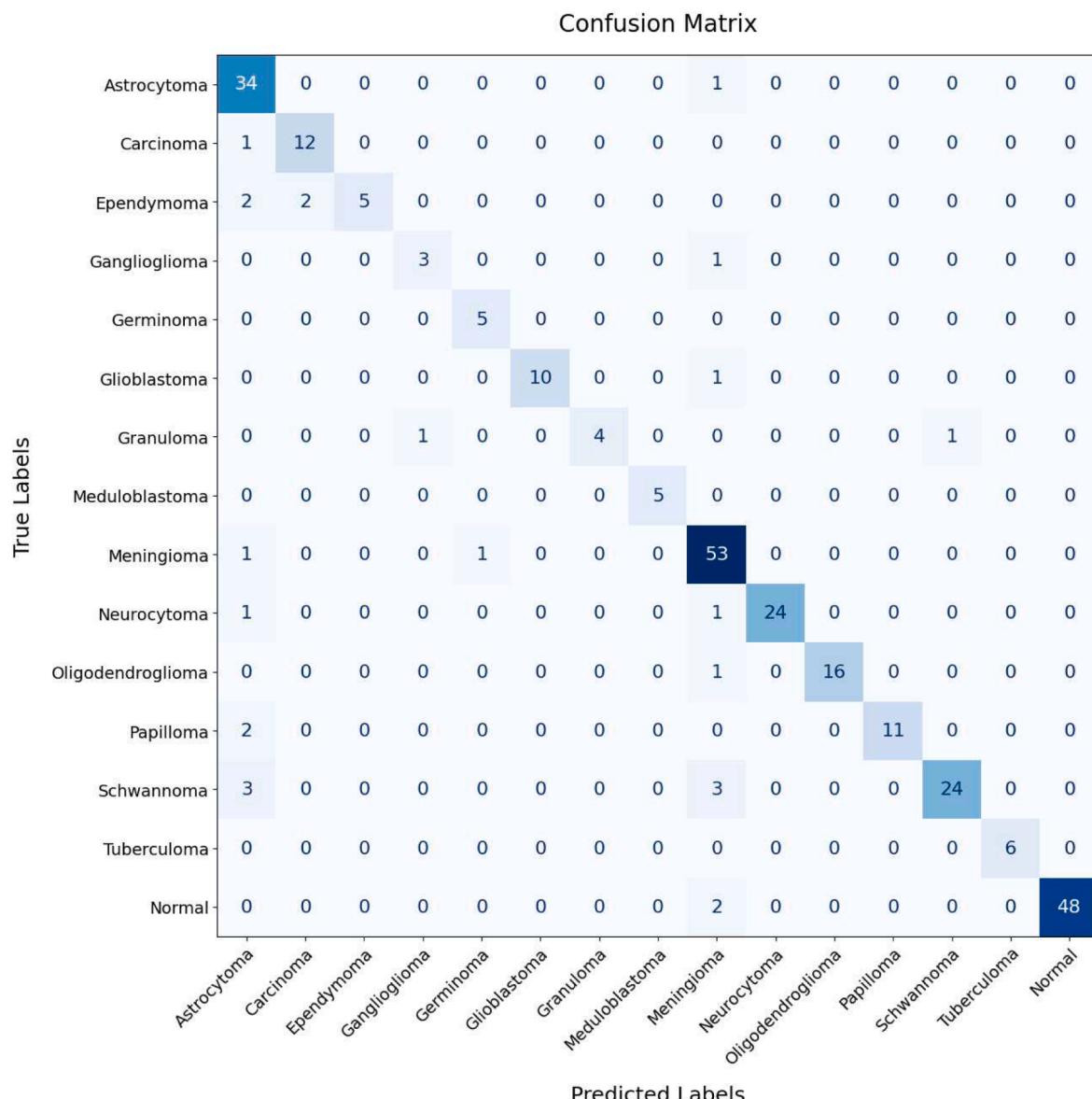
the number of layers (1 to 6) and the number of neurons per layer (1024, 512, 256, and 128). The optimal hyperparameter configuration identified through this process includes a dilation rate of 2 for the atrous bottleneck block, four atrous convolution layers within the ASPP module with dilation rates of 1, 2, 3, and 4, and three fully connected layers with 512, 256, and 128 neurons, respectively. This optimized configuration was found to yield the best performance on the validation set, demonstrating the effectiveness of the hyperparameter tuning process in enhancing the model's accuracy and generalization.

4.5. The proposed CerebralNet model's performance on the BM dataset

To assess the performance of the proposed model in a clinical setting, we evaluated its performance on the test set of the BM dataset. An analysis was conducted using a confusion matrix generated from

CerebralNet's predictions on the BM dataset (Fig. 6). This matrix provides valuable insights into the model's classification accuracy for each brain tumor class, including the healthy class, and effectively highlights instances of misclassification. This detailed analysis enables a thorough understanding of CerebralNet's strengths and weaknesses, identifying areas for potential improvement and guiding future refinements.

While the CerebralNet model demonstrates promising results, analysis of its performance on the imbalanced BM dataset, as visualized in Fig. 6, reveals certain limitations. Certain tumor classes are represented by a very limited number of samples (e.g., Ganglioglioma: 4, Ependymoma: 9, Germinoma and Medulloblastoma: 5, Granuloma and Tuberculoma: 6). This class imbalance can inadvertently bias the model towards learning features that are more prevalent in the majority classes, potentially leading to the neglect of subtle features associated with rare tumor types. This phenomenon can significantly impact the model's



Accuracy (%)	F1-Score (%)	Recall (%)	Precision (%)	MCC (%)	Training Time (Sec)
91.23	91.17	91.23	92.42	90.17	1392.47

Fig. 6. The CerebralNet model's performance on the BM dataset.

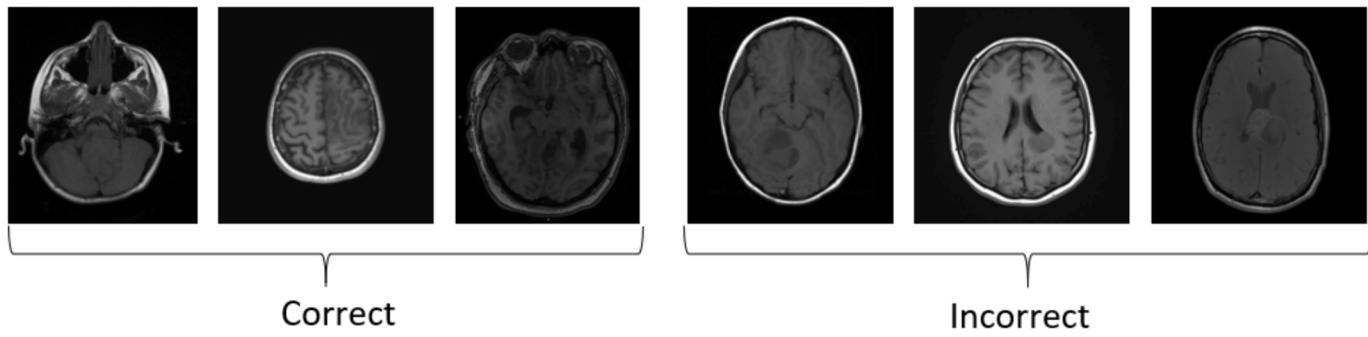


Fig. 7. The examples of both successful and unsuccessful classifications of brain MRI within the BM dataset.

ability to accurately detect and classify these less frequent but clinically significant tumor subtypes. Even though the precision appears high, the lower value of MCC suggests that the model might still be making some errors, particularly in correctly identifying the minority class (brain tumors) in the imbalanced dataset. Notably, the model exhibits reduced performance compared to its performance on the augmented datasets (discussed in the following section), highlighting the critical role of data augmentation in enhancing model robustness.

Fig. 7 highlights examples of both successful and unsuccessful classifications of brain MRI within the BM dataset.

4.6. The proposed CerebralNet model's performance on the ABM dataset

To introduce diversity into the BM dataset, a randomized augmentation strategy was employed. A pool of potential augmentation techniques, encompassing Contrast Limited Adaptive Histogram Equalization (CLAHE), Defocus, Downscale, Gaussian Blur, Horizontal Flip, Hue Saturation Value adjustments, Random Gamma, Transpose, Zoom Blur, and Pixel Dropout, was defined. For each brain MRI image within the BM dataset, a subset of randomly selected these augmentations (by using the proposed probabilistic augmentation selection strategy) was applied. The number of augmentations applied to each image was determined stochastically, introducing variability in the degree of augmentation. **Fig. 8** presents a visual examination of the CerebralNet model's training and validation performance on the augmented dataset.

Two key metrics are visualized in **Fig. 8**: (1) Epoch vs. Accuracy: This plot illustrates the evolution of both training and validation accuracy

across epochs. The increasing trend in accuracy for both datasets indicates effective learning, while the gap between training and validation accuracy provides insights into overfitting. (2) Epoch vs. Loss: This plot depicts the training and validation loss curves over epochs. The decreasing trend in both training and validation loss signifies that the model is effectively minimizing the error between its predictions and the true labels.

Fig. 9 presents the confusion matrix and associated performance metrics achieved from the model's evaluation on the test set of the ABM dataset.

The evaluation of the ABM dataset yielded impressive results, with the model demonstrating high accuracy (96.83 %), F1-score (96.82 %), recall (96.83 %), and precision (96.85 %). Notably, the model is correctly classifying almost all healthy brain scans with very less false positives. This crucial characteristic is of paramount importance in clinical settings, as minimizing false positives is critical to avoid unnecessary patient anxiety and invasive procedures. By accurately identifying all instances of brain tumors while effectively ruling out healthy cases, the model significantly contributes to reliable and safe clinical decision-making in brain tumor diagnosis.

Fig. 10 highlights examples of both successful and unsuccessful classifications of brain MRI within the ABM dataset.

Despite CerebralNet's overall high performance, the analysis of misclassified cases provides valuable insights into the inherent challenges of brain tumor detection and classification. Brain MRI images are complex, high-dimensional data with intricate patterns and subtle variations. Training a model to accurately recognize and differentiate between various tumor types within these complex images presents

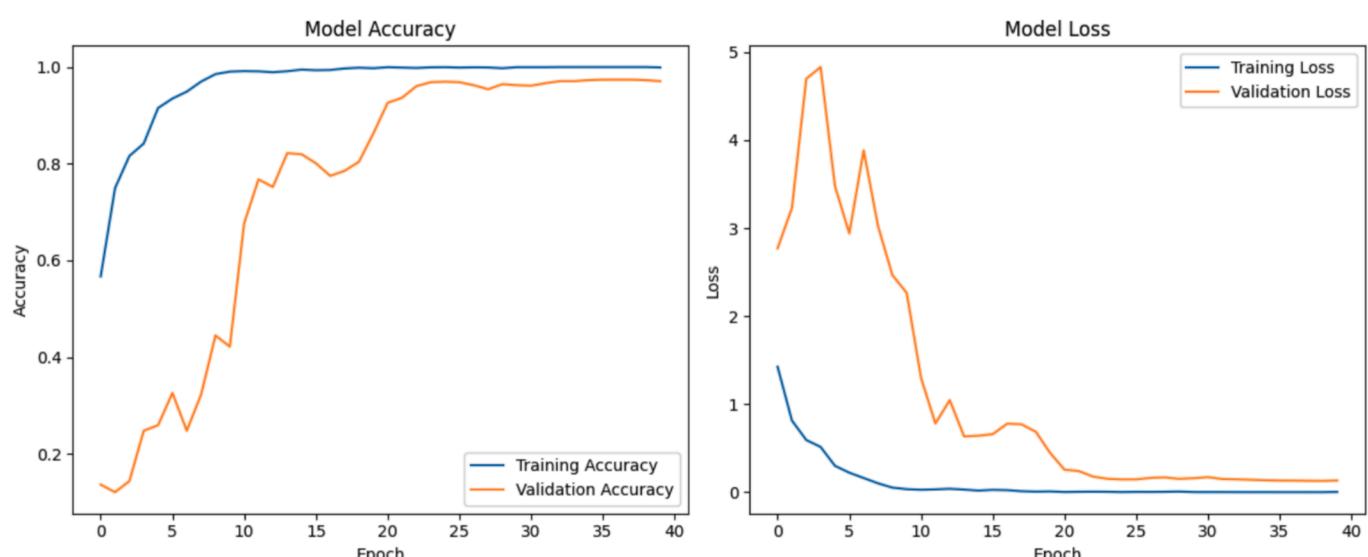
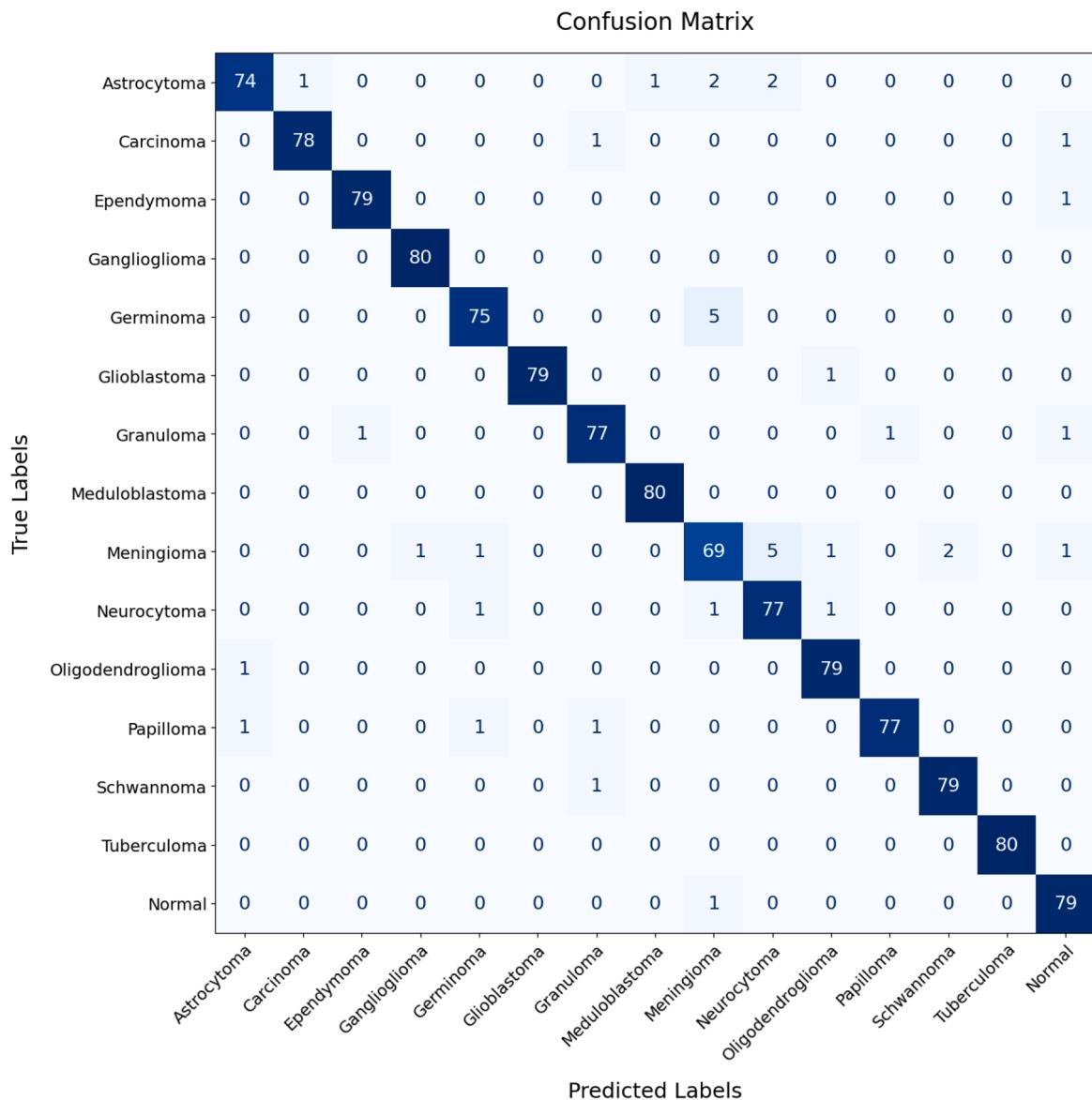


Fig. 8. The visual analysis of CerebralNet's training and validation performance on the ABM dataset.



Accuracy (%)	F1-Score (%)	Recall (%)	Precision (%)	MCC	Training Time (Sec)
96.83	96.82	96.83	96.85	96.6	1969.13

Fig. 9. The confusion matrix and associated performance metrics gained from CerebralNet's evaluation on the test set of the ABM dataset.

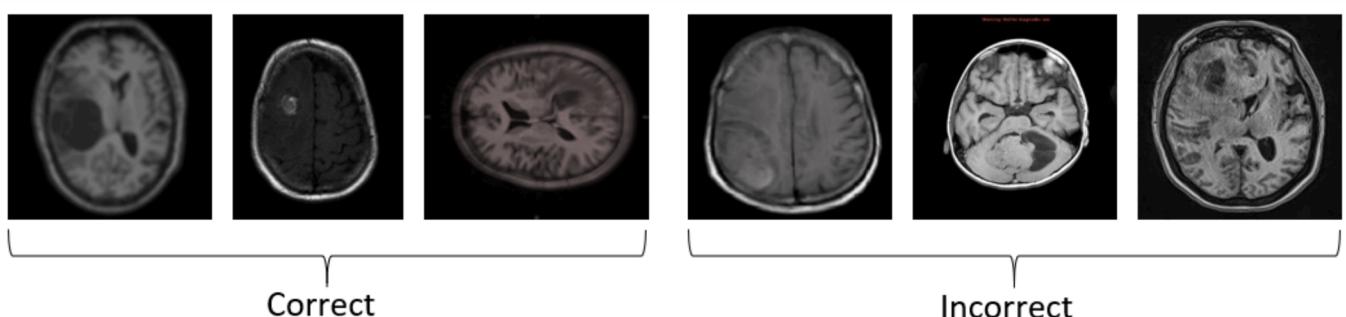


Fig. 10. The examples of both successful and unsuccessful classifications of brain MRI within the ABM dataset.

significant challenges. The model's performance can be impacted by the extraction of irrelevant or misleading features, leading to misclassifications. This highlights the crucial role of robust feature extraction in accurately identifying and characterizing brain tumors.

4.7. Comparison and evaluation of CerebralNet with transfer learning models

When dealing with imbalanced datasets, where the occurrence of the target class (in this case, the presence of a brain tumor) is relatively infrequent, precision emerges as a more critical performance metric than overall accuracy [42]. Precision, defined as the ratio of true positive predictions to the total number of predicted positive instances, directly measures the model's ability to correctly identify actual tumor cases among all instances predicted as positive. In the context of brain tumor detection and classification, a high-precision model is crucial to minimize the risk of false positives, which can lead to unnecessary anxiety, invasive procedures, and potentially harmful treatments for patients.

Accuracy, while seemingly a straightforward measure, can be misleading in imbalanced datasets. A model can achieve high accuracy simply by consistently predicting the majority class (in this case, the absence of a tumor), even if it performs poorly in detecting actual tumor cases. This emphasizes the importance of precision in clinical settings where misdiagnoses can have significant consequences.

To address the potential impact of class imbalance, this study included a baseline dataset (BM) with inherent class imbalances. The performance of the proposed CerebralNet model, along with other transfer learning models, was evaluated on both the BM dataset and an augmented dataset (ABM) to assess their performance under varying data distributions. Table 6 presents a detailed comparison of the models, including precision values, highlighting the proposed model's ability to accurately identify true positives while minimizing false positives, even in the presence of class imbalance.

From Table 6 it is clear that the proposed CerebralNet architecture achieved a significant improvement in precision on the ABM dataset (96.85 %) compared to its performance on the BM dataset (93.42 %). This improvement underscores the effectiveness of the data augmentation techniques employed in creating the ABM dataset and the model's enhanced ability to accurately identify true positive cases, minimizing the risk of misclassifying healthy subjects as having a brain tumor. This analysis reveals that while several transfer learning models, such as DenseNet121, DenseNet169, and InceptionV3, demonstrated competitive precision on the BM dataset, the proposed CerebralNet architecture consistently outperformed these models on the augmented ABM dataset. This improvement in precision on the augmented data highlights the model's enhanced robustness and generalizability, suggesting its

Table 6

The comparison of precision among the proposed CerebralNet model and other transfer learning models used in this study.

Transfer learning Model	Precision on the BM test set (%)	Precision on the ABM test set (%)
VGG16 [3,14]	92.83	92.4
VGG19 [9,10]	88.76	91.92
ResNet50 [16]	90.1	90.44
ResNet101 [9]	92.17	89.36
ResNet152 [10]	92.49	88.87
InceptionV3 [9]	92.67	94.49
MobileNetV2 [2,9]	89.78	95.52
MobileNetV3_Small [10]	95.13	88.46
MobileNetV3_Large [10]	92.42	89.77
DenseNet121 [15]	93.27	94.18
DenseNet169 [10]	93.19	95.2
DenseNet201 [15]	93.19	95.0
CerebralNet (Proposed Approach)	93.42	96.85

potential for improved performance in clinical scenarios where image quality and variability are common.

4.8. Understanding model predictions with LIME

Fig. 11 presents a series of LIME visualizations, showcasing the model's decision-making process for a representative sample from each class within the ABM dataset. For each sample, the figure displays the original brain MRI image alongside the highlight of the specific regions within the image that most significantly contribute to the model's prediction. The corresponding output provides visual confirmation of the model's focus on these critical regions. These visualizations offer valuable insights into the model's reasoning, revealing the key image features, such as tumor margins, internal textures, and surrounding anatomical structures, that drive the model's classification decisions. The use of the ABM dataset for these visualizations is particularly insightful as the diverse augmentations within the ABM dataset challenge the model to learn robust and generalizable features, providing a more comprehensive understanding of its decision-making process across a wider range of image variations.

5. Discussion

The primary motivation behind brain tumor detection and classification lies in the critical need for early and accurate diagnosis. Brain tumors can have a significant impact on an individual's health and quality of life, and timely intervention is crucial for successful treatment and improved patient outcomes. Traditional diagnostic methods, primarily reliant on visual interpretation of medical images like MRI scans by radiologists, face several limitations. These limitations include inter-observer variability, subjectivity in interpretation, and potential for human error, leading to inconsistencies in diagnosis and potential delays in treatment. Furthermore, manual analysis of large volumes of medical images is time-consuming and can be prone to fatigue, potentially impacting diagnostic accuracy.

To address the limitations of traditional diagnostic approaches and further enhance the performance of MobileNetV2, we incorporated the power of ASPP and Atrous Convolution blocks into the architecture. This strategic integration aimed to overcome the inherent limitations of standard convolutional networks in capturing multi-scale features within brain tumor images. By effectively capturing contextual information at multiple scales, the proposed model demonstrated significant improvements over the baseline MobileNetV2 architecture. The integration of ASPP and Atrous Convolutions enabled the model to more accurately detect tumors of varying sizes and shapes, leading to enhanced sensitivity in identifying subtle and complex tumor patterns. This enhancement significantly improved the model's ability to differentiate between tumor types and healthy brain tissue, ultimately leading to more reliable and clinically impactful brain tumor detection and classification.

Table 7 presents the performance evaluation results of the proposed model on both the BM and ABM datasets.

Table 7 comprehensively summarizes the proposed model's impressive performance on both the BM and ABM datasets. Notably, the model achieved high accuracy, F1-score, recall and precision across both datasets, demonstrating its robustness and generalizability. A key strength of the CerebralNet model lies in the fact that almost no instances of healthy brain scans being misclassified as containing tumors. This critical characteristic minimizes the risk of false positives, a crucial factor in clinical settings, as it avoids unnecessary investigations and treatments for healthy individuals. Furthermore, the model demonstrated high recall, effectively identifying the majority of tumor instances within the dataset, minimizing the risk of missed diagnoses. Additionally, the model achieved high MCC values on both datasets, demonstrating its ability to accurately classify all four possible outcomes (true positives, true negatives, false positives, and false negatives),

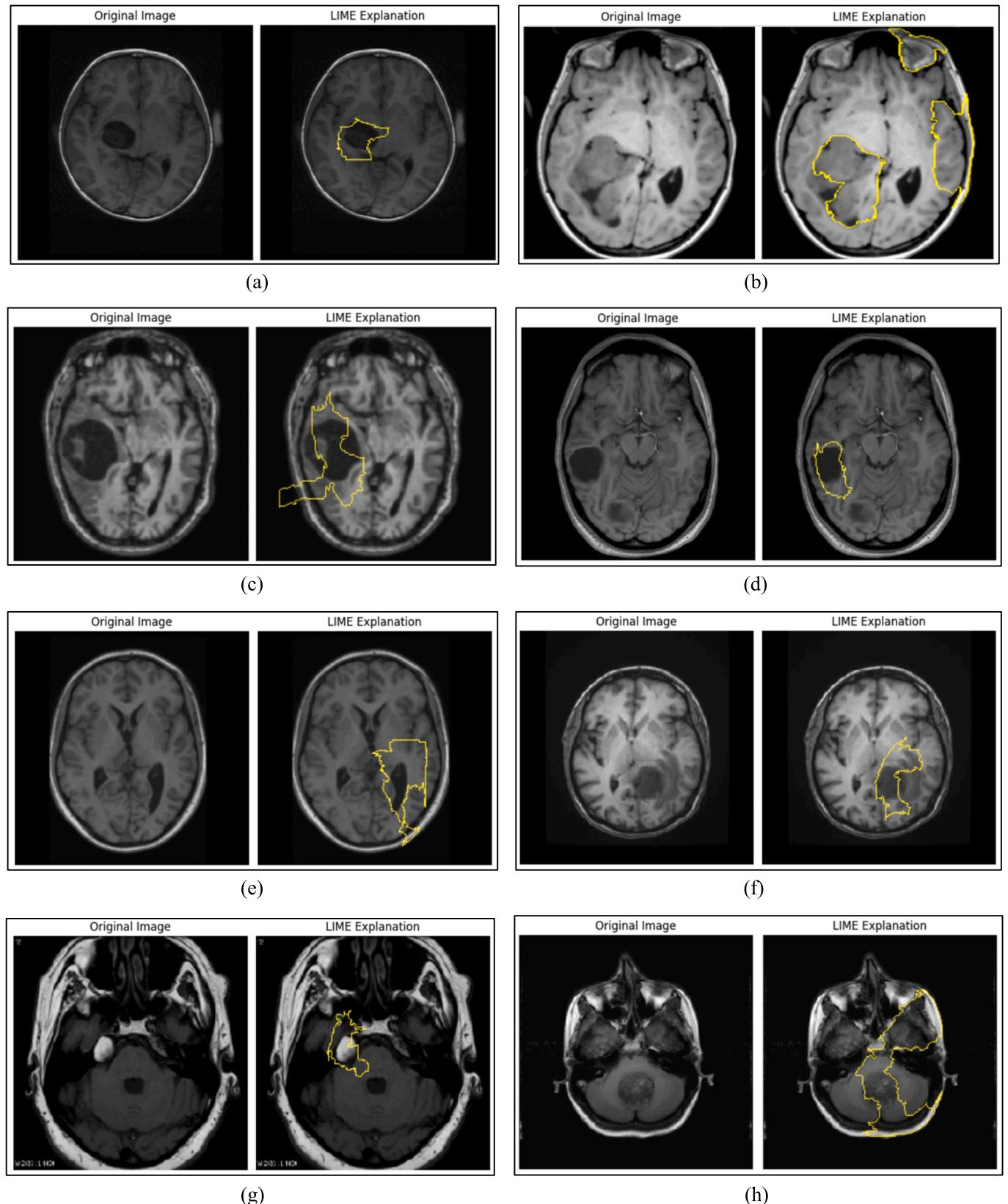


Fig. 11. LIME visualizations, showcasing CerebralNet's decision-making process for a representative sample from each class within the ABM dataset: (a) Astrocytoma, (b) Carcinoma, (c) Ependymoma, (d) Ganglioglioma, (e) Germinoma, (f) Glioblastoma, (g) Granuloma, (h) Medulloblastoma, (i) Meningioma, (j) Neurocytoma, (k) Oligodendrogioma, (l) Papilloma, (m) Schwannoma, (n) Tuberculoma, (o) Normal.

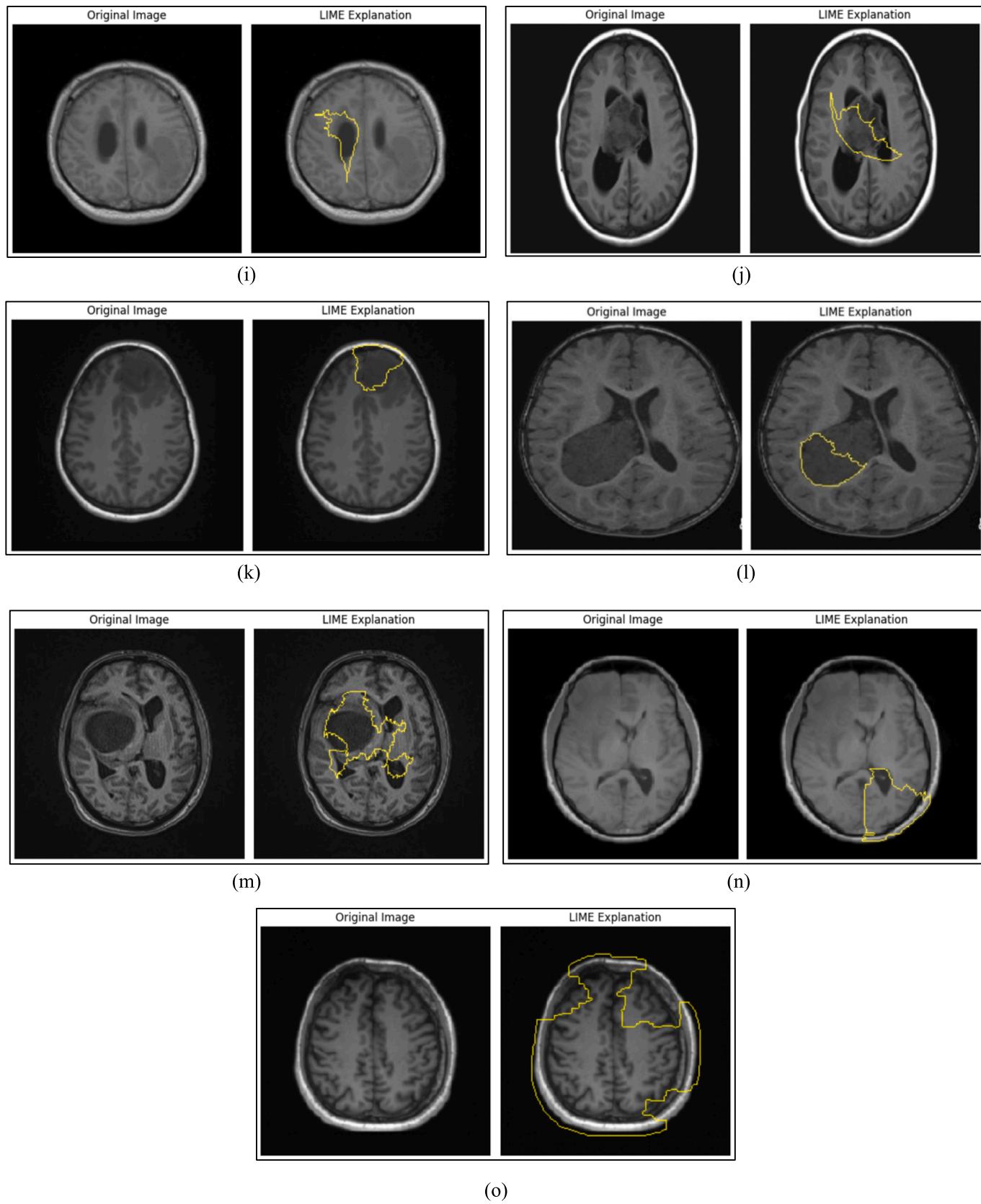


Fig. 11. (continued).

Table 7

The performance evaluation results of the proposed CerebralNet model on both the BM and ABM datasets.

Dataset	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	MCC (%)
BM	91.23	93.42	91.23	91.12	90.17
ABM	96.83	96.85	96.83	96.82	96.6

Table 8

The comparative analysis of the proposed CerebralNet model with other approaches used by researchers for brain tumor detection and classification.

Methodology Used	Accuracy on the BM Dataset (%)	Accuracy on the ABM Dataset (%)
The approach used in [16]	81.27	83.64
The approach used in [17]	89.36	90.13
The approach used in [18]	84.74	86.58
The approach used in [19]	82.43	83.91
The approach used in [20]	85.62	87.83
The approach used in [21]	89.14	91.79
The approach used in [22]	76.93	77.27
The approach used in [23]	93.33	89.42
The approach used in [24]	89.56	90.32
The approach used in [25]	90.45	88.52
The approach used in [26]	92.83	93.1
The approach used in [27]	91.14	91.87
The approach used in [28]	88.49	89.51
The approach used in [29]	90.73	92.67
The approach used in [30]	85.29	90.16
The approach used in [31]	87.94	88.61
The Proposed Approach (CerebralNet)	91.23	96.83

further solidifying its reliability and clinical relevance. These results collectively demonstrate the model's potential as a valuable tool for assisting clinicians in the accurate and efficient detection and classification of brain tumors.

A comparative analysis of the model's performance on the BM and ABM datasets reveals significant improvements across key metrics. Notably, on the ABM dataset, the model demonstrated superior performance in terms of accuracy (96.83 % compared to 91.23 % on the BM dataset), F1-score (96.82 % compared to 91.12 %), recall (96.83 % compared to 91.23 %), precision (96.85 % compared to 93.42 %), and MCC (96.59 % compared to 89.66 %). The overall performance gains observed on the ABM dataset underscore the effectiveness of the data augmentation strategies. These enhancements demonstrate the model's ability to effectively generalize and accurately classify brain tumors across a wider range of image variations, ultimately improving its robustness and clinical applicability.

Table 8 provides a comparative analysis of the proposed CerebralNet model with other approaches used by researchers for brain tumor detection and classification. The table presents the accuracy achieved by these different approaches on both the original (BM) and augmented (ABM) test datasets.

From **Table 8** it is clear that the proposed model demonstrated a significant performance improvement on the augmented ABM dataset compared to all other approaches listed, achieving an accuracy of 96.83 % on the ABM dataset, surpassing the highest accuracy of 93.33 % achieved by any other method on the BM dataset. On the original BM dataset, the proposed model achieved an accuracy of 91.23 %, demonstrating competitive performance compared to other approaches. However, the most significant improvement was observed in the augmented ABM dataset. This substantial performance gain on the ABM dataset underscores the effectiveness of the proposed model architecture and the data augmentation strategies employed. The integration of ASPP and Atrous Convolution blocks within the MobileNetV2 framework proved crucial in enhancing CerebralNet's capability to extract multi-scale features and capture intricate spatial relationships within brain

Table 9

The detailed analysis of the number of trainable parameters and computational demands for the CerebralNet model.

Dataset	Training time (sec)	Testing time (sec)	Trainable Parameters	Computational Resources	Floating Point Operations (FLOPs)
BM	1392.47	219	3,621,391	T4 GPU with 16 GB of GDDR6 memory	683 million
ABM	1969.13	276			

tumor images. Furthermore, the data augmentation techniques applied to create the ABM dataset, such as geometric transformations, noise injection, and intensity variations, significantly improved the model's robustness and generalization capabilities.

To support the real-world deployment of deep learning models for brain tumor detection and classification in clinical settings, it is crucial to consider a comprehensive set of metrics beyond mere accuracy. Key factors include the model's complexity, as reflected by the number of trainable parameters, as well as its computational demands, including training time and resource utilization. **Table 9** provides a detailed analysis of these crucial factors for the proposed model, offering a comprehensive assessment of its efficiency and practicality for clinical applications.

While our development environment utilized a T4 GPU with High RAM on Google Colab, which facilitated rapid prototyping and experimentation, we recognized that clinical settings have diverse resource constraints. The NVIDIA T4 GPU was considered well-suited for inference acceleration due to its Tensor Cores and efficient architecture, making it a viable option for deployment where timely results are crucial. We also ensured compatibility with NVIDIA GPUs, including the T4, which are prevalent in many clinical settings, by using TensorFlow frameworks and libraries that are optimized for NVIDIA hardware. We acknowledged the challenges of data heterogeneity in clinical settings and addressed this by evaluating the model's performance on the diverse ABM dataset to improve the model's robustness to variations in image acquisition. The analysis of training and testing times reported in **Table 9** reveals that while the ABM dataset, with its augmented images, improved model accuracy, it also increased training time. Training on the ABM dataset required approximately 41.41 % longer than training on the original BM dataset. This increase is expected as the augmented data significantly expands the dataset size, requiring the model to process a larger volume of information during training. However, the increase in training time is outweighed by the significant performance gains observed on the ABM dataset, demonstrating the effectiveness of data augmentation in enhancing model accuracy and robustness. We also quantified the model's computational complexity using FLOPs to ensure that inference could be performed within acceptable timeframes on hardware comparable to or exceeding the performance of a T4 GPU. A model with 683 million FLOPs represents a moderate computational demand.

The proposed CerebralNet model offers several potential advantages for clinical deployment in brain tumor detection. The 683 million FLOPs represent a manageable computational load for many modern clinical settings, particularly those with up-to-date GPU-equipped workstations. This suggests that real-time or near-real-time processing is achievable, facilitating timely clinical decision-making. Furthermore, the integration of LIME enhances the model's transparency, providing clinicians with insights into the factors driving its predictions. This explainability can foster trust and facilitate the adoption of CerebralNet in clinical practice. Finally, the study's emphasis on data augmentation, by enriching the dataset with a novel probabilistic approach, contributes to the model's robustness and generalizability, which are crucial for reliable performance across diverse patient populations and imaging conditions encountered in real-world clinical settings.

While the proposed CerebralNet model presents several advantages

for clinical deployment, there are also potential challenges that need to be addressed for successful implementation. A key practical consideration is hardware and infrastructure compatibility. CerebralNet, with its 683 million Floating Point Operations, necessitates a computational capacity that, while often available in modern radiology and pathology departments equipped with Graphics Processing Unit-equipped workstations, may require hardware upgrades in some clinical facilities to ensure efficient, real-time operation. Efficient integration of CerebralNet into existing clinical workflows, particularly with Picture Archiving and Communication Systems (PACS) and Electronic Health Record (EHR) systems, is also crucial, especially for the seamless transfer, processing, and presentation of LIME. Furthermore, scalability and deployment across multiple sites introduce operational complexities, including the need for robust infrastructure and ongoing support to distribute the model, manage updates, and maintain consistent performance and explanation quality across diverse clinical settings. Finally, regulatory compliance and certification, including adherence to standards set by bodies such as the Food and Drug Administration and Conformité Européenne marking, represent a significant step, requiring demonstration of CerebralNet's robustness and generalizability, particularly concerning its performance on data from varied sources.

6. Conclusions

This study presents the CerebralNet model, a new method to detect and classify brain tumors within MRI images. This model uses the power of deep learning to effectively identify subtle patterns and characteristics that distinguish healthy brain tissue from tumor regions. By analyzing intricate "deep features" extracted from the MRI images, the CerebralNet model aims to accurately classify brain scans as either tumor-type or healthy.

This research introduces a novel approach to brain tumor detection and classification that incorporates two key innovations: (1) Brain MRI Augmentation: To address the limitations of limited available brain MRI data, a comprehensive augmentation strategy was employed, encompassing techniques such as Contrast Limited Adaptive Histogram Equalization, Defocus, Downscale, Gaussian Blur, and various geometric transformations. This data augmentation strategy effectively simulates clinical imaging artifacts and variations, enhancing the model's robustness and generalizability. (2) CerebralNet Architecture: Building upon the efficient MobileNetV2 architecture, the proposed model incorporates ASPP and Atrous Convolutions to enhance feature extraction. ASPP allows the model to capture multi-scale contextual information, while Atrous Convolutions expand the receptive fields of the network, enabling the model to effectively detect tumors of varying sizes and accurately localize tumor boundaries.

To evaluate the proposed model's effectiveness in clinical settings, a detailed evaluation was directed utilizing two datasets: the original dataset (BM) and an augmented dataset (ABM). The BM dataset, characterized by inherent class imbalance, presented a challenging evaluation scenario. Despite this, the proposed model demonstrated robust performance on the BM dataset, highlighting its ability to effectively handle class imbalances. Furthermore, the inclusion of the ABM dataset, which incorporated a diverse set of image augmentations to simulate clinical imaging artifacts and variations, further enhanced the model's robustness and generalizability. The model demonstrated consistently high performance on both the BM and ABM datasets, suggesting its robustness to variations in image quality, a crucial aspect for reliable clinical applications. Clinical brain MRI scans can exhibit significant variability in image quality because of the factors such as patient movement, scanner variations, and image acquisition protocols. The model's ability to effectively handle these variations and maintain high performance, even in class imbalance, makes it a promising tool for reliable and accurate brain tumor detection and classification in clinical settings.

This study effectively utilized XAI (LIME) to gain insights into

CerebralNet's decision-making process. By utilizing LIME, we gained valuable insights into the model's decision-making process, identifying the specific brain MRI features that most significantly influence its predictions. This enhanced understanding of the model's behavior is crucial for building trust in its predictions and ensuring its reliable application in clinical settings.

While the proposed CerebralNet model demonstrates promising results in brain tumor detection and classification, several areas for future improvement can be identified. Firstly, the current model focuses on brain tumor detection and classification, determining the absence or presence of a tumor along with the tumor type. Future research could incorporate techniques for tumor localization, such as segmentation models, to pinpoint the exact location and extent of the tumor within the brain. This would provide valuable information for surgical planning and treatment guidance. Secondly, the current study primarily focuses on detecting a specific set of brain tumor types. Expanding the model's capabilities to encompass a wider range of tumor types, including rare and aggressive subtypes, would significantly enhance its clinical utility. Thirdly, this study primarily focused on utilizing T1 MRI images. However, incorporating additional imaging sequences, such as T2 and enhanced T1 images, will provide richer information about tissue characteristics and potentially improve the model's diagnostic accuracy. By using the complementary information provided by these different modalities, the model can potentially achieve more robust and accurate brain tumor detection and classification. Finally, the core architecture of the proposed model, combining the strengths of MobileNetV2 with advanced techniques like ASPP and Atrous Convolutions, has the potential to be adapted for other medical imaging tasks beyond brain tumor detection and classification, such as detecting abnormalities in other organs or diagnosing other neurological conditions. This would further demonstrate the versatility and generalizability of the proposed approach within the broader field of medical image analysis.

CRediT authorship contribution statement

Arth Agrawal: Writing – review & editing, Visualization, Validation, Software, Resources, Methodology, Investigation, Data curation. **Jyotismita Chaki:** Writing – review & editing, Writing – original draft, Validation, Supervision, Project administration, Methodology, Conceptualization.

Declaration of competing interest

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.

Data availability

Data will be made available on request.

References

- [1] Brain Tumor facts (<https://braintumor.org/brain-tumors/about-brain-tumors/brain-tumor-facts/>).
- [2] Brain Tumor Statistics (<https://www.aaroncohen-gadol.com/en/patients/brain-tumor/types/statistics>).
- [3] Brain Tumor Grades (<https://www.thebraintumourcharity.org/brain-tumour-diagnosis-treatment/how-brain-tumours-are-diagnosed/how-brain-tumours-are-graded/>).
- [4] C.K. Clayton, N. Loecher, R.T. Webster, Vaping behavior among adolescent and young adult cancer survivors: A scoping review, *Pediatr. Blood Cancer* 72 (1) (2025) e31367.
- [5] A. Almudayni, M. Alharbi, A. Chowdhury, J. Ince, F. Alablani, J.S. Minhas, A. Lechini-Visintini, E.M.L. Chung, Magnetic resonance imaging of the pulsing brain: a systematic review, *MAGMA* 36 (1) (2023) 3–14.
- [6] M. Ghorbian, S. Ghorbian, M. Ghobaei-arani, A comprehensive review on machine learning in brain tumor classification: taxonomy, challenges, and future trends, *Biomed. Signal Process. Control* 98 (2024) 106774.

- [7] Y. Pande, J. Chaki, Brain tumor detection across diverse MR images: An automated triple-module approach integrating reduced fused deep features and machine learning, *Results Eng.* 103832 (2024).
- [8] S. Maurya, S. Tiwari, M.C. Mothukuri, C.M. Tangeda, R.N.S. Nandigam, D. C. Addagiri, A review on recent developments in cancer detection using machine learning and deep learning models, *Biomed. Signal Process. Control* 80 (2023) 104398.
- [9] N. Remzan, K. Tahiry, A. Farchi, Advancing brain tumor classification accuracy through deep learning: harnessing radimagenet pre-trained convolutional neural networks, ensemble learning, and machine learning classifiers on MRI brain images, *Multimed. Tools Appl.* (2024) 1–29.
- [10] X. Lian, Y. Pang, J. Han, J. Pan, Cascaded hierarchical atrous spatial pyramid pooling module for semantic segmentation, *Pattern Recognit.* 110 (2021) 107622.
- [11] M. Khalid, S. Deivasigamani, V.S., S. Rajendran, An efficient colorectal cancer detection network using atrous convolution with coordinate attention transformer and histopathological images, *Sci. Rep.* 14 (1) (2024) 19109.
- [12] V. Hassija, V. Chamola, A. Mahapatra, A. Singal, D. Goel, K. Huang, S. Scardapane, I. Spinelli, M. Mahmud, A. Hussain, Interpreting black-box models: a review on explainable artificial intelligence, *Cogn. Comput.* 16 (1) (2024) 45–74.
- [13] K.N. Rao, O.I. Khalef, V. Krishnasree, A.S. Kumar, D.M. Alsekait, S.S. Priyanka, A. S. Alattas, D.S. AbdElminaam, An efficient brain tumor detection and classification using pre-trained convolutional neural network models, *Heliyon* 10 (17) (2024).
- [14] A.S. Musallam, A.S. Sherif, M.K. Hussein, A new convolutional neural network architecture for automatic detection of brain tumors in magnetic resonance imaging images, *IEEE Access* 10 (2022) 2775–2782.
- [15] A. Priya, V. Vasudevan, Brain tumor classification and detection via hybrid alexnet-gru based on deep learning, *Biomed. Signal Process. Control* 89 (2024) 105716.
- [16] A.H. Khan, S. Abbas, M.A. Khan, U. Farooq, W.A. Khan, S.Y. Siddiqui, A. Ahmad, Intelligent model for brain tumor identification using deep learning, *Appl. Comput. Intell. Soft Comput.* 2022 (1) (2022) 8104054.
- [17] M.S.I. Khan, A. Rahman, T. Debnath, M.R. Karim, M.K. Nasir, S.S. Band, A. Mosavi, I. Dehzangi, Accurate brain tumor detection using deep convolutional neural network, *Comput. Struct. Biotechnol. J.* 20 (2022) 4733–4745.
- [18] R. Vankdothu, M.A. Hameed, H. Fatima, A brain tumor identification and classification using deep learning based on CNN-LSTM method, *Comput. Electr. Eng.* 101 (2022) 107960.
- [19] M.I. Mahmud, M. Mamun, A. Abdelgawad, A deep analysis of brain tumor detection from mr images using deep learning networks, *Algorithms* 16 (4) (2023) 176.
- [20] H.M. Rai, K. Chatterjee, 2D MRI image analysis and brain tumor detection using deep learning CNN model LeU-Net, *Multimed. Tools Appl.* 80 (28) (2021) 36111–36141.
- [21] M.A. Khan, A. Khan, M. Alhaisoni, A. Alqahtani, S. Alsabai, M. Alharbi, N.A. Malik, R. Damaševičius, Multimodal brain tumor detection and classification using deep saliency map and improved dragonfly optimization algorithm, *Int. J. Imaging Syst. Technol.* 33 (2) (2023) 572–587.
- [22] S. Anantharajan, S. Gunasekaran, T. Subramanian, R. Venkatesh, MRI brain tumor detection using deep learning and machine learning approaches, *Meas.: Sens.* 31 (2024) 101026.
- [23] S.K. Mathivanan, S. Sonaimuthu, S. Murugesan, H. Rajadurai, B.D. Shivahare, M. A. Shah, Employing deep learning and transfer learning for accurate brain tumor detection, *Sci. Rep.* 14 (1) (2024) 7232.
- [24] E. Hassan, H. Ghadiri, Advancing brain tumor classification: A robust framework using EfficientNetV2 transfer learning and statistical analysis, *Comput. Biol. Med.* 185 (2025) 109542.
- [25] A.B. Subba, A.K. Sunaniya, Computationally optimized brain tumor classification using attention based GoogLeNet-style CNN, *Expert Syst. Appl.* 260 (2025) 125443.
- [26] S. Patil, D. Kirange, Ensemble of deep learning models for brain tumor detection, *Procedia Comput. Sci.* 218 (2023) 2468–2479.
- [27] S. Asif, M. Zhao, F. Tang, Y. Zhu, An enhanced deep learning method for multi-class brain tumor classification using deep transfer learning, *Multimed. Tools Appl.* 82 (2023) 31709–31736.
- [28] A.K. Sharma, A. Nandal, A. Dhaka, L. Zhou, A. Alhudhaif, F. Alenezi, K. Polat, Brain tumor classification using the modified ResNet50 model based on transfer learning, *Biomed. Signal Process. Control* 86 (2023) 105299.
- [29] S. Maqsood, R. Damaševičius, R. Maskeliūnas, Multi-modal brain tumor detection using deep neural network and multiclass SVM, *Medicina* 58 (8) (2022) 1090.
- [30] R. Appiah, V. Pulletikurthi, H.A. Esquivel-Puentes, C. Cabrera, N.I. Hasan, S. Dharmarthine, L.J. Gomez, L. Castillo, Brain tumor detection using proper orthogonal decomposition integrated with deep learning networks, *Comput. Methods Programs Biomed.* 250 (2024) 108167.
- [31] R. Tanone, L.H. Li, S. Saifullah, ViT-CB: Integrating hybrid Vision Transformer and CatBoost to enhanced brain tumor detection with SHAP, *Biomed. Signal Process. Control* 100 (2025) 107027.
- [32] Brain Tumor MRI Images 44 Classes (<https://www.kaggle.com/datasets/fernando2rad/brain-tumor-mri-images-44c>).
- [33] C.P. Jetlin, PyQDCNN: Pyramid QDCNN for multi-level brain tumor classification using MRI image, *Biomed. Signal Process. Control* 100 (2025) 107042.
- [34] R. Gayathiri, S. Santhanam, C-SAN: Convolutional stacked autoencoder network for brain tumor detection using MRI, *Biomed. Signal Process. Control* 99 (2025) 106816.
- [35] M.J. Adamu, H.B. Kauwua, L. Qiang, C.O. Nyatenga, A. Younis, M. Fahad, S. Dauya, Efficient and accurate brain tumor classification using hybrid mobileNetV2-support vector machine for magnetic resonance imaging diagnostics in neoplasms, *Brain Sci.* 14 (12) (2024) 1178.
- [36] M.Z. Asghar, F.R. Albogamy, M.S. Al-Rakhami, J. Asghar, M.K. Rahmat, M. M. Alam, A. Lajis, H.M. Nasir, Facial mask detection using depthwise separable convolutional neural network model during COVID-19 pandemic, *Front. Public Health* 10 (2022) 855254.
- [37] A. Halder, D. Dey, Atrous convolution aided integrated framework for lung nodule segmentation and classification, *Biomed. Signal Process. Control* 82 (2023) 104527.
- [38] G.M. Foody, Challenges in the real world use of classification accuracy metrics: from recall and precision to the Matthews correlation coefficient, *PLoS One* 18 (10) (2023) e0291908.
- [39] M.K. Panda, A. Sharma, V. Bajpai, B.N. Subudhi, V. Thangaraj, V. Jakhetiya, Encoder and decoder network with ResNet-50 and global average feature pooling for local change detection, *Comput. Vis. Image Underst.* 222 (2022) 103501.
- [40] N. Bacanin, C. Stoean, M. Zivkovic, M. Rakic, R. Strulak-Wójcikiewicz, R. Stoean, On the benefits of using metaheuristics in the hyperparameter tuning of deep learning models for energy load forecasting, *Energies* 16 (3) (2023) 1434.
- [41] M. Bentoumi, M. Daoud, M. Benouali, A. Taleb Ahmed, Improvement of emotion recognition from facial images using deep learning and early stopping cross validation, *Multimed. Tools Appl.* 81 (21) (2022) 29887–29917.
- [42] R. İncir, F. Bozkurt, Improving brain tumor classification with combined convolutional neural networks and transfer learning, *Knowl.-Based Syst.* 111981 (2024).