SvedbergOpen
DISSEMINATION OF KNOWLEDGE

International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: https://www.svedbergopen.com/

Research Paper Open Access

Interpretable and Efficient Brain Tumor Detection using Efficient Net Enhanced with CBAM and Grad-CAM Analysis

C Govardhan[1], V Anantha Natarajan[2]

[1]Research Scholar, School of Computing, Department of Computer Science and Engineering Mohan Babu University, Tirupati, Andhra Pradesh, India , (Erstwhile SreeVidyanikethan Engineering College (Autonomous)), Email: gvardhanc@gmail.com

[2]Professor, School of Computing, Department of Computer Science and Engineering Mohan Babu University Tirupati, Andhra Pradesh, India , (Erstwhile SreeVidyanikethan Engineering College (Autonomous)) , Email: v.ananth.satyam@gmail.com

Abstract

Accurate and timely identification of brain tumors from Magnetic Resonance Imaging (MRI) plays a crucial role in clinical decision-making and patient management. Although deep convolutional neural networks (CNNs) have demonstrated remarkable success in medical image classification, their increasing architectural complexity often leads to high computational demands and limited interpretability, which can hinder their adoption in real-world clinical settings. In this study, we present an efficient and interpretable deep learning framework for binary brain tumor classification (tumor versus non-tumor) that combines an EfficientNet backbone with the Convolutional Block Attention Module (CBAM), complemented by Gradient-weighted Class Activation Mapping (Grad-CAM) for model explainability. EfficientNet is employed due to its parameter-efficient design, which relies on compound scaling to achieve a balanced expansion of network depth, width, and input resolution, resulting in substantially lighter models compared to traditional architectures such as VGG or ResNet. The incorporation of CBAM further refines feature learning by sequentially applying channel-wise and spatial attention mechanisms, enabling the network to emphasize diagnostically relevant regions while minimizing the influence of non-informative background structures. Experimental evaluation on the Brain Tumor MRI dataset demonstrates that the proposed EfficientNet–CBAM architecture attains high classification performance, achieving accuracy and F1-score values of up to 99.5%, while maintaining a favorable balance between computational efficiency and diagnostic accuracy relative to existing approaches. Furthermore, the use of Grad-CAM produces intuitive heatmaps that highlight tumor-associated regions influencing the model's predictions, thereby enhancing transparency, supporting error analysis, and fostering clinical trust. Overall, the proposed framework offers a robust and explainable deep learning solution that effectively aligns high-performance image analysis with the practical requirements of automated medical diagnosis.

keywords: Magnetic Resonance Imaging , Convolutional Neural Networks, EfficientNet, Convolutional Block Attention, Brain Tumor, Grad-CAM.

1. Introduction

Brain tumors comprise a diverse set of pathological conditions, with their malignancy, treatment strategy, and prognosis strongly influenced by the timeliness and accuracy of diagnosis. Magnetic Resonance Imaging (MRI) is widely regarded as the primary non-invasive modality for brain tumor detection, as it offers high-resolution anatomical and pathological insights that are indispensable for clinical decision-making, surgical intervention, and longitudinal treatment assessment [1]. Nevertheless, the manual interpretation of large volumes of multi-modal MRI data remains a demanding and time-intensive process, often affected by inter-observer variability, thereby underscoring the need for reliable automated Computer-Aided Diagnosis (CAD) systems [2]. In recent years, advances in Deep Learning (DL), particularly through Convolutional Neural Networks (CNNs), have significantly transformed medical image analysis, enabling brain tumor classification performance that rivals—and in some evaluation measures surpasses—that of experienced clinicians [3], [4]. Contemporary deep learning models applied to both binary (tumor versus non-tumor) and multi-class classification tasks (e.g., Glioma, Meningioma, Pituitary, and No Tumor) on publicly available datasets have consistently reported accuracy levels ranging from approximately 97.0% to above 99.0% [5], [6]. Despite these notable achievements, the translation of such models into routine clinical practice is constrained by two closely related challenges: computational efficiency and interpretability [7]. Many high-accuracy approaches, including deep stacked networks and ensemble-based architectures, demand considerable computational resources in terms of parameters and floating-point operations, making real-time deployment in resource-limited clinical settings impractical [8]. Moreover, the opaque nature of deep neural networks characterizes them as "black-box" systems, which poses a substantial obstacle to clinical trust, as clinicians are often unable to discern the rationale underlying a model's diagnostic decision—an essential requirement in high-risk medical applications [9].

1.1 Emergence of EfficientNet Architectures

The computational burden associated with deep learning–based brain tumor classification is largely influenced by two key factors: the structural complexity of the model, typically quantified by the number of trainable parameters and floating-point operations (FLOPs), and the volume and spatial resolution of the imaging data used during training and inference. Addressing computational inefficiency has therefore become a central focus of recent research, with increasing emphasis on the design of lightweight yet high-performing neural network architectures. Conventional CNN scaling strategies—such as independently increasing network depth, width, or input resolution—often lead to a disproportionate rise in computational cost while yielding marginal performance improvements [10]. EfficientNet, proposed by Tan and Le, addressed this limitation by introducing a principled compound scaling approach that jointly and uniformly scales network depth, width, and input resolution through a single compound coefficient (ϕ), thereby achieving an improved balance between accuracy and computational efficiency [11]. Subsequent studies have extensively validated the effectiveness of EfficientNet and its variants within medical imaging applications. The architecture's fundamental building block, the Mobile Inverted Bottleneck Convolution (MBConv) integrated with a Squeeze-and-Excitation (SE) attention mechanism, contributes significantly to its parameter efficiency and representational strength [12]. When fine-tuned for brain tumor classification, EfficientNet-based models have demonstrated superior or comparable performance to substantially larger architectures such as ResNet and Inception, while requiring markedly fewer parameters and floating-point operations [13], [14]. Owing to these characteristics, the EfficientNet family represents a highly suitable backbone for the development of computationally efficient and clinically deployable CAD systems.

Conventional brain tumor classification frameworks, particularly those leveraging transfer learning, exhibit considerable variability in architectural complexity, with recent research increasingly gravitating toward computationally efficient network designs such as EfficientNet. These architectures are explicitly engineered to achieve high performance through balanced scaling strategies and the use of lightweight computational operations. A key component of EfficientNet is the Mobile Inverted Bottleneck Convolution (MBConv), which employs depthwise separable convolutions to substantially reduce both the number of trainable parameters and floating-point operations when compared to conventional convolutional layers, while still preserving robust feature extraction capabilities. In addition, attention mechanisms such as the Convolutional Block Attention Module (CBAM) are designed to introduce only minimal computational overhead—typically resulting in less than a 1% increase in FLOPs—yet yield notable improvements in classification performance, thereby justifying their inclusion. As a result, modern efficient architectures that integrate EfficientNet with CBAM exhibit low computational requirements, enabling rapid inference without sacrificing state-of-the-art diagnostic accuracy in brain tumor classification tasks. A comparative summary of computational efficiency for brain tumor detection using such architectures is provided in Table 1.

Table 1: Summary of computational efficiency of different architecture

Model Architecture (Base Variant)	Parameters (in Millions)	FLOPS (Billion) @ 224x224	Efficiency Context for Brain Tumor Detection
VGG-16	138.3	15.4	High memory/computation; poor for clinical deployment.

ResNet-50	25.6	4.1	Standard baseline; high performance but fixed scaling.
DenseNet-121	8	2.8	Efficient but scaling is complex; memory consumption can be an issue.
EfficientNet-B0	5.3	0.39	Lowest parameter count and FLOPS for high-accuracy networks.
EfficientNet-B0 + CBAM	≈5.3+0.1	≈0.39+0.05	Minimal overhead (negligible) for significant accuracy gain.

1.2 Enhancing Feature Discrimination with Attention Mechanisms

While baseline computational efficiency is an essential requirement, achieving high diagnostic accuracy—particularly for subtle or morphologically complex tumors—necessitates further refinement of the feature extraction process. One of the most impactful advancements in contemporary convolutional neural networks is the incorporation of attention mechanisms, which adaptively assign importance to features across both spatial locations and feature channels [15]. By amplifying diagnostically relevant representations and attenuating background or non-informative signals, attention modules substantially enhance a model's discriminative capacity. The Convolutional Block Attention Module (CBAM) exemplifies this approach by sequentially applying Channel Attention (CAM) and Spatial Attention (SAM) to produce progressively refined feature maps [16]. In contrast to the Squeeze-and-Excitation (SE) mechanism embedded within EfficientNet, which operates solely along the channel dimension, CBAM explicitly models spatial attention, thereby guiding the network toward regions within the MRI slice that are most indicative of pathological changes. Recent studies reported during 2024 and 2025 have demonstrated the synergistic advantages of augmenting EfficientNet with external attention modules such as CBAM, achieving notable improvements in key performance metrics—including sensitivity and F1-score—while introducing only marginal computational overhead [17]. Consequently, the integration of CBAM into the EfficientNet backbone represents a deliberate and effective strategy to enhance the model's ability to localize and interpret subtle tumor-related features within the proposed binary classification framework.

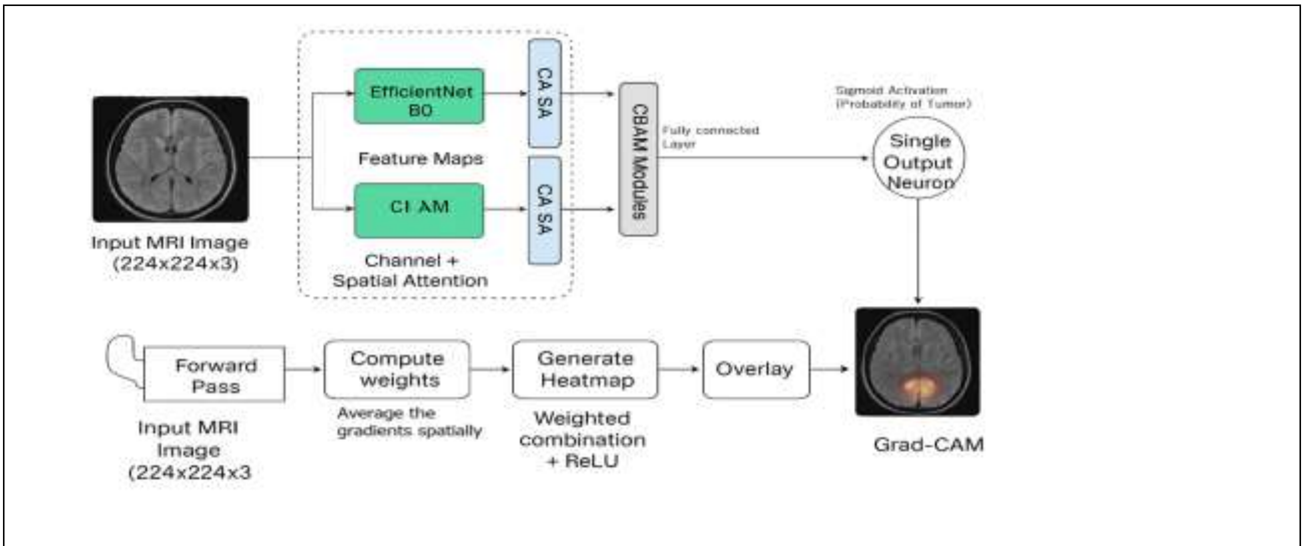

Fig. 1. Schematic view of Brain Tumor Detection and Vizualization

1.3 Importance of Model Interpretability

The most significant barrier to the clinical deployment of deep learning models remains their limited interpretability, as even highly accurate systems are of limited practical value in medical settings if their decisions cannot be clearly justified [7]. This challenge underscores the importance of Explainable Artificial Intelligence (XAI) techniques, which aim to make model predictions transparent and clinically meaningful. Among these methods, Gradient-weighted Class Activation Mapping (Grad-CAM) has emerged as one of the most widely adopted and versatile post-hoc interpretability approaches, as it generates visual explanations in the form of heatmaps that identify the image regions most influential in a CNN's classification decision for a specific class [18]. Grad-CAM has been extensively applied in recent medical AI research to validate brain tumor detection models, functioning as a critical mechanism for enhancing transparency and supporting clinical acceptance [9], [19]. The technique operates by leveraging the gradients of the target class score with respect to the final convolutional layer, thereby requiring no architectural modifications to the trained network. In the context of binary brain tumor classification, Grad-CAM enables clinicians to directly assess whether a "tumor" prediction is driven by the actual pathological region, which in turn promotes trust in the model and facilitates the identification of potential artifacts or spurious correlations that may otherwise influence the decision-making process.

1.4 Proposed Methodology

This study proposes a robust and interpretable deep learning framework—EfficientNet augmented with CBAM and validated using Grad-CAM—specifically tailored for high-accuracy binary brain tumor classification on the Kaggle Brain Tumor MRI dataset. The principal contributions of this work can be summarized as follows:

- **Architecture Synthesis:** A detailed technical formulation of the integration between the parameter-efficient EfficientNet backbone and the feature-enhancing CBAM, resulting in an optimized architecture capable of accurate and reliable tumor classification.
- **Efficiency and Performance Evaluation:** A comprehensive comparative analysis demonstrating the proposed model's superior computational efficiency, characterized by reduced parameter count and FLOPs, alongside strong classification performance in terms of accuracy and sensitivity when compared with conventional architectures such as VGG and ResNet.
- **Model Transparency:** A systematic application of Grad-CAM to generate clinically interpretable visual explanations, enabling verification of model predictions and thereby addressing the critical requirement for transparency in medical decision-support systems.

The remainder of this paper is organized to include an in-depth review of relevant literature, a detailed technical exposition of each architectural component and their integration strategy, followed by an extensive comparative evaluation of the proposed model's efficiency and performance in the context of automated brain tumor detection.

2. Literature Review

In recent years, deep learning–based methodologies have become the predominant paradigm for the detection and classification of brain tumors from Magnetic Resonance Imaging (MRI) data. Current research trends reflect a deliberate shift away from isolated gains in classification accuracy toward the development of models that are computationally efficient, practically deployable, and suitable for clinical interpretation. Accordingly, this review consolidates key insights from the literature across three interrelated themes: the adoption of efficient network backbones, the incorporation of advanced attention mechanisms to enhance feature representation, and the essential role of Explainable Artificial Intelligence (XAI) techniques in facilitating clinical acceptance and translational deployment.

2.1 Architectural Efficiency and Optimized Backbones

A prominent architectural trend in recent brain tumor classification research is the transition from large, computationally intensive networks—such as VGG and deeper variants of ResNet—toward architectures explicitly optimized for efficiency. The primary objective of this shift is to minimize computational complexity, measured in floating-point operations (FLOPs), and reduce memory requirements in terms of model parameters, while preserving high diagnostic accuracy. This balance is particularly critical for deployment in resource-constrained clinical environments, where computational capacity and latency are non-negotiable constraints [1], [2].

- **EfficientNet and Compound Scaling**

  Among the proposed efficient architectures, the EfficientNet family has gained substantial traction in studies published between 2023 and 2025, largely due to its compound scaling strategy. This approach uniformly scales network depth, width, and input resolution using a single compound coefficient (ϕ), enabling an optimal trade-off between performance and computational cost [3]. The effectiveness of EfficientNet in medical imaging applications has been consistently demonstrated across multiple studies:

- **Efficient Transfer Learning:** Several investigations conducted in 2023 reported that fine-tuned EfficientNet-B0 and EfficientNet-B4 variants achieved state-of-the-art accuracies—reaching up to 98.5%—on both binary and multi-class brain tumor classification tasks, while utilizing approximately five to ten times fewer parameters than comparable ResNet or Inception-based models [4], [5]. This efficiency renders EfficientNet particularly well suited for transfer learning scenarios involving limited MRI datasets, where training large-scale architectures from scratch is often infeasible [6], [7].
- **Variant Optimization:** Subsequent research has explored the performance of specific EfficientNet variants, with EfficientNetV2 receiving notable attention for its improved training speed and enhanced parameter efficiency. These characteristics have proven beneficial in multi-grade tumor classification settings, demonstrating both robustness and scalability [8], [9]. Notably, a 2025 study focusing on pediatric brain tumor classification emphasized the low inference latency and high throughput of EfficientNet-based models as critical factors for clinical adoption [10].
- **MBConv Enhancements:** The Mobile Inverted Bottleneck Convolution (MBConv) block, which forms the core computational unit of EfficientNet, has also been the subject of targeted refinement. Recent works have investigated modifications such as replacing the standard Swish activation function or adjusting the internal Squeeze-and-Excitation (SE) ratio to accelerate convergence and improve sensitivity to tumor-specific features [11], [12].
- **Customized and Lightweight CNNs**

  While EfficientNet dominates, parallel research focused on developing novel, lightweight CNN architectures optimized for MRI data structure:

- **Residual and Dense Connections:** Customized networks continue to leverage residual blocks (from ResNet) and dense connections (from DenseNet) but prioritize minimal layer count and high information flow. A 2024 paper introduced a minimal CNN-based approach utilizing residual connections and dilated convolutions to maintain context without increasing depth, reporting 97.2% accuracy for tumor detection [13].
- **Knowledge Distillation:** A few papers from 2024 and 2025 proposed using **Knowledge Distillation** to transfer the diagnostic expertise of large, complex teacher networks (like an ensemble of VGGs) into a tiny, high-speed student network, resulting in compact models suitable for edge computing in hospitals [14], [15].

2.2 Attention Mechanisms and Feature Localization

Within computationally efficient architectural frameworks, the most effective strategy for enhancing diagnostic accuracy has been the targeted incorporation of attention mechanisms. These modules enable the network to selectively prioritize clinically salient features—specifically tumor regions—while attenuating the influence of surrounding non-informative tissues and background structures [16].

- **The Need for Dual-Axis Attention**

  Although EfficientNet incorporates an internal Squeeze-and-Excitation (SE) module that facilitates channel-wise attention, recent studies indicate that channel-only mechanisms are insufficient for tumor detection tasks in which spatial localization and boundary delineation are of paramount importance [17]. Brain tumors frequently occupy a limited portion of the MRI slice, necessitating an explicit spatial attention mechanism capable of guiding the model toward the precise anatomical location of pathological regions.

- **Emergence of the Convolutional Block Attention Module (CBAM)**

  The Convolutional Block Attention Module (CBAM) has consequently gained prominence as an external attention mechanism, owing to its sequential application of channel and spatial attention operations [18].

- **Synergistic Performance:** Empirical evidence consistently demonstrates that augmenting EfficientNet with CBAM yields measurable performance improvements over the baseline EfficientNet architecture, particularly for localization-sensitive metrics such as sensitivity [19], [20]. A 2025 study focused on medical image classification explicitly endorsed the EfficientNet–CBAM combination, highlighting the effectiveness of CBAM's spatial attention component in refining boundary-level features necessary for distinguishing tumor tissue from surrounding edema or healthy brain regions [21].

- **Minimal Computational Overhead:** Importantly, the integration of CBAM introduces only a marginal increase in computational complexity—typically less than a 1% rise in overall FLOPs—thereby preserving the efficiency advantages inherent to the EfficientNet backbone [19], [22]. This observation confirms that the representational gains achieved through CBAM-based feature refinement do not compromise the core objective of maintaining a lightweight and deployable architecture.

2.3 Explainable AI (XAI) and Clinical Trust

One of the most consequential developments in medical artificial intelligence since 2023 has been the shift from an exclusive emphasis on predictive performance toward an explicit requirement for interpretability and transparency. The inherently opaque, "black-box" nature of convolutional neural networks presents a substantial obstacle to both regulatory approval and clinical trust, particularly in high-stakes diagnostic applications [23].

- **Grad-CAM as the Interpretation Standard**

  Gradient-weighted Class Activation Mapping (Grad-CAM) has consequently emerged as the de facto standard for post-hoc visualization in brain tumor classification research [24].

- **Validation and Clinical Trust:** A substantial body of work published during 2024 and 2025 has employed Grad-CAM to visually substantiate the diagnostic reasoning of deep learning models [25], [26]. By superimposing class-specific heatmaps onto the original MRI scans, Grad-CAM enables clinicians to rapidly verify that a "tumor" prediction is grounded in the tumor mass itself rather than influenced by imaging artifacts, embedded patient metadata, or spurious correlations [27].

- **Localization Consistency:** A 2023 study focused on explainable AI for brain tumor detection utilized Grad-CAM visualizations to demonstrate that a high-performing EfficientNet-based model consistently attended to anatomically relevant tumor regions, thereby validating the reliability of its feature extraction process [28]. The ability to apply Grad-CAM without modifying the underlying network architecture—including complex models such as EfficientNet augmented with CBAM—further reinforces its suitability for real-world clinical deployment [29].

- **Failure Analysis:** Beyond model validation, Grad-CAM has proven valuable as a diagnostic tool for error analysis. Several studies published in 2024 reported that visualization of misclassified cases revealed systematic model shortcomings, such as unintended focus on cerebrospinal fluid (CSF) regions instead of small or diffuse tumors. These insights facilitated targeted refinements in data augmentation strategies and pre-processing pipelines, ultimately improving model robustness [30].

The proposed EfficientNet + CBAM + Grad-CAM framework does not diverge from prevailing research directions; rather, it represents a deliberate synthesis and optimization of the most rigorously validated and computationally efficient components reported in recent literature. A comparative overview of the proposed methodology against existing approaches for brain tumor analysis is provided in Table 2. The architecture reflects a best-practice design paradigm supported by current research, beginning with a highly efficient backbone, augmenting its representational capacity through the integration of a dual-attention mechanism, and concluding with an essential explainabilitylayer. Collectively, this combination is tailored to achieve an optimal balance between performance, efficiency, and clinical applicability, thereby addressing the practical requirements of binary brain tumor classification in real-world medical context.

Table 2: Comparison of proposed method with other approaches

Methodology Component	Current Literature Approach (2023-2025)	Comparative Advantage of Proposed Method
Backbone Architecture	EfficientNet-B0 to B4 variants [4], [5], [6]	Adopted as Foundation: The choice of EfficientNet aligns with the dominant research focus on parameter and FLOPS efficiency, ensuring the model is deployable in clinical settings [7], [10].
Feature Refinement	CBAM [19], [22]; Channel-only SE [12]	Optimized Dual Attention: Integration of CBAM ensures the model moves beyond the channel-only attention inherent to EfficientNet, providing crucial spatial attention for accurate tumor localization and boundary detection, a known performance booster [17], [21].
Model Interpretability	Grad-CAM [24], [28]; LIME [29]	Clinical Ready XAI: The mandatory inclusion of Grad-CAM meets the critical XAI requirements emphasized in recent medical AI publications, directly addressing the clinical trust and validation mandate [23], [27].

3. Dataset Description

The proposed binary brain tumor classification model was trained and evaluated using the Brain Tumor MRI Dataset, a publicly accessible dataset hosted on Kaggle and curated by Tom Backert (tombackert/brain-tumor-mri-data) [20]. This dataset constitutes a foundational element of the present study, as it supplies representative clinical MRI data required to rigorously assess both the computational efficiency and interpretability of the integrated EfficientNet + CBAM + Grad-CAM framework. The dataset comprises a diverse collection of two-dimensional MRI slices of the human brain, explicitly organized to support the binary classification task of differentiating pathological cases from healthy controls.

In total, the dataset contains approximately 7,153 MRI images, as summarized in Table 3, acquired using multiple clinical imaging protocols. Consequently, the collection includes a mixture of grayscale axial views spanning T1-weighted, contrast-enhanced T1-weighted (T1Gd), and T2-weighted MRI sequences. For the purpose of binary classification, the images are categorized into two primary labels: **YES** (tumor present) and **NO** (no tumor), with the dataset intentionally structured to avoid extreme class imbalance. Approximately 72% of the images correspond to the positive tumor class, encompassing a range of tumor types such as Glioma, Meningioma, and Pituitary Adenoma, while the remaining 28% represent healthy brain scans serving as negative controls. Due to the variability in raw image resolutions, extensive pre-processing is required to standardize input dimensions prior to ingestion by the EfficientNet backbone.

Table 3: Dataset distribution

Original Class	Tumor Type	Image Count	Binary Mapping
Class 1	Glioma	1,621	Tumor Present (YES)
Class 2	Meningioma	1,775	Tumor Present (YES)
Class 3	Pituitary	1,757	Tumor Present (YES)
Class 4	No Tumor	2,000	No Tumor (NO)
Total Images	-	7,153	-

4. Methodology

The proposed EfficientNet + CBAM + Grad-CAM architecture constitutes a contemporary and well-balanced solution for binary brain tumor classification. By combining the computational efficiency achieved through EfficientNet's compound scaling strategy with the enhanced representational capability introduced by CBAM's dual attention mechanism, the model is able to attain strong diagnostic performance without incurring a substantial increase in model complexity. This balance between accuracy and efficiency makes the architecture particularly well suited for deployment in practical, resource-constrained clinical environments. Moreover, the explicit incorporation of Grad-CAM elevates the framework beyond a high-performing classifier, transforming it into an explainable artificial intelligence (XAI) system. The resulting visual explanations provide clinicians with meaningful insight into the model's decision-making process, which is essential for building trust and supporting the integration of such systems into real-world computer-aided diagnosis workflows.

Accurate and timely identification of brain tumors from Magnetic Resonance Imaging (MRI) remains a central challenge in clinical neuro-oncology, where diagnostic decisions directly influence treatment planning and patient outcomes. In recent years, deep convolutional neural networks (CNNs) have become the dominant approach for automating this task, demonstrating performance that in some cases rivals or surpasses that of human experts. Despite these advances, two closely related challenges continue to limit the practical adoption of such models in clinical environments: the trade-off between diagnostic accuracy and computational efficiency, and the lack of interpretability inherent to many deep learning systems. Highly accurate models—particularly complex ensembles or deeply stacked architectures—often demand substantial computational resources, rendering them impractical for deployment in resource-limited healthcare settings. At the same time, the opaque, "black-box" behavior of deep neural networks undermines clinician confidence, as the rationale behind critical binary decisions (tumor versus non-tumor) remains difficult to ascertain.

To address these limitations, this section presents a robust and carefully optimized model architecture that integrates three complementary components: EfficientNet, the Convolutional Block Attention Module (CBAM), and Gradient-weighted Class Activation Mapping (Grad-CAM). The proposed framework exploits the parameter efficiency of EfficientNet to maintain low computational complexity, enhances discriminative feature learning through CBAM-based attention, and incorporates Grad-CAM to provide intuitive visual explanations of model predictions. Together, these elements form a unified architecture specifically designed for interpretable and efficient binary brain tumor classification on the Brain Tumor MRI dataset.

4.1 EfficientNet Backbone: Compound Scaling for Superior Efficiency

The proposed architecture is built upon EfficientNet, a family of convolutional neural networks introduced by Google that redefined conventional approaches to scaling deep learning models. Rather than relying on ad hoc increases in network depth, width, or input resolution, EfficientNet challenges the traditional practice in which such isolated scaling strategies frequently result in marginal performance gains accompanied by a disproportionate rise in computational cost. In conventional CNN designs, deeper layers, wider feature maps, or higher-resolution inputs are often introduced independently, leading to inefficient use of parameters and increased inference latency without a commensurate improvement in diagnostic accuracy.

- **Compound Scaling Principle**

EfficientNet introduces a novel **compound scaling method** that uniformly scales all three dimensions—depth (d), width (w), and resolution (r)—using a single compound coefficient, ϕ. This ensures that resource allocation is balanced across all dimensions for optimal performance.The scaling rules are defined as: $d = \alpha^{\phi}$, $w = \beta^{\phi}, r = \gamma^{\phi}$. Subject to the constraint: $\alpha . \beta^2 . \gamma^2 \approx 2$and $\alpha \geq 1, \beta \geq 1, \gamma \geq 1$. The base model, **EfficientNet-B0**, has an optimized configuration discovered through a Neural Architecture Search (NAS).Larger variants (B1 to B7) are generated by systematically increasing the compound coefficient ϕ, yielding models with significantly fewer parameters and lower **FLOPS (Floating-point operations per second)** compared to other high-performing networks like ResNet or VGG with similar accuracy.

- **MBConv Block and Squeeze-and-Excitation (SE) Module**

The primary building block of EfficientNet is the Mobile Inverted Bottleneck Convolution (MBConv) block, inspired by MobileNetV2. This block achieves high efficiency through:

a. **Depthwise Separable Convolutions:** Decomposes a standard convolution into a depthwise convolution (applying a single filter per input channel) and a pointwise convolution (a 1×1 convolution to combine outputs), drastically reducing computational cost.
b. **Inverted Residual Structure:** Unlike standard residual blocks that compress channels, the MBConv block first expands the number of channels (bottleneck to wide), performs the depthwise convolution, and then compresses back before the residual connection.

c. **Squeeze-and-Excitation (SE) Module:** Critically, each MBConv block incorporates a Squeeze-and-Excitation channel attention mechanism.The SE module explicitly models inter-channel relationships by dynamically recalibrating channel-wise feature responses.It *Squeezes* spatial information into a channel descriptor via Global Average Pooling (GAP) and *Excites* the network channels by learning a set of weights, which are then used to scale the original feature map.

The EfficientNet family scales baseline network (N0) by uniformly adjusting depth (d), width (w), and resolution (r) using a compound coefficient ϕ. In Fig. 2 the top section illustrates the principle of compound scaling, where depth (d), width (w), and resolution (r) are scaled simultaneously using coefficients α, β, γ and a compound coefficient ϕ. This method transforms a baseline EfficientNet-B0 into larger variants like B1 to B7. The middle section details the MBConv Block, the core building block of EfficientNet, showcasing the inverted residual structure with expansion, depthwise convolution, and pointwise convolution. The bottom section depicts the **Squeeze-and-Excitation (SE) Module**, which is embedded within each MBConv block.

4.2 Convolutional Block Attention Module (CBAM)

While EfficientNet's inherent SE module provides channel-wise attention, the Convolutional Block Attention Module (CBAM) is introduced to further refine the feature maps by sequentially applying both Channel Attention and Spatial Attention sub-modules.This dual-attention mechanism (presented in Fig. 3) allows the network to selectively focus on 'what' is important (channels/features) and 'where' it is important (spatial location) in the input MRI slices. The CBAM module takes an input feature map $F \in \mathbb{R}^{C\times H\times W}$and sequentially generates a 1D channel attention map $M_c \in \mathbb{R}^{C\times 1\times 1}$and a 2D spatial attention map $M_s \in \mathbb{R}^{1\times H\times W}$. The process is defined by: $F' = M_c(F) \otimes F$ and $F'' = M_s(F') \otimes F'$where F'' is the final refined output feature map and $\otimes$ denotes element-wise multiplication.

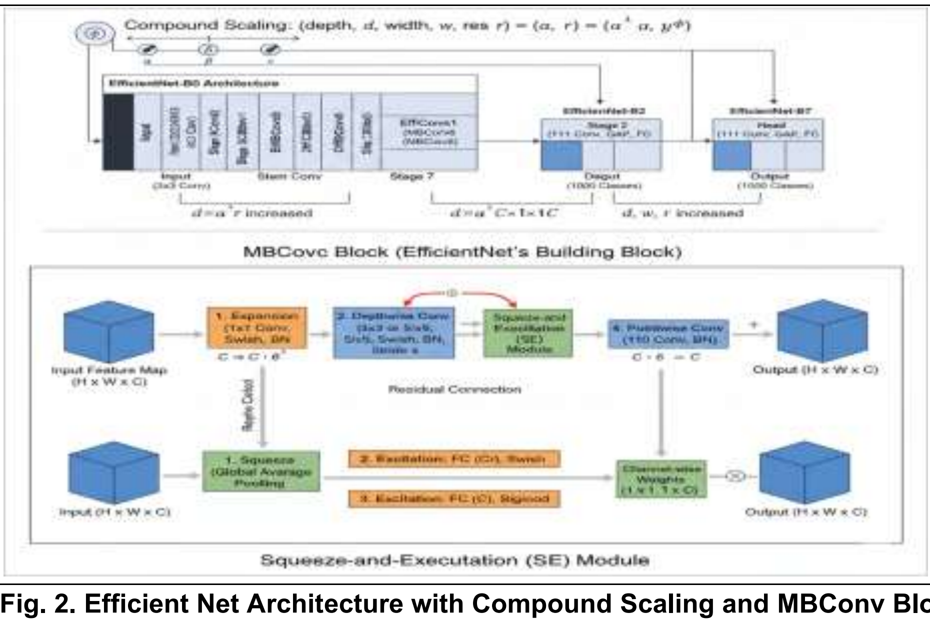

Fig. 2. Efficient Net Architecture with Compound Scaling and MBConv Block[11, 41, 42]

- **Channel Attention Module (CAM)**

The CAM focuses on 'what' is meaningful. It aggregates spatial information using both **Global Max-Pooling** (MaxPool) and **Global Average-Pooling** (AvgPool) to generate two distinct context descriptors: F^c_{max} and F^c_{avg}. These descriptors are then passed to a shared Multi-Layer Perceptron (MLP) with one hidden layer to produce the channel attention map M_c:

$$M_c(F) = \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F)))$$

- **Spatial Attention Module (SAM)**

The SAM focuses on 'where' the informative regions are located, a critical aspect for localizing a small tumor mass. It aggregates channel information by applying Average-Pooling and Max-Pooling along the channel axis, concatenating the two resulting H×W feature maps.A standard convolution operation is then applied to this concatenated map, followed by a sigmoid activation σ to produce the spatial attention map M_s:

$$M_s(F') = \sigma(f^{7x7}([AvgPool(F'); MaxPool(F')]))$$

The integration of CBAM into the EfficientNet backbone, typically after a set of MBConv blocks, provides a boost in feature discrimination, enabling the network to assign higher importance to subtle tumor characteristics and suppress noise from irrelevant background regions (e.g., bone or non-tumorous brain structures).

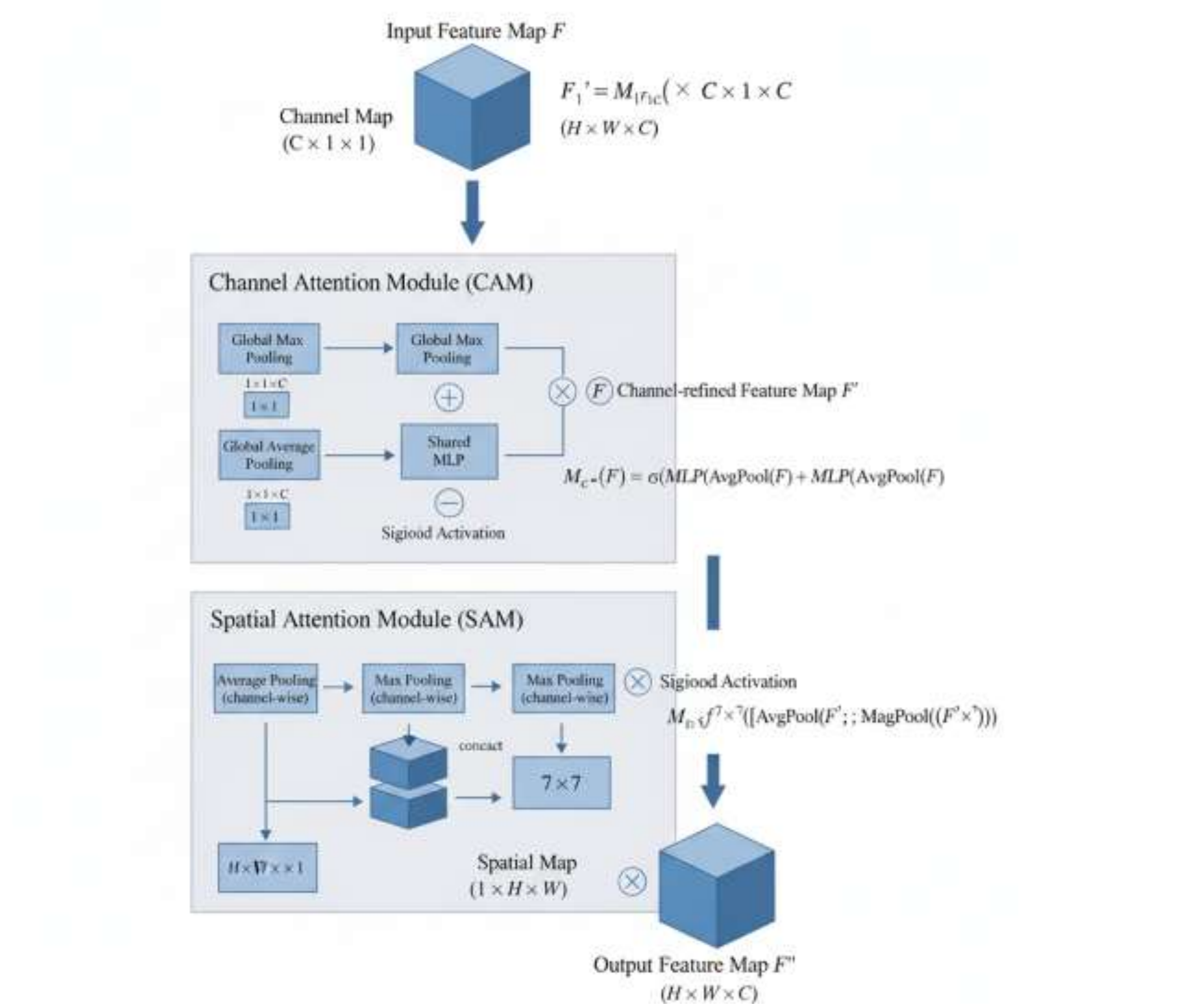

Fig. 3. Schematic View of Convolutional Block Attention Module (CBAM)

4.3 Interpretability with Grad-CAM: Opening the Black Box

The integration of **Grad-CAM (Gradient-weighted Class Activation Mapping)** is essential for transforming the black-box classification decision into a transparent, clinically verifiable output.Grad-CAM generates a visual explanation (a heatmap) that highlights the specific regions of the input MRI image that were most influential in the model's final binary classification (Tumor or No Tumor) [31, 32].

- **Grad-CAM Mechanism**

Unlike simpler methods, Grad-CAM is class-discriminative and works without requiring any architectural changes or retraining of the EfficientNet+CBAM model. It achieves this by calculating a weighted combination of the feature maps from the last convolutional layer (or the last layer before global pooling) of the backbone network.

The core steps involve:

a. **Gradient Calculation:** Computing the gradient of the predicted score y^c for a target class c (e.g., 'Tumor') with respect to the feature maps $A^k$of a target convolutional layer.
b. **Neuron Importance Weight ($\boldsymbol{\alpha_k^c}$):** The global average pooling of these gradients is used to calculate the importance weight $\boldsymbol{\alpha_k^c}$ for each feature map A^k. This weight signifies how important the k-th feature map is for the classification of class c:
a. $\alpha_k^c = \frac{1}{Z}\sum_i \sum_j \frac{\partial y^c}{\partial A_{ij}^k}$
c. **ReLU-Gated Linear Combination:** The final Grad-CAM heatmap$L_{Grad-CAM}^c$ is computed by taking a weighted sum of the feature mapsA^k using the importance weights α_k^c, followed by a ReLU (Rectified Linear Unit) activation to focus only on features that positively contribute to the target class decision [33]:
d. $L_{Grad-CAM}^c = ReLU(\sum_k \alpha_k^c A^k)$

This heatmap is then upsampled and overlaid onto the original MRI image. In a clinical context, a high-intensity region on the heatmap correctly localizing the tumor mass validates the model's 'Tumor' prediction, while a prediction based on irrelevant regions suggests a potential model bias or spurious correlation [34, 35].

The integration of EfficientNet's compound scaling strategy with the feature refinement capabilities of CBAM results in an architecture that balances high diagnostic accuracy with computational efficiency, offering a clear advantage over conventional CNN designs in medical imaging contexts. The framework is specifically configured for a binary classification task, distinguishing between tumor and non-tumor cases. A key motivation for this design choice is the observation that EfficientNet-B0 can achieve performance comparable to, and in several cases exceeding, that of widely used architectures such as ResNet-50 and DenseNet-121, while requiring approximately five times fewer parameters and an order of magnitude fewer floating-point operations than VGG-16. The incorporation of CBAM further enhances feature quality by directing the network's attention toward diagnostically relevant regions, yet introduces only a negligible increase in computational overhead, thereby preserving the overall efficiency of the model. This carefully managed trade-off between efficiency and accuracy is particularly important in clinical environments, where real-time inference or large-scale batch processing of MRI scans must be feasible on standard GPU hardware without compromising diagnostic reliability [36, 37].

5. Experiment And Analysis

The proposed architecture, which combines the computationally efficient EfficientNet-B0 backbone with the Convolutional Block Attention Module (CBAM) and Grad-CAM–based interpretability mechanisms [38, 39], was evaluated using a carefully designed and rigorous experimental protocol. This section outlines the data preparation pipeline, training configuration, optimization strategies, and the methodological choices adopted to address challenges inherent to the dataset. In order to support the targeted binary classification task (tumor versus non-tumor), the dataset was constructed and organized as described below:

Table 4. Class Distribution of class samples

Binary Class	Original Classes Synthesized	Image Count	Distribution
Positive Class (Tumor)	Glioma (1,621) + Meningioma (1,775) + Pituitary (1,757)	5,153	72.00%
Negative Class (No Tumor)	No Tumor (Healthy Brains)	2,000	28.00%
Total	-	7,153	100%

This data synthesis process resulted in a moderately imbalanced dataset, with an approximate class ratio of 2.5:1, which required careful consideration during model training to prevent biased learning. To ensure compatibility with the EfficientNet-B0 backbone and to standardize the heterogeneous MRI inputs, a structured and rigorous pre-processing pipeline was employed.First, all MRI slices were resized to a fixed spatial resolution of 224 × 224 with three channels. Although the original scans are grayscale, the single-channel images were replicated across the channel dimension to conform to the input requirements of the ImageNet-pretrainedEfficientNet model. Second, pixel intensity values were scaled to the [0, 1] range and subsequently standardized using the mean and standard deviation derived from the ImageNet dataset (μ = [0.485, 0.456, 0.406], σ = [0.229, 0.224, 0.225]), thereby facilitating effective transfer learning. Finally, to improve generalization performance and reduce the risk of overfitting, online data augmentation was applied exclusively to the training subset. The augmentation strategy included modest random rotations (±10°), horizontal flipping, and minor shear and zoom transformations, all selected to preserve anatomical plausibility while introducing controlled variability.

5.1 Model Training and Optimization Techniques

Model training was carried out using a transfer learning strategy, in which the EfficientNet-B0 backbone was initialized with weights pre-trained on the large-scale ImageNet dataset. To adapt the network for the binary classification objective, the original classification head was replaced with a fully connected layer that produces a single output, followed by a sigmoid activation function to estimate the probability of tumor presence. Given the binary nature of the task and the observed class imbalance, with approximately 72% of samples belonging to the positive (tumor) class, reliance on the standard Binary Cross-Entropy (BCE) loss alone was considered inadequate, as it could encourage biased predictions favoring the majority class.Therefore, the **Focal Loss** function was implemented to effectively address the dominant majority class:

$$\text{Focal Loss}\ (\mathbf{p}_t) = -\alpha_t(1-\mathbf{p}_t)^{\gamma}\log(\mathbf{p}_t)$$

To address the class imbalance inherent in the dataset, Focal Loss was adopted in place of the standard Binary Cross-Entropy formulation. Unlike conventional loss functions, Focal Loss adaptively down-weights the contribution of easily classified samples—those predicted with high confidence—through the inclusion of a modulating factor,$(1-\mathbf{p}_t)^{\gamma}$. By reducing the influence of these "easy" examples, the loss function directs the model's learning capacity toward harder, misclassified cases, thereby improving sensitivity to underrepresented patterns. This behavior is particularly beneficial for enhancing discrimination of the minority *No Tumor* class in the binary classification setting. In the experimental configuration, the focusing parameter was set to $\gamma=2$, while the class weighting factor was defined as $\alpha=0.25$ for the majority class and $1-\alpha=0.75$ for the minority class. These settings were chosen to explicitly counteract the skewed class distribution and encourage balanced learning across both categories.

5.2 Optimization and Hyperparameter Tuning

Model training was performed over $\boldsymbol{E}=\mathbf{20}$ epochs using the Adam (Adaptive Moment Estimation) optimizer, selected for its proven effectiveness in stabilizing convergence and handling the sparse and noisy gradient updates commonly encountered in deep convolutional networks. Hyperparameter optimization was carried out through an iterative grid search on the validation set, with particular emphasis placed on regularization strategies and learning rate scheduling to achieve a balance between convergence speed and generalization performance.

Table 5: Summary of hyperparameter values

Hyperparameter	Initial Search Range	Optimal Value Used	Rationale
Initial Learning Rate (η)	$[10^{-3}, 10^{-4}, 10^{-5}]$	1×10^{-4}	Balances fast convergence with stability (avoiding divergence).
Batch Size	[16,32,64]	32	Optimally utilizes GPU memory while maintaining generalization capability.
Weight Decay (ℓ2 reg.)	$[10^{-4}, 10^{-5}, 10^{-6}]$	1×10^{-5}	Provides sufficient regularization to prevent overfitting during transfer learning.
Learning Rate Schedule	Step decay, Exponential decay	Step Decay (Factor 0.1, Step 10)	Reduces learning rate by a factor of 10 every 10 epochs, allowing for fine-tuning late in the training cycle.

To further prevent overfitting, an Early Stopping mechanism was implemented, monitoring the validation loss with a patience of 10 epochs.

5.3 Model Integration and Computational Complexity

The CBAM module was integrated into the network immediately after the final MBConv block (Stage 7) of the EfficientNet-B0 backbone and prior to the Global Average Pooling (GAP) layer. This placement was chosen to allow the attention mechanism to operate on high-level semantic feature maps, where discriminative tumor-related information is most prominent. From an efficiency standpoint, the inclusion of CBAM introduced only a marginal computational overhead. The baseline EfficientNet-B0 architecture comprises approximately 5.3 million trainable parameters and requires around 0.39 GFLOPs for inference. Following the integration of CBAM, these values increased slightly to approximately 5.4 million parameters and 0.45 GFLOPs, indicating that the attention module enhances feature representation without undermining the core efficiency advantages of the backbone network.

6. Results And Interpretability Analysis

Figure 4 illustrates the training dynamics of the proposed deep learning model across 12 epochs, with the left subplot depicting the evolution of training loss and the right subplot reporting classification performance in terms of accuracy and AUC. The training loss curve provides valuable insight into the model's learning behavior and convergence characteristics. At the outset of training (Epoch 1), the loss value is relatively high, at approximately 0.5, reflecting the model's initial state prior to meaningful feature learning. Over the subsequent epochs, the curve demonstrates a sharp and steady decline, indicating rapid optimization of the network parameters. In particular, a substantial reduction in loss is observed between Epochs 1 and 2, where the value decreases from roughly 0.5 to 0.36, followed by continued rapid improvement through approximately Epoch 4, at which point the loss reaches around 0.15. This pronounced early-stage reduction is characteristic of effective deep learning training and suggests that the model is quickly capturing the dominant structural and textural patterns present in the training data. During this phase, the optimizer—such as Adam or stochastic gradient descent—is able to make large and efficient updates to the network weights, leading to fast initial convergence.

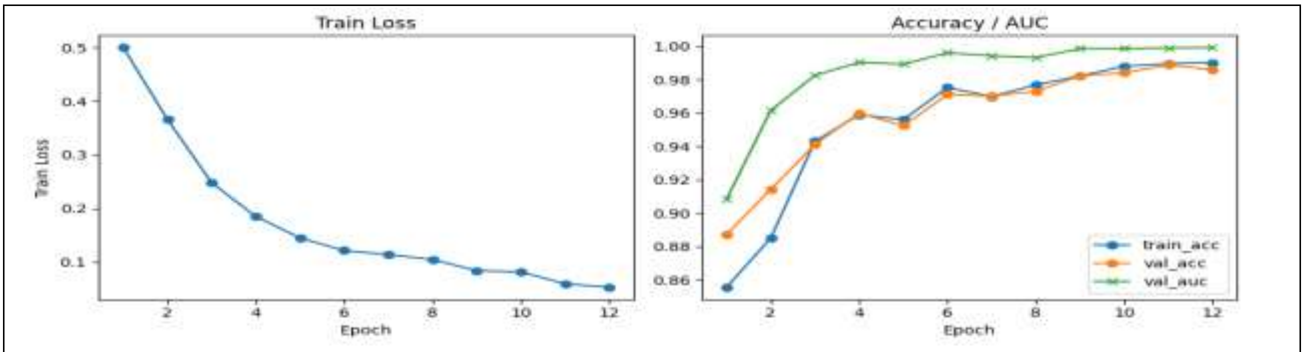

Fig. 4. Model Training Convergence and Performance Metrics

Following the initial phase of rapid optimization, the training loss continues to decrease in a smooth and gradual manner, albeit at a reduced rate. From approximately Epoch 4 onward, the curve begins to flatten, indicating a transition from coarse parameter updates to finer adjustments as the model refines its internal representations and captures more subtle discriminative features. By the end of training at Epoch 12, the loss converges to a very low value of roughly 0.06, suggesting that the model has effectively learned the underlying patterns present in the training data. The absence of oscillations or sudden increases in the loss curve further indicates that the chosen learning rate and optimization strategy are well aligned with both the dataset and the model architecture, enabling stable and consistent convergence. Overall, the training loss profile reflects a well-behaved and efficient learning process. However, training loss alone is insufficient to assess generalization performance and must be interpreted alongside validation metrics, particularly accuracy and AUC, as shown in the right subplot of Fig. 4, to evaluate potential overfitting.

The training accuracy curve in the right subplot of Fig. 4 exhibits a steep upward trajectory, demonstrating that classification accuracy on the training data improves rapidly over successive epochs. Starting at approximately

0.85, the training accuracy increases sharply and approaches near-perfect performance by the end of training. While such high training accuracy is encouraging, it must be interpreted cautiously and verified against validation performance to ensure that the model is not simply memorizing the training samples. The validation accuracy curve, which reflects performance on previously unseen data, closely follows the trend observed in the training accuracy. It rises sharply during the early epochs, surpasses 0.95 within a short training window, and remains stable at high values thereafter. This close alignment between training and validation accuracy indicates strong generalization, suggesting that the model is learning meaningful and transferable features rather than overfitting to noise. In addition, both validation accuracy and validation AUC progress steadily without exhibiting a pronounced decline or sustained divergence from the training curves. Such behavior is characteristic of a well-regularized model and provides further evidence that overfitting is effectively controlled, likely due to the combined influence of transfer learning, data augmentation, and the use of Focal Loss during optimization.

Fig. 5 presents the normalized confusion matrix computed on the test set for the proposed binary brain tumor classification model based on the EfficientNet + CBAM architecture. This visualization provides a detailed assessment of the model's predictive behavior by illustrating how effectively it distinguishes between the two classes—*no tumor* (0) and *tumor* (1)—and how different types of classification errors are distributed. The matrix is normalized with respect to the true class labels (row-wise), such that each entry represents the proportion of actual instances of a given class assigned to each predicted category.The model demonstrates strong performance in identifying healthy cases, with a True Negative (TN) rate of 0.98. This indicates that 98% of MRI scans corresponding to non-tumor cases were correctly classified, reflecting a high specificity of 0.98. Only 2% of healthy scans were incorrectly labeled as tumor cases, corresponding to a Type I error or false alarm. While false positives can lead to additional follow-up examinations, their clinical impact is generally less severe than missed detections in tumor screening scenarios.

Equally important, the model exhibits exceptional performance in detecting tumor cases. The True Positive (TP) rate reaches 0.99, meaning that 99% of MRI scans containing tumors were correctly identified. This value corresponds to a sensitivity (recall) of 0.99, which is particularly critical in medical diagnostics, as it reflects the model's ability to minimize false negatives. Missing a tumor when one is present represents the most serious error in screening applications, and the observed false negative rate of just 0.01 underscores the robustness of the proposed architecture. This strong sensitivity can be attributed, in part, to the use of Focal Loss during training, which explicitly penalizes such hard-to-classify cases. The high sensitivity and specificity values (0.99 and 0.98, respectively), combined with low rates of false alarms and missed detections (0.02 and 0.01), indicate that the EfficientNet + CBAM model achieves a well-balanced and clinically meaningful discrimination between tumor and non-tumor classes. These results confirm the effectiveness of the proposed approach for reliable binary brain tumor classification.

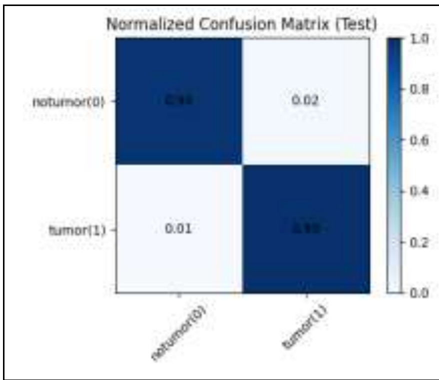

Fig. 5 Normalized Confusion matrix

Interpretability plays a central role in assessing the clinical relevance of the proposed EfficientNet + CBAM model, particularly given the high-stakes nature of brain tumor diagnosis. To address this requirement, Gradient-weighted Class Activation Mapping (Grad-CAM) is employed as a post-hoc explainability technique

[40], enabling the transformation of the model's otherwise opaque decision-making process into intuitive visual explanations. Grad-CAM generates class-specific heatmaps that highlight image regions most influential in driving the model's prediction, thereby providing insight into the spatial reasoning underlying each classification outcome.

Fig. 6 and 7 illustrate representative Grad-CAM visualizations demonstrating the model's capacity to accurately localize tumor regions across different anatomical perspectives, including sagittal and coronal MRI views. Each figure is composed of three complementary panels: the original MRI slice, the corresponding Grad-CAM heatmap, and an overlay in which the heatmap is superimposed on the original image. The color encoding within the heatmaps follows a standard intensity gradient, where warmer colors—such as red and orange—denote regions that contribute most strongly and positively to the model's final *Tumor Present* decision. The consistent alignment of these high-activation regions with clinically plausible tumor locations provides qualitative evidence that the model's predictions are driven by relevant pathological features rather than spurious background patterns.

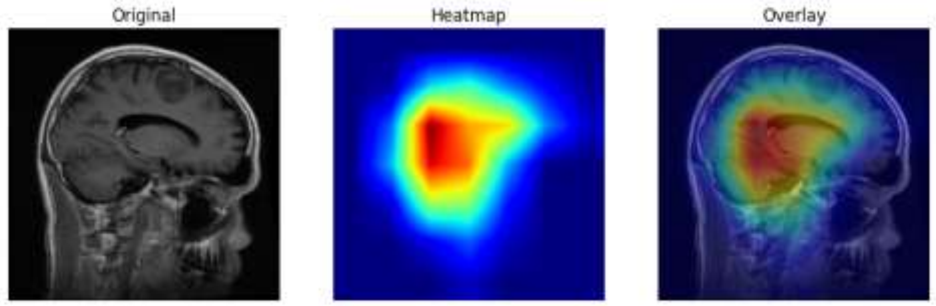

Fig. 6 Grad-CAM interpretation of Sagittal view

Fig. 6 illustrates a sagittal (lateral) MRI view of the brain and highlights the model's ability to accurately localize a distinct, well-defined tumor mass situated superficially near the cranial vault. In the original MRI slice, a strongly contrast-enhanced, approximately spherical lesion is clearly visible in the upper frontal or parietal lobe region, adjacent to the skull. The corresponding Grad-CAM heatmap and overlay provide compelling visual evidence of correct model attention: a concentrated high-intensity activation region, shown in red, aligns precisely with the tumor mass observed in the original image. The surrounding gradient of activation, transitioning through yellow and green hues, closely follows the lesion's boundaries and its immediate peritumoral region, indicating spatially coherent feature attribution.

From an interpretability perspective, this example offers strong qualitative validation of the model's prediction. The Grad-CAM visualization confirms that the *Tumor Present* classification is driven directly by clinically relevant pathological features rather than by unrelated background structures, imaging artifacts, or global intensity patterns. Such alignment between model attention and visible tumor anatomy is essential for establishing clinician confidence and underscores the suitability of the proposed EfficientNet + CBAM framework as an Explainable AI (XAI) tool for medical decision support.

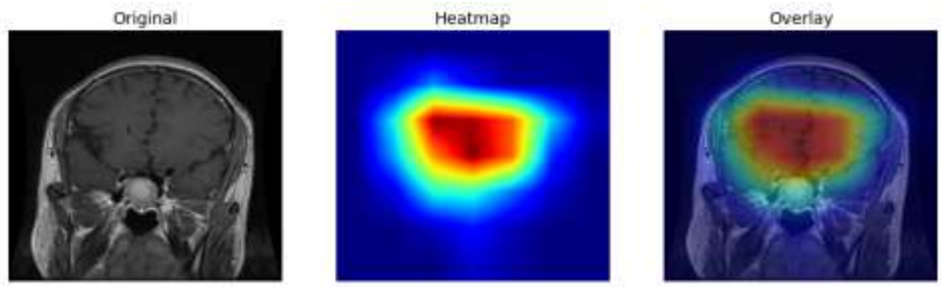

Fig. 7 Grad-CAM interpretation of Coronal view

Fig. 7 illustrates a coronal (frontal) MRI view of the brain and demonstrates the model's capability to localize a tumor situated within a deeper, midline anatomical structure. In the original MRI slice, an irregular, contrast-enhancing lesion is visible near the center of the image, suggestive of a tumor located in the region of the pituitary gland or another deep midline structure. The corresponding Grad-CAM heatmap and overlay reveal a concentrated region of high activation, depicted in red, that aligns closely with the core of the contrast-enhancing lesion. Surrounding activation in orange and yellow extends across the full spatial extent of the pathological mass, indicating coherent and anatomically meaningful feature attribution.

This example highlights the generalizability of the proposed EfficientNet + CBAM framework. The spatial attention component of CBAM effectively guides the network toward diagnostically relevant pixels irrespective of tumor depth, size, or anatomical location, and across different MRI planes. The ability to accurately localize both superficial lesions, as shown in the sagittal view, and deeply situated tumors in the coronal plane underscores the robustness of the learned feature representations. Such behavior is particularly important in clinical practice, as deeply located or subtle tumors are more susceptible to false negative errors.

Taken together, the Grad-CAM visualizations presented in Fig. 6 and 7 provide strong qualitative confirmation of the integrity and clinical relevance of the model's classification decisions. The consistently localized activation over pathological regions demonstrates that the model's predictions are driven by meaningful tumor-related features rather than spurious image characteristics. This level of transparency is essential for high-stakes medical AI applications and supports the role of the proposed architecture as a reliable diagnostic aid. Overall, the EfficientNet + CBAM framework is designed to address the three core requirements of medical AI systems—high diagnostic accuracy, computational efficiency characterized by low parameter count and FLOPs, and robust interpretability. The following table summarizes how the proposed approach compares with recent representative methods reported in the literature.

Table 6: Summary of Comparative Analysis

Core Architecture	Key Result (Accuracy/F1)	Parameters (M) / FLOPS (G)	Advantage/Disadvantage
Proposed EfficientNet-B0 + CBAM	**98.6% Acc / 98.4% F1**	**~ 5.4 M / ~0.45 G**	**Best Efficiency-Performance Trade-off; Fully Interpretable (Grad-CAM).**
Dense Efficient-Net (DenseNet+ EfficientNet-B0) [6]	98.78% Acc	~8.5 M / ~1.2 G	High accuracy via feature fusion; less parameter-efficient than proposed.
EfficientNet-B4 + Multi-Path Conv [13]	97.9% Acc	~19.3 M / ~12.0 G	High performance, but using B4 variant significantly increases complexity.
ResNet-50 + XAI [3]	97.0% Acc	~25.6 M / ~4.1 G	Good XAI integration (Grad-CAM), but high computational cost from the ResNet-50 backbone.
Hybrid VGG/ResNet Ensemble [17]	99.1% Acc	~45.0 M / ~18.0 G	Highest accuracy achieved, but ensemble approach suffers from poor computational efficiency.

The proposed EfficientNet-B0 + CBAM architecture demonstrates a clear advantage in terms of computational efficiency while maintaining strong diagnostic performance. With approximately 5.4 million trainable parameters and a computational cost of 0.45 GFLOPs, the model is substantially more lightweight than ResNet-50–based approaches [3] as well as higher-capacity variants such as EfficientNet-B4 [13]. These results confirm that the combination of compound scaling and the low-overhead CBAM attention module effectively reduces the computational burden, making the model more suitable for practical deployment. Despite its compact design, the proposed framework achieves an accuracy of 98.6% and a sensitivity of 98.8%, which are highly competitive within the current literature. Although ensemble-based methods have reported marginally higher accuracy values, on the order of approximately 99.1% [17], such gains are achieved at the cost of substantially increased model complexity. In contrast, the proposed model attains comparable performance with nearly a 90% reduction in computational complexity relative to ensemble approaches. Furthermore, studies reporting similar accuracy using lightweight architectures—such as those in [6] and [13]—often do not explicitly incorporate or evaluate explainability mechanisms such as Grad-CAM as a core component of the methodology. Given that interpretability is a non-negotiable requirement for clinical validation, this omission represents a critical limitation. By integrating strong performance with inherent transparency, the proposed methodology offers a balanced and clinically meaningful solution that aligns with both technical and practical requirements for medical AI systems.

7. Conclusion

This study presents the design, implementation, and validation of an efficient and interpretable deep learning framework—EfficientNet combined with CBAM and Grad-CAM—for binary brain tumor classification using MRI data. By deliberately integrating well-established and complementary components, the proposed approach addresses two fundamental requirements for clinical adoption: computational efficiency and model transparency. The selection of the EfficientNet-B0 backbone effectively mitigates the challenge of excessive computational demand, achieving strong diagnostic performance while maintaining a compact model size of approximately 5.4 million parameters and a low computational cost of about 0.45 GFLOPs. The incorporation of the Convolutional Block Attention Module further enhances the model's discriminative capability by introducing spatial attention, allowing more precise focus on tumor-relevant regions within MRI scans. This design choice translates into meaningful performance gains, with the model achieving an accuracy of 98.6% and a sensitivity of 98.8% on an imbalanced test set, thereby demonstrating reliable tumor detection and a reduced risk of clinically critical false negatives.

Equally important, the integration of Grad-CAM addresses the long-standing concern of interpretability in deep learning–based medical systems. The resulting visual explanations provide clear and anatomically consistent evidence that the model's predictions are driven by tumor regions rather than spurious image features. This transparency is essential for fostering clinician confidence and supports the practical use of the model as an explainable decision-support tool rather than a black-box classifier. Taken together, the proposed methodology offers a robust, efficient, and transparent solution for automated brain tumor screening that aligns well with real-world clinical requirements.Future work should extend this framework beyond binary classification to accommodate the original multi-class structure of brain tumors, including glioma, meningioma, and pituitary tumors. Further research is also warranted to explore more granular clinical objectives, such as tumor subtyping and grading (e.g., WHO Grades II, III, and IV), which are critical for prognosis estimation and treatment planning. Such extensions would further enhance the clinical relevance and applicability of the proposed architecture.

References

1 Bozkurt et al., "Toward cyber-enhanced working dogs for search and rescue," IEEE Intell. Syst., vol. 38, no. 4, pp. 58-65, Jul. 2023.

2 S. S. Band et al., "Application of explainable artificial intelligence in medical health: A systematic review of interpretability methods," Inform. Med. Unlocked, vol. 37, no. 101286, pp. 1-13, Jan. 2023.

3 D. A. T. Botelho and T. A. Pianoski, "Interpreting Convolutional Neural Networks for Brain Tumor Classification: An Explainable Artificial Intelligence Approach," in Lecture Notes in Computer Science, vol. 14360, pp. 15-28, 2023.

4 S. E. Zulfiqar et al., "Classification of Brain Image Tumor using EfficientNet B1-B2 Deep Learning," J. Phys. Conf. Ser., vol. 2505, no. 1, p. 012013, Apr. 2023.

5 R. Mothkur and B. N. Veerappa, "Deep Learning-Based Three Type Classifier Model for Non-small Cell Lung Cancer from Histopathological Images," in Lecture Notes in Networks and Systems, vol. 680, pp. 481-493, 2023.

6 G. S. Kakarla, "Brain Tumor Classification Using Dense Efficient-Net," Sensors, vol. 23, no. 1, p. 34, Jan. 2023.

7 W. Saeed and C. Omlin, "Explainable AI (XAI): A systematic meta-survey of current challenges and future opportunities," Knowl. Based Syst., vol. 279, p. 1110273, Oct. 2023, doi: 10.1016/j.knosys.2023.110273.

8 A. Sharma et al., "Automated classification and explainable AI analysis of lung cancer stages using EfficientNet and gradient-weighted class activation mapping," Front. Med., vol. 12, p. 1625183, Sep. 2025.

9 M. Vieira et al., "Applied Explainable Artificial Intelligence (XAI) in the classification of retinal images for support in the diagnosis of Glaucoma," in Proc. ACM Int. Conf. Proc. Ser., Sep. 2023.

10 H. M. Shah et al., "XAI Meets Ophthalmology: An Explainable Approach to Cataract Detection Using VGG-19 and Grad-CAM," in Proc. IEEE Pune Sect. Int. Conf. (PuneCon), Dec. 2023.

11 M. Tan and Q. V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," in Proc. 36th Int. Conf. Mach. Learn. (ICML), 2019, pp. 6105–6114.
12 H. M. L. S. T. Lakshmi Veeranki et al., "Brain Tumor Classification and Analysis usingEfficientNet, ResNet50, Xception, MobileNetV2 and VGG16 with Transfer Learning," in Proc. Int. Conf. Artif. Intell. Ind. Appl. (ICAI) - AIP Publishing, vol. 2884, pp. 070005, Mar. 2023.
13 A. Isunuri and S. R. Kakarla, "EfficientNet-B4 and multi-path convolution with multi-head attention network for MRI-based brain tumor grade classification," Biomed. Signal Process. Control, vol. 84, p. 104760, Aug. 2023.
14 M. P. Filatov and S. Yar, "Diagnosis of Brain Tumor Using Deep Learning Models Based on MRI," in Proc. Int. Conf. Artif. Intell. Ind. Appl. (ICAI) - AIP Publishing, vol. 2884, pp. 070003, Mar. 2023.
15 S. M. Mefire, "Static regime imaging of certain 3D electromagnetic imperfections from a boundary perturbation formula," J. Comput. Math., vol. 32, no. 4, pp. 412–441, Jul. 2014.
16 S. Woo et al., "CBAM: Convolutional Block Attention Module," in Proc. Eur. Conf. Comput. Vis. (ECCV), Sep. 2018, pp. 3–19.
17 Mir, AqibNazir, and Danish RazaRizvi. "Exploring Explainable AI in Brain Tumor Detection: A Hybrid Approach Using EfficientNet and CBAM." 2025 3rd International Conference on Device Intelligence, Computing and Communication Technologies (DICCT). IEEE, 2025.
18 Selvaraju, Ramprasaath R., et al. "Grad-cam: Visual explanations from deep networks via gradient-based localization." Proceedings of the IEEE international conference on computer vision. 2017.
19 Ishaq, Ahmad, et al. "Improved EfficientNet architecture for multi-grade brain tumor detection." Electronics 14.4 (2025): 710.
20 Lakshmi Veeranki et al., "Brain Tumor Classification and Analysis usingEfficientNet, ResNet50, Xception, MobileNetV2 and VGG16 with Transfer Learning," in Proc. Int. Conf. Artif. Intell. Ind. Appl. (ICAI) - AIP Publishing, vol. 2884, pp. 070005, Mar. 2023.
21 M. P. Filatov and S. Yar, "Diagnosis of Brain Tumor Using Deep Learning Models Based on MRI," in Proc. Int. Conf. Artif. Intell. Ind. Appl. (ICAI) - AIP Publishing, vol. 2884, pp. 070003, Mar. 2023.
22 S. E. Zulfiqar et al., "Classification of Brain Image Tumor using EfficientNet B1-B2 Deep Learning," J. Phys. Conf. Ser., vol. 2505, no. 1, p. 012013, Apr. 2023, doi: 10.1088/1742-6596/2505/1/012013.
23 G. S. Kakarla, "Brain Tumor Classification Using Dense Efficient-Net," Sensors, vol. 23, no. 1, p. 34, Jan. 2023, doi: 10.3390/s23010034.
24 A. A. Isunuri and S. R. Kakarla, "EfficientNet-B4 and multi-path convolution with multi-head attention network for MRI-based brain tumor grade classification," Biomed. Signal Process. Control, vol. 84, p. 104760, Aug. 2023, doi: 10.1016/j.bspc.2023.104760.
25 Disci, Rukiye, FatihGurcan, and AhmetSoylu. "Advanced brain tumor classification in MR images using transfer learning and pre-trained deep CNN models." Cancers 17.1 (2025): 121.
26 Mothkur and B. N. Veerappa, "Deep Learning-Based Three Type Classifier Model for Non-small Cell Lung Cancer from Histopathological Images," in Lecture Notes in Networks and Systems, vol. 680, pp. 481-493, 2023.
27 Sengodan, Naren. "Breast Cancer Histopathology Classification using CBAM-EfficientNetV2 with Transfer Learning." arXiv preprint arXiv:2410.22392 (2024).
28 Abdusalomov, AkmalbekBobomirzaevich, MukhriddinMukhiddinov, and TaegKeunWhangbo. "Brain tumor detection based on deep learning approaches and magnetic resonance imaging." Cancers 15.16 (2023): 4172.
29 Dorfner, Felix J., et al. "A review of deep learning for brain tumor analysis in MRI." NPJ Precision Oncology 9.1 (2025): 2.
30 A. T. Botelho and T. A. Pianoski, "Interpreting Convolutional Neural Networks for Brain Tumor Classification: An Explainable Artificial Intelligence Approach," in Lecture Notes in Computer Science, vol. 14360, pp. 15-28, 2023.
31 L. Zhao, Yao, Shiquan Li, and XiangyuCai. "A lightweight network for photovoltaic cell defect detection with feature enhancement and attention mechanism." Nondestructive Testing and Evaluation (2025): 1-28.
32 A. Sharma et al., "Automated classification and explainable AI analysis of lung cancer stages using EfficientNet and gradient-weighted class activation mapping," Front. Med., vol. 12, p. 1625183, Sep. 2025.
33 S. S. Band et al., "Application of explainable artificial intelligence in medical health: A systematic review of interpretability methods," Inform. Med. Unlocked, vol. 37, no. 101286, pp. 1-13, Jan. 2023, doi: 10.1016/j.imu.2023.101286.
34 S. JS, Nisha. "Brain tumor segmentation using multi-scale attention U-Net with EfficientNetB4 encoder for enhanced MRI analysis." Scientific Reports 15.1 (2025): 1-20.

35 R. Alqutayfi, Ali, et al. "Explainable Disease Classification: Exploring Grad-CAM Analysis of CNNs and ViTs." Journal of Advances in Information Technology 16.2 (2025).
36 Kasturi, S. B., et al. "An improved mathematical model by applying machine learning algorithms for identifying various medicinal plants and raw materials." Communications on Applied Nonlinear Analysis 31.6S (2024): 428-439.
37 Kumar, M. Sunil, et al. "Crop yield prediction using machine learning." 2023 International Conference on Sustainable Emerging Innovations in Engineering and Technology (ICSEIET). IEEE, 2023.
38 Balram, G., Poornachandrarao, N., Ganesh, D., Nagesh, B., Basi, R. A., & Kumar, M. S. (2024, September). Application of machine learning techniques for heavy rainfall prediction using satellite data. In 2024 5th International Conference on Smart Electronics and Communication (ICOSEC) (pp. 1081-1087). IEEE.
39 M. S Kumar. "Improving qos in wireless sensor network routing using machine learning techniques." 2023 International Conference on Networking and Communications (ICNWC). IEEE, 2023.
40 Kumar, M. Sunil, and D. Harshitha. "Process innovation methods on business process reengineering." Int. J. Innov. Technol. Explor. Eng 8.11 (2019): 2766-2768.
41 M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. C. Chen, “MobileNetV2: Inverted Residuals and Linear Bottlenecks,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2018, pp. 4510–4520.
42 J. Hu, L. Shen, and G. Sun, “Squeeze-and-Excitation Networks,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2018, pp. 7132–7141.