RESEARCH ARTICLE OPEN ACCESS

Deep Learning and Explainability in Brain Tumor Classification: A Comprehensive MRI-Based Review (2011–2025)

Shubham Porte*, Shanu Kuttan Rakesh **
*(M. Tech. Scholer, CSE Department, Chouksey Engineering College, Bilaspur (C.G.)
Email: shubhamporte0@gmail.com)
** (Associate Professor, CSE Department, Chouksey Engineering College, Bilaspur (C.G.)
Email: shanu.kuttan28@gmail.com)

--************************----------------------------------

Abstract:

One of the most difficult tasks in medical imaging is diagnosing brain tumors, primarily because tumor variability and the complexity of MRI scans. Over the last decade, research has shifted from conventional machine learning models to more advanced deep learning and transfer learning architectures. These approaches have consistently reported promising results, with accuracies ranging from 80% to nearly 98% in classification tasks. At the same time, XAI, or explainable artificial intelligence, has become a critical component, enabling clinicians to visualize and validate automated predictions. Grad-CAM, SHAP, and LIME are several techniques that help close the gap between algorithmic output and medical reasoning, thereby improving trust and clinical usability. This review compiles studies published between 2011 and 2025, examining their methodologies, strengths, and limitations. It also highlights major challenges, including dataset limitations, class imbalance, and lack of integration into real clinical workflows. The review concludes that future systems must strike a balance between predictive accuracy and interpretability to support safe and effective adoption in healthcare practice.

.

Keywords* — *Brain tumor diagnosis, Magnetic Resonance Imaging (MRI), Deep learning, Transfer learning, Explainable AI (XAI), Interpretability in healthcare.

--************************----------------------------------

I. INTRODUCTION

Brain tumors remain among the most serious and life-threatening neurological disorders worldwide, with high mortality and morbidity rates if not detected at an early stage. Accurately and timely making a diagnosis is essential to improving patient survival and informing treatment planning. Magnetic Resonance Imaging (MRI) is considered the gold standard for brain tumor diagnosis because of its ability to provide high-resolution, non-invasive visualization of soft tissues and abnormal brain structures [1]. However, the manual interpretation of MRI scans are a time-consuming procedure that depends heavily on radiologists' expertise. Inter-observer variability and the large volume of medical images often lead to delays or inconsistent diagnoses, highlighting the need for automated and reliable diagnostic solutions [2].

Medical image analysis has been completely transformed by the explosive expansion of artificial intelligence (AI) and intense learning. Convolutional Neural Networks (CNNs) have proven to be highly effective at identifying intricate patterns and minute details in medical scans. which are often difficult for the human eye to discern [3]. Among various CNN architectures, VGG16, ResNet, and EfficientNet are widely applied for tumor detection and classification tasks due to their proven ability in feature extraction and transfer learning [4]. By leveraging these models, researchers have achieved classification accuracies exceeding 90% in distinguishing tumor forms such as pituitary tumors, meningiomas, and gliomas [5]. This demonstrates the strong potential of AI-driven methods to complement radiologists, reduce workload, and improve diagnostic efficiency in clinical workflows.

Despite their high predictive performance, most deep learning models are considered "black boxes," making it challenging to understand how they make decisions. In medical settings, this lack of transparency poses a substantial obstacle to clinical implementation of automated predictions, as healthcare professionals require justification for their use before integrating them into treatment decisions [6]. Artificial intelligence that can be explained (XAI) addresses this problem by offering visual and analytical explanations for model predictions. Gradient-weighted Class Activation Mapping (Grad-CAM) is one technique that highlights discriminative regions in MRI scans that contribute to tumor classification, enabling clinicians to validate whether the model focuses on medically relevant features [7]. Similarly, methods like SHAP and LIME provide feature-level interpretability, fostering greater trust and accountability. Integrating explainability with high classification accuracy bridges the gap between AI innovation and real-world medical practice, ensuring both reliability and clinical acceptance [8].

II. EXISTING APPROACHES FOR BRAIN TUMOR CLASSIFICATION

A. TRADITIONAL MACHINE LEARNING APPROACHES

Early works (2011–2015) on brain tumor classification relied heavily on hand-crafted feature extraction, such as texture (GLCM, LBP), shape, and wavelet features. Classifiers like Support Vector Machines (SVMs), Random Forests, and KNN were widely applied [2]. These approaches required domain expertise to select features and often performed well on small, curated datasets. However, their limitations included:

- Strong dependency on manual feature engineering,
- Poor generalization on unseen or heterogeneous data,
- Difficulty handling multi-class classification, especially when including a “no tumor” category [5].

B. DEEP LEARNING

Since 2015, Medical image analysis has been revolutionized by Convolutional Neural Networks (CNNs), which automatically learn hierarchical characteristics from MRI scans [4]. Architectures such as VGG16, ResNet, DenseNet, Inception, and EfficientNet attained cutting-edge results in tumorcategorization.

Recent studies reported classification accuracies exceeding 90–96% for multi-class problems involving glioma, meningioma, pituitary tumor, and no tumor [3,9]. Key trends include:

- Hybrid and ensemble CNNs (e.g., combining VGG16 and ResNet50),
- Advanced data augmentation (rotation, flipping, brightness/contrast adjustment),
- Reporting with confusion matrices, ROC-AUC, and class-wise F1-scores,
- Integration of Explainable AI (Grad-CAM) to highlight tumor-relevant MRI regions and improve clinical trust [10].

C. TRANSFER LEARNING IN MEDICAL IMAGING

Due to the scarcity of annotated medical data, **transfer learning (TL)** has become the mainstream approach [11]. Pre-trained models (ImageNet-trained VGG, ResNet, EfficientNet) are fine-tuned for MRI classification.

Advantages include:

- Faster convergence and reduced training cost,
- Better generalization even on limited data.
- Easier integration with explainability tools.
Recent studies (2023–2025) increasingly combine **transfer learning with XAI (Grad-CAM, SHAP, LIME)** to make CNN predictions both accurate and interpretable, a key requirement for deployment in clinical workflows [6].

Table 1 provides a consolidated overview of 30 representative studies published between 2011 and 2025 on brain tumor classification using MRI. The table highlights the transition from traditional machine learning (feature-based SVM, KNN, RF) to deep learning architectures (CNN, VGG16, ResNet, DenseNet) and, more recently, transfer learning with explainable AI (Grad-CAM, SHAP, LIME). For each study, the methodology, dataset, class setup, explainability technique, and major findings are summarized, demonstrating how research has progressively moved toward multi-class classification, higher accuracy, and clinical interpretability.

Table 1. Overview of major approaches (2011–2025) for brain tumor classification using MRI.

Year	Focus / Study	Method / Architecture	Dataset	Classes	XAI Used	Key Findings
2011 [5]	Feature-based CAD	GLCM + SVM	Local MRI	2–3	–	Early ML approach; feature dependence [1]

2012 [4]	Wavelet features	Wavelet + KNN	Private	2–3	–	Limited generalization
2014 [11]	Survey & algorithm	CAD + ML	Mixed	3	–	Early review; CAD pipelines dominate [2]
2015 [2]	Deep learning intro	VGG16	Public	3	–	Start of CNN shift [3]
2016 [3]	CNN vs ML	CNN vs SVM/RF	BRATS	3	–	CNN superior
2017 [10]	Explainability	Grad-CAM introduced	–	–	Grad-CAM	Foundation for XAI [6]
2018 [13]	Early TL	VGG/ResNet TL	Kaggle	3	–	Faster convergence
2019 [9]	Data augmentation survey	Augmentation strategies	–	–	–	Became standard [7]
2020 [14]	DL tumor classification	Custom CNN/ResNet	Mixed MRI	3–4	–	90%+ accuracy [4]
2021 [15]	Comparative CNNs	VGG, ResNet, DenseNet	Open	3–4	–	CNN families compared [5]
2021 [16]	Medical XAI survey	XAI in healthcare	–	–	Grad-CAM/SHAP	Trust & interpretability [8]
2023 [17]	Modified VGG19	VGG19 + Augmentation	MRI diverse	3	Grad-CAM	Improved generalization
2023 [18]	Hybrid CNN-SVM	CNN features + SVM	Kaggle	3	–	Hybrid boosts accuracy
2024 [19]	Explainable CNN	Grad-CAM on MRI	Public	3–4	Grad-CAM	High performance + XAI
2024 [20]	Transfer learning survey	Multiple CNNs	Several	3–4	–	TL boosts efficiency
2024 [21]	ResNet50 + Grad-CAM	ResNet50 TL	MRI	3–4	Grad-CAM	Visual focus validated
2024 [22]	Modified InceptionV3	InceptionV3 + XAI	Public	3–4	Grad-CAM	Strong multi-class
2024 [23]	DenseNet vs VGG	DenseNet121, VGG16	MRI	3–4	–	Trade-offs reported
2024 [24]	Region-specific study	TL + XAI	Bangladesh MRI	3	Grad-CAM	Clinical applicability
2024 [25]	CAD with XAI	Mixed CNN models	Multi-datasets	3–4	Grad-CAM	XAI mainstream
2025 [26]	Hybrid VGG16	VGG16 + hybrid	Public	3–4	Grad-CAM	Accuracy + interpretability
2025 [27]	Ensemble CNNs	VGG, ResNet, EffNet	Multiple	3–4	Grad-CAM	Ensemble stability
2025 [28]	Explainable CNN pipeline	CNN + SHAP/LIME	MRI	3–4	SHAP/LIME	Deeper insights
2025 [29]	DBN-VGG16	DBN + VGG16	MRI	3–4	Grad-CAM	Better hierarchical learning

2025 [30]	Maxout-VGG16	VGG16 + Maxout	MRI	3–4	Grad-CAM	Handles difficult samples
2025 [31]	Ensemble VGG+ResNet	Ensemble CNNs	Kaggle	3–4	Grad-CAM	Reduced overfitting
2025 [32]	Explainable CNN (XAI)	CNN + key features	MRI	3–4	Grad-CAM	Clinically verifiable
2025 [33]	Fusion CNNs	VGG, MBNet, EffNet	MRI	3–4	Grad-CAM	Fusion improves accuracy
2025 [34]	Hybrid CNN-VGG16	Conf. paper	MRI	3–4	Grad-CAM	Conference-level results
2025 [35]	Biomedical hybrid	VGG-hybrid	MRI	3–4	Grad-CAM	Journal-grade robustness

III. EXPLAINABLE AI IN BRAIN TUMOR DIAGNOSIS

A. *Grad-CAM*

One of the most widely adopted explainability Gradient-weighted Class Activation Mapping is a technique used in medical image analysis (Grad-CAM). It works by producing heatmaps

that highlight the most discriminative regions in MRI scans, influencing the model's classification decision [7]. In brain tumor studies, Grad-CAM has been extensively used with CNN-based models such as VGG16, ResNet, and Inception to verify whether the network is focusing on tumor-affected regions rather than irrelevant background areas [16]. These heatmaps provide radiologists with visual confirmation, bridging the gap between model predictions and clinical reasoning. Numerous studies have reported that Grad-CAM not only improves trust in AI predictions but also helps detect cases where the network may misclassify due to confounding features, making it an essential tool for clinical validation [8].

B. SHAP and LIME

Beyond Grad-CAM, model-agnostic approaches such as LIME (Local Interpretable Model-agnostic Explanations) and SHAP (SHapley Additive exPlanations) have been used to brain tumor diagnosis. Unlike Grad-CAM, which is designed primarily for CNNs, SHAP and LIME can interpret a broader range of machine learning and deep learning models. SHAP uses cooperative game theory to assign contribution values to each feature, thereby explaining how different inputs (e.g., pixel intensities, image features) contribute to a prediction [36]. LIME, on the other hand, generates simplified local surrogate models that approximate the decision-making process of complex models for specific predictions [37]. In medical imaging, SHAP and LIME are especially useful for quantifying feature-level importance, complementing the region-based insights provided by Grad-CAM.

Despite the impressive accuracy of deep learning models, their adoption in clinical settings remains limited due to the "black-box" nature of AI systems [6]. In healthcare, accuracy alone is insufficient—clinicians require transparency, accountability, and justification before relying on automated predictions for diagnosis or treatment planning. Explainability tools such as Grad-CAM, SHAP, and LIME enable practitioners to cross-check AI decisions against medical expertise. This is crucial for:

- Building clinical trust: Radiologists can verify if the model is attending to medically relevant tumor regions.
- Improving patient safety: Misclassifications can be detected early when heatmaps or feature explanations reveal inconsistencies.
- Supporting education and training: Visual explanations help medical trainees understand both tumor characteristics and AI decision-making.
- Facilitating regulatory approval: Interpretable AI models are more likely to meet ethical and legal standards for deployment in healthcare systems [22].

In summary, interpretability transforms AI from a "black box" into a clinically reliable decision-support tool, ensuring that automated systems assist rather than replace human expertise.

IV. COMPARATIVE ANALYSIS OF STUDIES

Across reviewed studies from 2011 to 2025, reported classification performance for brain-tumor detection and multi-class classification (glioma, meningioma, pituitary, no-tumor) typically lies in the **80%–98%** range for overall accuracy. Older, feature-based ML methods commonly reported accuracies near the lower end (≈80–88%), especially on small or private datasets. With the widespread adoption of CNNs, transfer learning, and ensemble strategies (2016 onward), many works began reporting **90%+** accuracy in multi-class settings, with top-performing studies (often using ensembles or heavy augmentation + TL) claiming accuracies up to **96–98%**. However, reported peak accuracies should be interpreted cautiously — variations in dataset size, class balance, pre-processing, cross-validation protocol, and test set composition all strongly affect reported numbers. The literature matrix below (20 selected studies) summarizes reported or typical accuracy ranges and the presence of explainability techniques.

A. STRENGTHS AND LIMITATIONS

Strengths observed across the literature

- Automatic feature learning: Deep CNNs eliminate manual feature engineering and learn hierarchical representations robust to some image variability.
- Transfer learning gains: ImageNet-pretrained backbones reduce training time and improve performance on small datasets.
- Augmentation & ensembles: Aggressive augmentation and model ensembles improve generalization and stability.
- Explainability integration: Grad-CAM (and increasingly SHAP/LIME) directly addresses clinical trust issues by producing interpretable visualizations.

Common limitations

- Dataset heterogeneity & size: Many studies rely on small, single-center, or merged public datasets; external, multi-center validation is often missing.
- Class imbalance & evaluation reporting: Several works omit class-wise metrics, report only accuracy (not macro-F1 or per-class recall), or use inconsistent cross-validation — complicating fair comparisons.
- Overfitting risk: High reported accuracies sometimes arise from leakage, insufficient cross-validation, or non-independent test sets.
- XAI limitations: Grad-CAM provides region-level saliency but not rigorous causal explanations; SHAP/LIME add feature-level insight but are computationally expensive and need careful interpretation for images.

Recent best practices emphasize balanced evaluation: reporting accuracy plus macro-F1, per-class recall/precision, and ROC-AUC; performing strict cross-validation and external testing; and integrating XAI tools (Grad-CAM for spatial validation, SHAP/LIME for feature-level explanations) to make high-accuracy models clinically actionable. Successful pipelines typically follow: (1) careful preprocessing + augmentation, (2) transfer learning + selective fine-tuning, (3) class imbalance handling (class weights or focal loss), and (4) XAI validation where heatmaps are corroborated with expert annotations. Combining performance metrics with interpretability improves model transparency, helps detect failure modes, and increases the likelihood of clinical acceptance.

Table 2. Comparative summary of 20 studies (2011–2025) on MRI-based brain tumor classification with reported accuracy and explainability.

Year	Study (short)	Method / Architecture	Dataset (type)	Reported Accuracy / Range	XAI used	Key strength / limitation
2014	El-Dahshan et al.	Feature-based CAD + ML (texture, wavelet + SVM)	Mixed/private	~80–85%	No	Early CAD pipeline; limited generalization. [5]
2015	Simonyan & Zisserman (VGG intro)	VGG family (foundation)	N/A (method paper)	N/A	N/A	Architectural foundation for TL in medical imaging. [4]
2017	Selvaraju et al. (Grad-CAM)	Grad-CAM (explainability method)	N/A	N/A	Grad-CAM	Introduced practical visual

						saliency for CNNs. [7]
2019	Shorten & Khoshgoftaar	Data augmentation survey	N/A	N/A	N/A	Augmentation strategies standardized for medical DL. [13]
2020	Afshar et al.	Capsule networks (CNN variant)	Public MRI	82–90%	No/limited	Better spatial relationships; complex training. [11]
2020	Rehman et al.	Deep learning (custom CNN)	Mixed MRI	85–92%	No	Demonstrated DL advantage over classical ML. [2]
2021	Iqbal et al.	Comparative DL (VGG/ResNet/DenseNet)	Open sets	88–93%	No	Comparative benchmarks across backbones. [3]
2023	El-Dahshan et al. (VGG19 modified)	Modified VGG19 + augmentation	Public MRI	90–95%	Grad-CAM	Improved generalization with preprocessing. [13]
2023	Rani & Kaur	Hybrid CNN + SVM	Kaggle MRI	89–94%	No	Hybrid approach boosts accuracy on small sets. [14]
2024	Jalal et al.	Transfer learning, fine-tuned CNN	Public	90–96%	No/Grad-CAM	High accuracy using TL backbones. [9]
2024	Khan et al.	Explainable DL with Grad-CAM	Public MRI	91–95%	Grad-CAM	Strong XAI focus; visual validation with clinicians. [15]
2024	Guluwadi	ResNet50 + Grad-CAM	Public MRI	92–95%	Grad-CAM	ResNet backbone + XAI; robust localization. [24]
2024	Ullah et al.	Modified InceptionV3 + Grad-CAM	Conference / public	90–94%	Grad-CAM	Shows TL + XAI synergy. [23]
2024	Shamshad et al.	Transfer learning study across models	Multiple MRI sets	88–95%	Varies	Systematic TL comparison; model efficiency focus. [31]
2024	Masab et al.	DenseNet121 vs VGG16 vs custom CNN	Public MRI	89–94%	No	Architecture trade-offs discussed. [32]
2024	Begum & Kalilulah	VGG16 + MobileNetV2 fusion	Conference	90–95%	No	Lightweight model options for deployment. [33]
2024	Mitra et al.	Modified VGG16	Conference	91–96%	No / Grad-CAM	Optimized VGG16 yields high accuracy. [37]
2025	Sánchez-Moreno et al.	Ensemble CNNs + XAI	Public	92–97%	Grad-CAM, SHAP	Ensemble + XAI improves both accuracy &interpretability. [8]

2025	Sharma & Rajalakshmi	VGG16 segmentation + classification	Journal	93–97%	Grad-CAM	Strong augmentation + segmentation boost results. [16]
2025	Chikhale & Kakani	Ensemble VGG16 + ResNet50	Springer conf.	92–98%	Grad-CAM	Ensemble shows best reported top accuracies; needs external validation. [30]

V. CHALLENGES AND FUTURE DIRECTIONS

A. DATASET LIMITATION

One of the persistent challenges in brain tumor classification research is the limited availability of large-scale, well-annotated, and standardized MRI datasets. Most existing studies rely on a small number of public repositories or locally collected scans, which often differ in acquisition parameters, image quality, and labeling standards. This heterogeneity restricts the generalizability of models across institutions. Furthermore, patient privacy and ethical restrictions make it difficult to share medical data widely, thereby slowing progress toward robust and clinically reliable systems. Creating multi-institutional benchmark datasets with consistent annotations is essential to overcoming this barrier.

B. CLASS IMBALANCE

Brain tumor datasets often suffer from imbalanced class distributions, where some tumor subtypes (such as gliomas) are represented in much larger numbers compared to others (such as meningiomas or pituitary tumors). In addition, the inclusion of a “no tumor” category can further distort class proportions. This imbalance skews training, causing models to favor majority classes while underperforming on minority ones. Several approaches—such as data augmentation, weighted loss functions, and generative methods for synthetic data—have been proposed, but a universally accepted strategy is still lacking. Ensuring balanced representation remains a priority to achieve fair and reliable predictions across all tumor categories.

C. 3D MRI and MULTIMODEL IMAGING

Most current classification pipelines are based on 2D MRI slices, which limits the spatial context captured from volumetric scans. Tumor morphology, however, is inherently three-dimensional, and ignoring inter-slice continuity can lead to loss of critical diagnostic information. Recent works highlight the potential of 3D CNNs and volumetric processing for more comprehensive feature extraction. Additionally, combining MRI with other imaging modalities (e.g., PET, CT, spectroscopy) or incorporating clinical metadata (patient history, genetic markers) can enable multimodal classification systems. Such approaches may provide a more holistic view of tumor characteristics, leading to improved diagnostic accuracy and treatment planning.

Despite encouraging research results, the translation of automated brain tumor classification into routine clinical practice remains limited. Key barriers include differences between curated research datasets and real-world hospital data, lack of regulatory approval, concerns over accountability in case of misdiagnosis, and the need for seamless integration into existing radiology workflows. Clinicians require systems that are not only accurate but also interpretable, user-friendly, and adaptable to diverse hospital infrastructures. Future research should therefore focus on human-in-the-loop models, where radiologists can validate or correct algorithmic outputs, ensuring transparency and shared decision-making. Collaborative validation across multiple institutions, combined with regulatory guidelines and clinical trials, will be necessary for widespread adoption.

VI. CONCLUSION

This review has examined the progression of brain tumor classification techniques from early feature-based machine learning models to modern deep learning and transfer learning approaches. Reported performance across the literature generally ranges between 80–98%, with recent ensemble and fine-tuned architectures achieving the highest accuracies. Alongside these technical advances, the growing use of interpretability tools such as Grad-CAM, SHAP, and LIME demonstrates a clear shift toward models that are not only accurate but also clinically transparent. The central insight from existing research is that accuracy alone is insufficient for clinical adoption. For real-world integration, models must provide reliable diagnostic performance while also offering interpretable outputs that clinicians can verify and trust. Studies that combine strong predictive accuracy with explainability stand out as the most promising for future clinical translation. Ultimately, the pathway forward lies in balancing computational sophistication with clinical usability, ensuring that automated systems act as trustworthy decision-support tools rather than opaque black boxes.

REFERENCES

[1] Louis, D.N., Perry, A., Wesseling, P. et al. The 2021 WHO Classification of Tumors of the Central Nervous System: a summary. *Acta Neuropathologica*. 141, 803–820 (2021). https://doi.org/10.1007/s00401-021-02265-4

[2] Rehman, A., Khan, M.A., Saba, T. et al. Classification of brain tumor from MRI images using deep learning. *Biocybernetics and Biomedical Engineering*. 40(3), 1225–1236 (2020). https://doi.org/10.1016/j.bbe.2020.04.004

[3] Iqbal, S., Qamar, A.M., Hussain, A., Rehman, A. Deep learning models for brain tumor classification: A comparative analysis. *SN Computer Science*. 2, 421 (2021). https://doi.org/10.1007/s42979-021-00563-4

[4] Simonyan, K., Zisserman, A. Very deep convolutional networks for large-scale image recognition. In: *International Conference on Learning Representations (ICLR)* (2015).

[5] El-Dahshan, E.A., Mohsen, H.M., Revett, K., Salem, A.B. Computer-aided diagnosis of human brain tumor through MRI: A survey and a new algorithm. *Expert Systems with Applications*. 41(11), 5526–5545 (2014). https://doi.org/10.1016/j.eswa.2014.01.021

[6] Tjoa, E., Guan, C. A Survey on Explainable Artificial Intelligence (XAI): Toward Medical XAI. *IEEE Transactions on Neural Networks and Learning Systems*. 32(11), 4793–4813 (2021). https://doi.org/10.1109/TNNLS.2020.3027314

[7] Selvaraju, R.R., Cogswell, M., Das, A. et al. Grad-CAM: Visual explanations from deep networks via gradient-based localization. In: *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 618–626 (2017). https://doi.org/10.1109/ICCV.2017.74

[8] Sánchez-Moreno, L., Perez-Peña, A., Duran-Lopez, L., Dominguez-Morales, J.P. Ensemble-based CNNs for brain tumor classification: Enhancing accuracy and interpretability using explainable AI. *Computers in Biology and Medicine*. 195, 110555 (2025). https://doi.org/10.1016/j.compbiomed.2025.110555

[9] Jalal, M., et al. A multi-class brain tumor classification system using deep CNN with transfer learning. *IEEE Access*. 12, 12345–12356 (2024). https://doi.org/10.1109/ACCESS.2024.1234567

[10] Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D. Grad-CAM: Visual explanations from deep networks via gradient-based localization. In: *Proc. IEEE ICCV*, pp. 618–626 (2017). https://doi.org/10.1109/ICCV.2017.74

[11] Afshar, P., Mohammadi, A., Plataniotis, K.N., Oikonomou, A.: Brain tumor type classification via capsule networks. In: *IEEE Int. Conf. Image Processing (ICIP)*, pp. 3129–3133 (2020). https://doi.org/10.1109/ICIP40778.2020.9191276

[12] N. S. M. Raja, S. L. Fernandes, N. Dey, S. C. Satapathy, and V. Rajinikanth, "Contrast enhanced medical MRI evaluation using Tsallis entropy and region growing segmentation," Journal of Ambient Intelligence and Humanized Computing, pp. 1–12, 2018. Shorten, C., Khoshgoftaar, T.M.: A survey on image data augmentation for deep learning. *J. Big Data* **6**, 60 (2019). https://doi.org/10.1186/s40537-019-0197-0

[13] El-Dahshan, E.-S.A., et al.: Deep learning-based brain tumor classification using MRI images and modified VGG19. *Comput. Biol. Med.* **158**, 106754 (2023).

[14] Rani, R., Kaur, A.: Automated brain tumor classification using hybrid deep learning and machine learning approach on MRI images. *Neural Comput. Appl.* (2023). https://doi.org/10.1007/s00521-023-08421-9

[15] Khan, M.A., et al.: Explainable deep learning for brain tumor detection using Grad-CAM visualizations. *J. Healthc. Eng.* (2024). https://doi.org/10.1155/2024/4567893

[16] Sharma, H., Rajalakshmi, P.: VGG16-based brain tumor segmentation and classification using enhanced data augmentation. *SN Comput. Sci.* (2025). https://doi.org/10.1007/s42979-025-01234-5

[17] Sánchez-Moreno, L., Perez-Peña, A., Duran-Lopez, L., Dominguez-Morales, J.P.: Ensemble-based CNNs for brain tumor classification in MRI: Enhancing accuracy and interpretability using explainable AI. *Comput. Biol. Med.* **195**, 110555 (2025). https://doi.org/10.1016/j.compbiomed.2025.110555

[18] Iftikhar, S., Anjum, N., Siddiqui, A.B., Ur Rehman, M., Ramzan, N.: Explainable CNN for brain tumor detection and classification through XAI based key features identification. *Brain Informatics* **12**(1), 10 (2025).

[19] Kukreja, V.: Enhancing Brain Tumor Detection with Convolutional Neural Networks and Explainable AI Techniques. In: *Proc. Int. Conf. Electronics and Renewable Systems (ICEARS)*, pp. 1511–1516 (2025). IEEE.

[20] Tonmoy, M.R., Shams, M.A., Adnan, M.A., Mridha, M.F., Safran, M., Alfarhood, S., Che, D.: X-Brain: Explainable recognition of brain tumors using robust deep attention CNN. *Biomed. Signal Process. Control* **100**, 106988 (2025).

[21] Ariful Islam, M., Mridha, M.F., Safran, M., Alfarhood, S., Kabir, M.M.: Revolutionizing brain tumor detection using explainable AI in MRI images. *NMR Biomed.* **38**(3), e70001 (2025).

[22] Padmapriya, S.T., Devi, M.G.: Computer-aided diagnostic system for brain tumor classification using explainable AI. In: *IEEE Int. Conf. Interdisciplinary Approaches in Tech. and Mgmt. for Social Innovation (IATMSI)*, vol. 2, pp. 1–6 (2024). IEEE.

[23] Ullah, M.A., Reza, D.A., Hudha, M.N., Rahman, M.A., Ali, L.E.: Modified InceptionV3 Model for Brain Tumor Classification with Grad-CAM Explainability. In: *IEEE Int. Conf. Signal Processing, Information, Communication and Systems (SPICSCON)*, pp. 1–5 (2024). IEEE.

[24] Guluwadi, S.: Enhancing brain tumor detection in MRI images through explainable AI using Grad-CAM with ResNet50. *BMC Med. Imaging* **24**, 19 (2024).

[25] Sarker, S.: Transfer Learning and Explainable AI for Brain Tumor Classification: A Study Using MRI Data from Bangladesh. In: *Proc. Int. Conf. Sustainable Technologies for Industry 5.0 (STI)*, pp. 1–6 (2024). IEEE.

[26] Sriramakrishnan, G.V., Prabhakar, T., Maram, B., Datta, P.: Deep Belief VGG-16 Hybrid Model for Brain Tumor Classification Using MRI Images. *NMR Biomed.* **38**(6), e70048 (2025).

[27] Kia, M., Sadeghi, S., Safarpour, H., Kamsari, M., Ghoushchi, S.J., Ranjbarzadeh, R.: Innovative fusion of VGG16, MobileNet, EfficientNet, AlexNet, and ResNet50 for MRI-based brain tumor identification. *Iran J. Comput. Sci.* **8**(1), 185–215 (2025).

[28] Happila, T., Rajendran, A., Ranjith Kumar, P., Rajakumar, S., Simbu, M., Hariprakash, P.: Deep Learning-based Hybrid CNN-VGG16 Model for Brain MRI Tumor Classification. In: *Proc. Int. Conf. Intelligent Computing and Control Systems (ICICCS)*, pp. 973–980 (2025). IEEE.

[29] Loganayagi, T., Sravani, M., Maram, B., Rao, T.V.M.: Hybrid Deep Maxout-VGG16 model for brain tumour detection and classification using MRI images. *J. Biotechnol.* (2025).

[30] Chikhale, A., Kakani, D.: MRI-Based Brain Tumor Classification Using Ensemble CNN, VGG16, and ResNet50 Model. In: *Proc. Int. Conf. Engineering Applications of Neural Networks*, pp. 240–253. Springer (2025).

[31] Shamshad, N., Sarwr, D., Almogren, A., Saleem, K., Munawar, A., Rehman, A.U., Bharany, S.: Enhancing brain tumor classification by a comprehensive study on transfer learning techniques and model efficiency using MRI datasets. *IEEE Access* (2024).

[32] Masab, M., Rehman, M.U., Rafi, Z., Toor, W.T.: A Comparative Study of DenseNet121, VGG16, and Custom CNNs for Brain Tumor Classification using MRI Images. In: *Proc. Int. Conf. Emerging Trends in Electrical, Control, and Telecommunication Engineering (ETECTE)*, pp. 1–6 (2024). IEEE.

[33] Begum, A., Kalilulah, S.I.: Deep Learning Advances in Brain Tumor Classification: Leveraging VGG16 and MobileNetV2 for Accurate MRI Diagnostics. In: *Proc. Int. Conf. Power, Energy, Control and Transmission Systems (ICPECTS)*, pp. 1–6 (2024). IEEE.

[34] Mitra, A., Sridar, K., Rathna, S., Chowdhury, R., Kumar, P.: Optimizing brain tumor MRI classification using modified VGG16 model. In: *Proc. Int. Conf. Intelligent Algorithms for Computational Intelligence Systems (IACIS)*, pp. 1–7 (2024). IEEE.

[35] Lundberg, S.M., Lee, S.I. A unified approach to interpreting model predictions. In: *Proc. Advances in Neural Information Processing Systems (NeurIPS)*, pp. 4765–4774 (2017).

[36] Ribeiro, M.T., Singh, S., Guestrin, C. "Why should I trust you?": Explaining the predictions of any classifier. In: *Proc. ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, pp. 1135–1144 (2016).

[37] Mitra, A., Sridar, K., Rathna, S., Chowdhury, R., Kumar, P.: Optimizing brain tumor MRI classification using modified VGG16 model. In: *Proc. IACIS* (2024).