International Journal of Computer Science Trends and Technology (IJCST) – Volume 12 Issue 2, Mar - Apr 2024

RESEARCH ARTICLE OPEN ACCESS

# Exploring the Effectiveness of Deep Learning-based Explainable AI Techniques in Image Modality

**Medisetti sri sarika [1], Madugula Murali Krishna [2]**

[1] PG Scholar, Dept. of CSE, Sri Sivani College of Engineering, Srikakulam, AP, India.
[2] Assistant Professor, Dept. of CSE, Sri Sivani College of Engineering, Srikakulam, AP, India.

## ABSTRACT

Explainable Artificial Intelligence (XAI) holds significant promise in the field of medical imaging, particularly in enhancing the interpretability and trustworthiness of deep learning models. By unveiling previously concealed information, XAI facilitates a more comprehensive understanding of model outputs, crucial for testing and validation. This study introduces a novel approach to diagnosing brain tumors using MRI data, integrating explainable methodologies with classification models. Through deep learning techniques, our framework aims to improve the accuracy of brain tumor detection while enhancing the interpretability of results. The primary focus lies in providing a more thorough interpretation of technologically advanced model outputs, thereby fostering trust among medical professionals. By elucidating the inner workings of these models, we not only increase interpretability but also streamline the diagnostic process, ultimately improving patient care. This holistic approach not only enhances the reliability of deep learning models in medical imaging but also has the potential to revolutionize healthcare delivery, ensuring more successful patient outcomes.

***Keywords***: Explainable Artificial Intelligence (XAI), Medical imaging, Deep learning models, Brain tumor diagnosis Interpretability

## I. INTRODUCTION

In the realm of medical imaging, the integration of artificial intelligence (AI) has ushered in a new era of diagnosis and treatment. Among the various facets of AI, Explainable Artificial Intelligence (XAI) stands out as a pivotal tool in unraveling the complexities of deep learning models, particularly in the context of medical image analysis. This introduction delves into the significance of XAI in enhancing the interpretability of deep learning models, focusing on its application in the diagnosis of brain tumors using magnetic resonance imaging (MRI) data. The structure of brain is shown in Figure 1. The narrative unfolds against the backdrop of burgeoning technological advancements and the pressing need for reliable and interpretable AI solutions in healthcare.

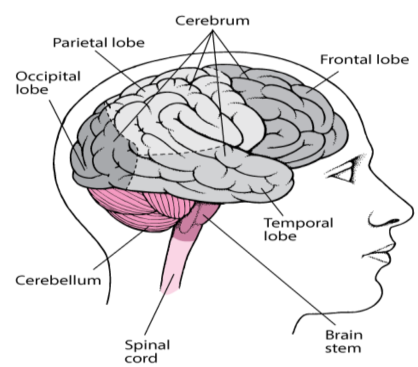


Fig 1: Structure of human brain

Medical imaging serves as a cornerstone in modern healthcare, offering invaluable insights into the human body's intricacies. From X-rays to MRI scans, these imaging modalities provide clinicians with non-invasive means to visualize internal structures, detect abnormalities, and guide therapeutic interventions. While magnetic resonance imaging (MRI) employs various sequences, it is the T1-weighted and T2-weighted sequences that find the most frequent application in clinical imaging (Figure 2). However, the sheer volume and complexity of imaging data present formidable challenges in interpretation, often leading to diagnostic errors and treatment delays. In this context, the integration of AI, particularly deep learning algorithms, has emerged as a transformative force, promising to revolutionize medical imaging practices.

Deep learning, a subset of AI inspired by the structure and function of the human brain's neural networks, has demonstrated remarkable proficiency in analyzing complex datasets, including medical images. By leveraging large-scale datasets and sophisticated algorithms, deep learning models can autonomously extract features, recognize patterns, and make predictions with unprecedented accuracy. Consequently, they hold immense potential in augmenting the capabilities of healthcare providers, enhancing diagnostic accuracy, and optimizing patient care pathways.

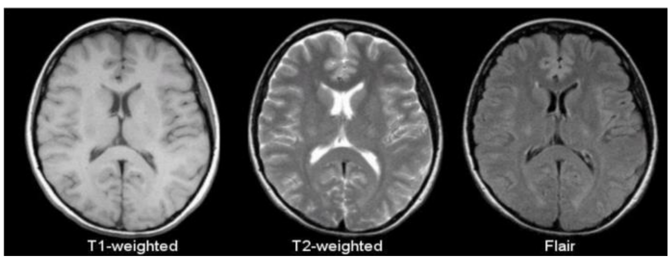


Fig.2 – T1, T2, Flair image

However, despite their undeniable efficacy, deep learning models are often perceived as "black boxes," wherein the decision-making process remains opaque and inscrutable. This inherent lack of transparency poses significant challenges in the healthcare domain, where interpretability and accountability are paramount. Imagine a scenario where a deep learning algorithm detects a brain tumor on an MRI scan. While the prediction may be accurate, the clinician is left grappling with the question: "Why did the algorithm arrive at this conclusion?" Without a clear understanding of the underlying rationale, clinicians may hesitate to trust AI-driven diagnoses, thereby impeding the adoption of these technologies in clinical practice.

Explainable Artificial Intelligence (XAI), a paradigm aimed at demystifying the inner workings of AI systems and elucidating their decision-making processes. Unlike traditional "black box" approaches, XAI techniques strive to provide interpretable explanations for AI-generated outputs, empowering end-users to comprehend, scrutinize, and ultimately trust the AI's recommendations. In the context of medical imaging, XAI assumes heightened significance, offering a pathway towards transparent and accountable AI-driven diagnostics.

The quest for explainability in medical imaging is underscored by the inherent complexity and high-stakes nature of diagnostic decision-making. Consider the diagnosis of brain tumors, a task fraught with challenges due to the diverse morphological characteristics and subtle variations in imaging appearances. While experienced radiologists excel in discerning these nuances, the sheer volume of imaging studies and the ever-increasing workload can strain human cognitive capacities, leading to diagnostic errors and oversights. Herein lies the promise of AI-enabled solutions, which can serve as invaluable decision support tools, augmenting human expertise and mitigating diagnostic errors.

However, the integration of AI into clinical workflows necessitates a paradigm shift in how we perceive and interact with these technologies. Beyond mere accuracy metrics, the trustworthiness, reliability, and interpretability of AI-driven diagnoses emerge as critical determinants of their real-world utility. It is here that XAI assumes a transformative role, bridging the gap between AI algorithms and end-users by providing intuitive and transparent explanations for model predictions.

Against this backdrop, our research endeavors focus on harnessing the synergistic potential of deep learning and XAI to advance the field of medical imaging, with a specific emphasis on the diagnosis of brain tumors using MRI data. By developing novel frameworks that combine state-of-the-art classification models with explainable methodologies, we aim to enhance the accuracy and interpretability of AI-driven brain tumor diagnostics. Through rigorous experimentation and validation, we seek to demonstrate the efficacy and reliability of our proposed approach, paving the way for its integration into clinical practice.

Central to our research is the belief that transparency breeds trust, and interpretability fosters adoption. By unraveling the intricacies of deep learning models and providing clinicians with actionable insights into AI-generated diagnoses, we aim to empower healthcare providers to make informed decisions, optimize treatment pathways, and ultimately improve patient outcomes. Our overarching goal is not merely to develop technologically advanced solutions but to bridge the gap between AI innovation and clinical utility, ushering in a new era of AI-enabled healthcare delivery.

## II. RESEARCH GAP

Despite the rapid advancements in medical imaging technology and artificial intelligence (AI), there exists a conspicuous research gap in the integration of Explainable Artificial Intelligence (XAI) with deep learning models for the diagnosis of brain tumors using MRI data. While deep learning algorithms have demonstrated remarkable efficacy in automated image analysis, their "black box" nature impedes their widespread adoption in clinical practice. The lack of interpretability and transparency poses significant challenges in the validation, trustworthiness, and regulatory approval of AI-driven diagnostic systems. Consequently, there is a pressing need to bridge this gap by developing novel frameworks that not only achieve high diagnostic accuracy but also provide intuitive explanations for model predictions.

Existing literature in the field of medical imaging predominantly focuses on the development and evaluation of deep learning models for various diagnostic tasks, including brain tumor detection. While these studies report promising results in terms of sensitivity and specificity, they often fall short in addressing the interpretability and explainability aspects of AI-driven diagnoses. Few studies have attempted to integrate XAI techniques with deep learning models in medical imaging, and even fewer have specifically targeted brain tumor diagnosis using MRI data. Thus, the research gap lies in the intersection of deep learning, XAI, and medical imaging, particularly in the context of brain tumor detection, where the interpretability of AI-driven diagnoses is of paramount importance.

## III. CONCEPTUAL FRAMEWORK

The conceptual framework of this study revolves around the integration of deep learning algorithms with XAI techniques for the diagnosis of brain tumors using MRI data. At its core, the framework comprises two interrelated components: the deep learning model and the XAI methodology. The deep learning model serves as the computational backbone, leveraging convolutional neural networks (CNNs) or other advanced architectures to extract meaningful features from MRI images and make diagnostic predictions. Concurrently, the XAI methodology complements the deep learning model by providing transparent and interpretable explanations for the model's predictions.

Within this framework, several key processes unfold. Firstly, the deep learning model is trained on a large dataset of labeled MRI scans, where it learns to discern patterns indicative of brain tumors. This training phase involves data preprocessing, model architecture selection, hyperparameter tuning, and optimization to maximize performance metrics such as accuracy, sensitivity, and specificity. Subsequently, the trained model undergoes evaluation using a separate test dataset to assess its generalization capabilities and diagnostic accuracy.

Simultaneously, the XAI methodology is employed to generate explanations for the deep learning model's predictions. This may involve post-hoc analysis techniques such as gradient-based attribution methods, saliency maps, or attention mechanisms, which highlight the regions of the input image that contribute most to the model's decision-making process. These explanations are then presented to end-users, typically radiologists or clinicians, to aid in their interpretation of the AI-generated diagnoses.

## IV. SCOPE OF THE STUDY

The scope of this study encompasses the development, implementation, and evaluation of a novel framework for the diagnosis of brain tumors using MRI data, integrating deep learning models with XAI techniques. The study focuses specifically on gliomas, the most common type of primary brain tumor, known for their diverse morphological characteristics and imaging appearances. MRI scans, including T1-weighted, T2-weighted, and contrast-enhanced sequences, serve as the primary modality for tumor detection and characterization.

The dataset used for model training and evaluation comprises anonymized MRI scans collected from multiple medical institutions, encompassing a diverse range of patient demographics, tumor types, and imaging protocols. The deep learning model is trained to classify MRI images into binary categories: tumor versus non-tumor, with an emphasis on achieving high diagnostic accuracy and robust generalization across different imaging modalities and patient cohorts.

The XAI methodology employed in this study focuses on generating interpretable explanations for the deep learning model's predictions, with the aim of enhancing the trustworthiness and interpretability of AI-driven diagnoses. Various XAI techniques, including gradient-based attribution methods, occlusion analysis, and attention mechanisms, are explored to elucidate the features and patterns influencing the model's decision-making process.

The evaluation of the proposed framework encompasses quantitative metrics such as accuracy, sensitivity, specificity, and area under the receiver operating characteristic curve (AUC-ROC), as well as qualitative assessments of the interpretability and clinical utility of the XAI-generated explanations. The study also includes comparative analyses with existing deep learning models and traditional machine learning algorithms to assess the incremental value of XAI in improving diagnostic performance and interpretability.

## V. OBJECTIVES

- Conduct an in-depth review of existing literature on deep learning, explainable AI, and image analysis techniques.
- Develop a comprehensive framework for integrating XAI methods into deep learning architectures for image analysis.
- Implement the explored XAI techniques and integrate them into existing deep learning models.
- Train deep neural networks on image datasets representing a real-world scenario.
- Evaluate the performance of XAI techniques in enhancing interpretability by measuring the quality of explanations generated.
- Compare the effectiveness of different XAI techniques in providing meaningful explanations for deep learning model predictions.
- Apply the developed XAI techniques to real-world image analysis tasks, such as medical image diagnosis or autonomous vehicle perception.

By achieving these objectives, the project endeavours to advance the field of explainable AI in image analysis, fostering a deeper understanding of complex neural networks and their applications. The outcomes of this research have the potential to enable safer, more reliable, and widely accepted AI systems for image-based tasks across various industries.

## VI. HYPOTHESIS

Based on the aforementioned research gap and conceptual framework, the following hypotheses are proposed:

1. Hypothesis 1: The integration of Explainable Artificial Intelligence (XAI) techniques with deep learning models enhances the interpretability and transparency of AI-driven diagnoses for brain tumor detection using MRI data.
2. Hypothesis 2: The proposed framework, combining deep learning algorithms with XAI methodologies, achieves superior diagnostic accuracy and robustness compared to conventional deep learning models and traditional machine learning approaches.
3. Hypothesis 3: XAI-generated explanations aid radiologists and clinicians in understanding the decision-making process of AI-driven diagnostic systems, thereby fostering trust, confidence, and acceptance of AI technologies in clinical practice.

## VII. RESEARCH METHODOLOGY

The brain tumor detection system is designed with key non-functional requirements to ensure its effectiveness, scalability, reliability, compatibility, maintainability, and regulatory compliance. Here's a summary:

Performance**:** The system can analyze and categorize medical images within 5 seconds under normal conditions. It can process at least 10 images concurrently without a noticeable slowdown, allowing for quick analysis even with high workloads.

Scalability**:** The architecture supports horizontal scaling, allowing for additional servers or computing nodes to accommodate increasing data volumes. This scalability enables a 15% yearly growth rate over the next three years. Horizontal scaling also improves fault tolerance and resilience, with the ability to handle increased workloads and system failures.

Reliability: The system maintains a high availability rate of at least 95%, even during maintenance. Robust error-handling minimizes false positives and negatives, contributing to dependable and accurate detection outcomes.

Compatibility: The system is compatible with all versions of operating systems used in medical imaging devices, allowing it to operate across various platforms without compatibility issues.

Maintainability**:** The codebase is thoroughly documented, following recognized coding standards. This documentation ensures easy collaboration among developers and supports seamless updates or enhancements. The design allows for integration of new features without extensive reengineering, facilitating system evolution in line with advancing detection techniques.

Regulatory Compliance**:** The research and development of the system strictly follow ethical guidelines and data protection regulations. This adherence ensures confidentiality and compliance with institutional policies, reinforcing trust and promoting responsible research practices.

The proposed brain tumor detection system uses Convolutional Neural Networks (CNNs) and the architecture is shown in Figure 3 and Figure 4 which is used to analyze MRI images and determine whether they contain brain tumors. The process begins with partitioning a dataset of 7023 brain MRI images into training and testing sets, using an 8:2 ratio. The images are resized to 150x150 pixels to ensure consistency. The system's methodology involves two main phases: Feature Extraction and Classification.

In the Feature Extraction phase, convolutional, pooling, and dropout layers are used to identify key features in the dataset. This phase uses Conv2D layers with increasing filters and max-pooling to reduce spatial dimensions. The Classification phase involves mapping these features to a set of class values to create a classifier that can categorize future instances. Dense layers evaluate the features' effectiveness in distinguishing between brain tumor cases and healthy tissue.

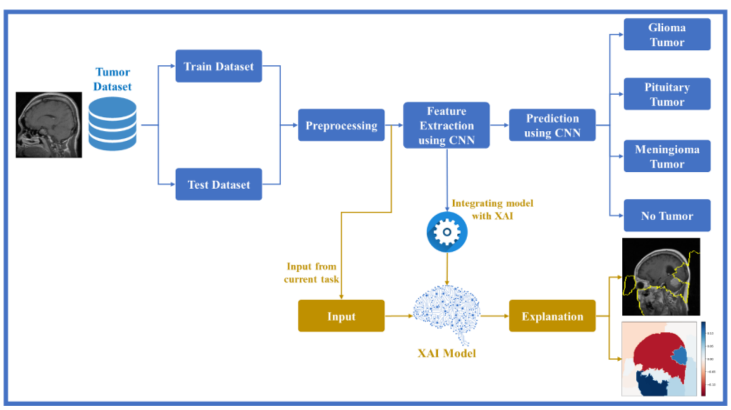


Fig 3: Architecture overview

The system uses Local Interpretable Model-Agnostic Explanations (LIME) to provide insights into its decision-making process. LIME generates local explanations for individual predictions made by the CNN, offering transparency and aiding in the interpretability of the model's output. This process involves selecting an instance, creating perturbations, recording predictions, fitting an interpretable model, assigning weights, and generating explanations to understand why the CNN made a specific prediction.

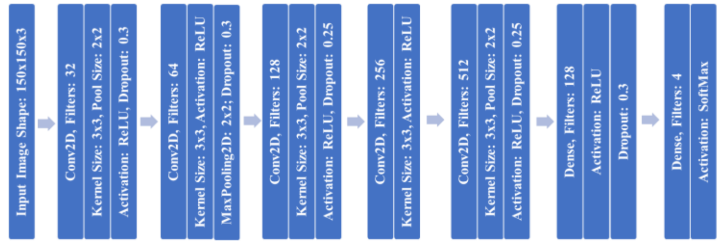


Fig 4: Proposed convolutional neural network architecture

The CNN's architecture includes five hidden Conv2D layers with varying filters, interspersed with max-pooling layers. A fully connected Dense layer with 128 neurons and dropout regularization reduces overfitting, while the output layer has four neurons to classify the four classes: glioma, meningioma, no tumor, and pituitary. The Rectified Linear Unit (ReLU) activation function is used throughout to introduce non-linearity and enhance the network's learning capability.

The formula for generating explanations using LIME can be expressed as:

$$\hat{f}_{(x)} = arg\ min_{g \in G} L(f, g, \pi_x) + \Omega(g)$$

Where:

- $\hat{f}_{(x)}$ is the interpretable model's prediction for the instance $x$.
- $f$ is the black-box model.
- $g$ is the interpretable model being trained (e.g., linear model, decision tree).
- $L(f, g, \pi_x)$ is a loss function that measures how well $g$ approximates $f$ in the local neighborhood of $x$, weighted by $\pi_x$.

- $\Omega(g)$ is a regularization term that penalizes complex models to promote simplicity and interpretability.

In summary, LIME is a versatile and powerful method for explaining the predictions of complex machine learning models. By providing local explanations that are both interpretable and faithful to the black-box model, LIME enables users to gain valuable insights into the decision-making processes of these models, contributing to greater transparency and trust in AI systems.

Overall, the proposed system integrates advanced CNN architectures with explainable AI methods to improve brain tumor detection's accuracy and transparency. It also aligns with the growing emphasis on explainable AI in medical imaging, making it a promising tool for both clinical and research applications in brain tumor diagnostics.

## VIII. RESULTS AND ANALYSIS

A common practice for evaluating machine learning algorithms involves splitting a dataset into three parts: training, validation, and test sets. Typically, this split follows a 65% training, 20% validation, and 15% test ratio. Randomizing the data before partitioning helps ensure each set reflects the overall data distribution, preventing biased results due to data leakage.

The training dataset is used to teach the machine learning model patterns and relationships within the data. The validation dataset assesses the model's performance and helps fine-tune hyperparameters, reducing the risk of overfitting. The test dataset serves as an independent set to evaluate the final model's performance, ensuring it can generalize to new data. The performance comparison is depicted in table 1.

Table 1 Performance comparison

| Model \ Metric | Precision | Recall | F1-Score | Accuracy | MSE | RMSE |
|---|---|---|---|---|---|---|
| InceptionV3 | 0.81192 | 0.78593 | 0.77257 | 0.78593 | 0.44954 | 0.67048 |
| VGG 16 | 0.86984 | 0.84839 | 0.85091 | 0.84839 | 0.44717 | 0.66871 |
| ResNet50 | 0.84839 | 0.86371 | 0.86341 | 0.86371 | 0.28790 | 0.53656 |
| **Proposed XDCNN** | **0.95124** | **0.95100** | **0.95106** | **0.95100** | **0.14701** | **0.38342** |

Among the evaluated models, the proposed XDCNN stands out as the top performer across all metrics. With a precision of approximately 95.12%, it demonstrates the highest accuracy in correctly identifying positive predictions, surpassing all other models. Furthermore, the XDCNN model achieves an impressive recall rate of around 95.10%, indicating its ability to capture nearly all actual positive samples. This exceptional balance between precision and recall is reflected in its F1-Score of approximately 95.11%, the highest among all models, suggesting a robust overall performance. Moreover, the XDCNN model exhibits outstanding accuracy, with approximately 95.10% of instances correctly classified, showcasing its efficacy in making accurate predictions. Additionally, the proposed XDCNN model boasts the lowest Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) values, signifying superior predictive capability with minimal errors in regression tasks. Overall, the proposed XDCNN model emerges as the most effective and reliable choice for the given task, outperforming established models like InceptionV3, VGG 16, and ResNet50 in various performance metrics. After training, the model is evaluated with the validation set to guide improvements or adjustments. The test set is exclusively for evaluating the model's generalization capability, not for training or validation. This approach ensures unbiased results by keeping the datasets independent.

The best model is chosen based on accuracy and low error rates from the validation dataset. This model undergoes final testing with the test dataset to confirm its reliability and generalization. If a model performs well on the test dataset, it's crucial to ensure no data leakage occurred from validation or test sets into the training set. Leakage could inflate the model's performance metrics, leading to inaccurate evaluations.

For brain tumor classification, the IEEE Dataport Brain Tumor dataset is used, with four classes: glioma, meningioma, no tumor, and pituitary tumor. While various classification models and methods exist, complex models often increase computational overhead. The proposed Explainable Deep Convolutional Neural Network (XDCNN) prioritizes simplicity and efficiency, offering higher accuracy and lower error rates with fewer layers than other models, reducing computational costs.

Explainable AI, specifically LIME, is integrated into the framework to enhance interpretability. Explainable AI results for sample input images is shown in Figure 5.

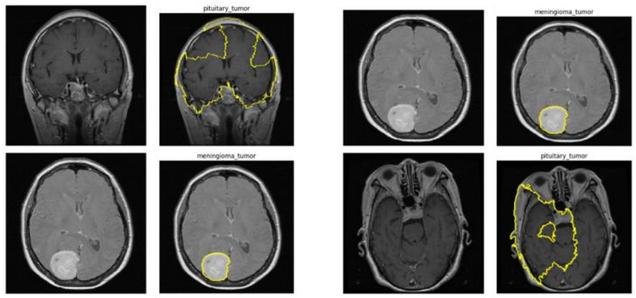


Fig 5– Explainable AI results for sample input images

This technique provides insights into the model's decision-making, crucial for medical applications. The results demonstrate improved accuracy and reduced error rates, along with detailed visualization for better tumor localization and assessment in MRI scans. The explainable AI approach assists clinicians in surgical planning and enhances diagnostic accuracy by detecting subtle features that may be missed by the human eye. This methodology contributes to better patient care and effective treatment strategies.

## IX. CONCLUSION

The proposed Explainable Deep Convolutional Neural Network (XDCNN) model for brain tumor detection has demonstrated significant promise in terms of performance, efficiency, and interpretability. The study's results indicate that the model offers high classification accuracy with reduced computational overhead compared to more complex models. This is achieved through a streamlined architecture that uses fewer layers while maintaining a high level of accuracy. The inclusion of explainable AI techniques like Local Interpretable Model-Agnostic Explanations (LIME) allows clinicians and researchers to understand the reasoning behind the model's predictions, which is particularly crucial in medical applications where accuracy and transparency are paramount.

The study utilized the IEEE Dataport Brain Tumor dataset to evaluate the XDCNN model's effectiveness in classifying MRI images into four categories: glioma, meningioma, no tumor, and pituitary tumor. The model's ability to deliver high performance while reducing Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) indicates its robustness in identifying brain tumors. Additionally, the study underscores the importance of maintaining separate training, validation, and test datasets to ensure unbiased results and reliable generalization.

The explainable AI aspect of the study provided valuable insights into the model's decision-making process. This feature enables medical professionals to validate the model's predictions, facilitating surgical planning and enhancing diagnostic accuracy. The study concludes that the proposed model's combination of high accuracy, reduced complexity, and interpretability can contribute to improved brain tumor detection, ultimately leading to more effective patient care and treatment strategies.

## X. LIMITATIONS

Despite the promising results achieved by the XDCNN model, there are several limitations to consider. Firstly, the dataset used in the study, although substantial, may not fully represent the diversity of real-world medical images. This limitation could affect the model's ability to generalize to different clinical settings, where variations in image quality, equipment, and patient demographics are common. As such, further studies should incorporate larger and more diverse datasets to ensure the model's robustness across various contexts.

Another limitation is the computational constraints of the study's hardware. Although the XDCNN model aims to reduce computational overhead, the computational resources required for training and evaluation could still be substantial, particularly in large-scale clinical environments. This limitation could hinder the model's scalability and integration into existing medical systems without adequate infrastructure.

Additionally, the explainable AI component, while offering valuable insights into the model's decision-making, may not always provide complete transparency. LIME, for instance, generates local explanations that might not capture the broader context of the model's predictions. This limitation could affect clinicians' ability to fully understand the underlying mechanisms of the model, leading to potential misinterpretations.

Lastly, the study's scope is focused on brain tumor detection. The model's adaptability to other medical imaging tasks or applications outside of brain tumor diagnosis has not been explored. This limitation restricts the study's applicability to a broader range of medical contexts, suggesting the need for further research to expand the model's utility.

## XI. IMPLICATIONS

The implications of this study are significant for both the medical and machine learning communities. For the medical field, the proposed XDCNN model presents an opportunity to enhance brain tumor detection's accuracy and reliability, leading to better patient outcomes. The model's high accuracy and explainable AI features can improve clinicians' confidence in automated diagnostic tools, fostering a more collaborative approach between human expertise and machine learning algorithms.

The reduced computational overhead and efficient architecture of the XDCNN model suggest that medical facilities can integrate this technology without significant disruptions to existing workflows. This efficiency could lead to faster diagnoses, allowing healthcare professionals to make timely treatment decisions, ultimately improving patient care.

From a research perspective, the study's emphasis on explainable AI addresses the growing need for transparency in machine learning models. By demonstrating that advanced classification models can be both accurate and interpretable, the study contributes to the ongoing discussion about the ethical use of AI in healthcare. It also underscores the importance of rigorous evaluation methods, emphasizing the need to maintain separate datasets to avoid biased results.

The study's limitations, particularly regarding data diversity and computational constraints, highlight areas for future research. These implications suggest that subsequent studies should focus on expanding the dataset and exploring the model's scalability to ensure its applicability across diverse medical settings. Additionally, researchers could investigate the potential of applying the XDCNN model to other medical imaging tasks, extending its benefits beyond brain tumor detection

## XII. REFERENCES

[1] Jemal, A.; Thomas, A.; Murray, T.; Thun, M. Cancer statistics. Ca-Cancer J. Clin. 2002, 52, 23–47.
[2] Miner, R.C. Image-guided neurosurgery. J. Med. Imaging Radiat. Sci. 2017, 48, 328–335. [CrossRef]
[3] Isensee, F.; Kickingereder, P.; Wick, W.; Bendszus, M.; Maier-Hein, K.H. Brain tumor segmentation and radiomics survival prediction: Contribution to the brats 2017 challenge. In International MICCAI Brainlesion Workshop; Springer: Berlin/Heidelberg, Germany, 2017; pp. 287–297.
[4] Yang, G.; Ye, Q.; Xia, J. Unbox the black-box for the medical explainable AI via multi-modal and multi-centre data fusion: A mini-review, two showcases and beyond. Inf. Fusion 2022, 77, 29–52. [CrossRef] [PubMed]
[5] Adadi, A.; Berrada, M. Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). IEEE Access 2018, 6, 52138–52160. [CrossRef]
[6] Gunning, D.; Stefik, M.; Choi, J.; Miller, T.; Stumpf, S.; Yang, G.-Z. XAI—Explainable artificial intelligence. Sci. Robot. 2019, 4, eaay7120. [CrossRef] [PubMed]
[7] Jayakrishna, M., M. Vijay, and Baseem Khan. "An Overview of Extensive Analysis of 3D Printing Applications in the Manufacturing Sector." Journal of Engineering 2023 (2023).
[8] Tonekaboni, S.; Joshi, S.; McCradden, M.D.; Goldenberg, A. What clinicians want: Contextualizing explainable machine learning for clinical end use. In Proceedings of the Machine Learning for Healthcare Conference, Ann Arbor, MI, USA, 8–10 August 2019; pp. 359–380.
[9] Messina, P.; Pino, P.; Parra, D.; Soto, A.; Besa, C.; Uribe, S.; Andía, M.; Tejos, C.; Prieto, C.; Capurro, D. A survey on deep learning and explainability for automatic report generation from medical images. ACM Comput. Surv. 2022, 54, 1–40. [CrossRef]
[10] Temme, M. Algorithms and transparency in view of the new general data protection regulation. Eur. Data Prot. Law Rev. 2017, 3, 473–485. [CrossRef]
[11] Jawarneh, M., Jayakrishna, M., Davuluri, S. K., Ramanan, S. V., Singh, P. P., & Joseph, J. A. (2023, February). Energy Efficient Lightweight Scheme to Identify Selective Forwarding Attack on Wireless Sensor Networks. In International Conference on Intelligent Computing and Networking (pp. 425-436). Singapore: Springer Nature Singapore.
[12] Zeiler, M.D.; Fergus, R. Visualizing and understanding convolutional networks. In Proceedings of the European Conference on Computer Vision, Zurich, Switzerland, 6–12 September 2014; pp. 818–833.
[13] Sundararajan, M.; Taly, A.; Yan, Q. Axiomatic attribution for deep networks. In Proceedings of the International Conference on Machine Learning, Sydney, Australia, 6–11 August 2017; pp. 3319–3328.
[14] Simonyan, K.; Vedaldi, A.; Zisserman, A. Deep inside convolutional networks: Visualising image classification models and saliency maps. arXiv 2013, arXiv:1312.6034.
[15] Springenberg, J.T.; Dosovitskiy, A.; Brox, T.; Riedmiller, M. Striving for simplicity: The all convolutional net. arXiv 2014, arXiv:1412.6806.
[16] Jayakrishna, M., et al. "Multi-scale Memory Residual Network Based Deep Learning Model for Network Traffic

Anomaly Detection." International Conference on Intelligent Computing and Networking. Singapore: Springer Nature Singapore, 2023.
[17] 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 Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 618–626.
[18] Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015; pp. 234–241.
[19] Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. arXiv 2014, arXiv:1409.1556.
[20] He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
[21] Jayakrishna, M., M. Vijay, and Baseem Khan. "An Overview of Extensive Analysis of 3D Printing Applications in the Manufacturing Sector." Journal of Engineering 2023 (2023).
[22] Tian, J.; Li, C.; Shi, Z.; Xu, F. A diagnostic report generator from CT volumes on liver tumor with semi-supervised attention mechanism. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Granada, Spain, 16–20 September 2018; pp. 702–710.
[23] Han, Z.;Wei, B.; Leung, S.; Chung, J.; Li, S. Towards automatic report generation in spine radiology using weakly supervised framework. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Granada, Spain, 16–20 September 2018; pp. 185–193.
[24] Teixeira, L.O.; Pereira, R.M.; Bertolini, D.; Oliveira, L.S.; Nanni, L.; Cavalcanti, G.D.; Costa, Y.M. Impact of lung segmentation on the diagnosis and explanation of COVID-19 in chest X-ray images. Sensors 2021, 21, 7116. [CrossRef]
[25] Ramzan, F.; Khan, M.U.G.; Iqbal, S.; Saba, T.; Rehman, A. Volumetric segmentation of brain regions from MRI scans using 3D convolutional neural networks. IEEE Access 2020, 8, 103697–103709. [CrossRef]
[26] Babu, V. Ram, M. Jaya Krishna, and A. Lakshumu Naidu. "Tribological Behaviour of Biodiesel and Metal Oxide Nanoparticles as Alternative Lubricant: A Pin-on-Disc Tribometer and Wear Study." Journal of Positive School Psychology (2022): 2066-2074.
[27] Yang, C.; Rangarajan, A.; Ranka, S. Visual explanations from deep 3D convolutional neural networks for Alzheimer's disease classification. In Proceedings of the AMIA Annual Symposium Proceedings, San Francisco, CA, USA, 3–7 November 2018; pp. 1571–1580.
[28] Wickstrøm, K.; Kampffmeyer, M.; Jenssen, R. Uncertainty and interpretability in convolutional neural networks for semantic segmentation of colorectal polyps. Med. Image Anal. 2020, 60, 101619. [CrossRef] [PubMed]
[29] Esmaeili, M.; Vettukattil, R.; Banitalebi, H.; Krogh, N.R.; Geitung, J.T. Explainable artificial intelligence for human-machine interaction in brain tumor localization. J. Pers. Med. 2021, 11, 1213. [CrossRef]
[30] Saleem, H.; Shahid, A.R.; Raza, B. Visual interpretability in 3D brain tumor segmentation network. Comput. Biol. Med. 2021, 133, 104410. [CrossRef]
[31] Natekar, P.; Kori, A.; Krishnamurthi, G. Demystifying brain tumor segmentation networks: Interpretability and uncertainty analysis. Front. Comput. Neurosci. 2020, 14, 6. [CrossRef] [PubMed]
[32] Adebayo, J.; Gilmer, J.; Muelly, M.; Goodfellow, I.; Hardt, M.; Kim, B. Sanity checks for saliency maps. Adv. Neural Inf. Process. Syst. 2018, 31, 9505–9515.
[33] Pereira, S.; Meier, R.; Alves, V.; Reyes, M.; Silva, C.A. Automatic brain tumor grading from MRI data using convolutional neural networks and quality assessment. In Understanding and Interpreting Machine Learning in Medical Image Computing Applications; Springer: Berlin/Heidelberg, Germany, 2018; pp. 106–114.
[34] Narayanan, B.N.; De Silva, M.S.; Hardie, R.C.; Kueterman, N.K.; Ali, R. Understanding deep neural network predictions for medical imaging applications. arXiv 2019, arXiv:1912.09621.
[35] Isensee, F.; Jäger, P.; Full, P.; Vollmuth, P.; Maier-Hein, K. nnU-Net for Brain Tumor Segmentation in Brainlesion: Glioma. In Proceedings of the Multiple Sclerosis, Stroke and Traumatic Brain Injuries-6th InternationalWorkshop, BrainLes, Lima, Peru, 4 October 2020.
[36] Yan, F.; Wang, Z.; Qi, S.; Xiao, R. A Saliency Prediction Model Based on Re-Parameterization and Channel Attention Mechanism. Electronics 2022, 11, 1180. [CrossRef]
[37] Ding, X.; Zhang, X.; Ma, N.; Han, J.; Ding, G.; Sun, J. Repvgg: Making vgg-style convnets great again. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 19–20 June 2021; pp. 13733–13742.
[38] Ding, X.; Chen, H.; Zhang, X.; Huang, K.; Han, J.; Ding, G. Re-parameterizing Your Optimizers rather than Architectures. arXiv 2022, arXiv:2205.15242.
[39] Krizhevsky, A.; Hinton, G. Learning Multiple Layers of Features from Tiny Images. Master's Thesis, University of Tront, Toronto, ON, Canada, 2009.
[40] Chattopadhay, A.; Sarkar, A.; Howlader, P.; Balasubramanian, V.N. Grad-cam++: Generalized gradient-

based visual explanations for deep convolutional networks. In Proceedings of the 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Tahoe, NV, USA, 12–15 March 2018; pp. 839–847.

[41] Bakas, S.; Reyes, M.; Jakab, A.; Bauer, S.; Rempfler, M.; Crimi, A.; Shinohara, R.T.; Berger, C.; Ha, S.M.; Rozycki, M. Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge. arXiv 2018, arXiv:1811.02629.

[42] Ge, C.; Gu, I.Y.-H.; Jakola, A.S.; Yang, J. Deep learning and multi-sensor fusion for glioma classification using multistream 2D convolutional networks. In Proceedings of the 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Honolulu, HI, USA, 18–21 July 2018; pp. 5894–5897.

[43] Rehman, A.; Khan, M.A.; Saba, T.; Mehmood, Z.; Tariq, U.; Ayesha, N. Microscopic brain tumor detection and classification using 3D CNN and feature selection architecture. Microsc. Res. Tech. 2021, 84, 133–149.

[44] Dixit, A.; Nanda, A. An improved whale optimization algorithm-based radial neural network for multi-grade brain tumor classification. Vis. Comput. 2022, 38, 3525–3540.