Automated Diagnosis of Brain Pathologies Using Deep Learning and CT Images

Dr. David Raj Micheal
Department of Mathematics
School of Advanced Sciences
Vellore Institute of Technology Chennai
Tamil Nadu - 600127
davidraj.micheal@vit.ac.in

Jeeva D
Department of Mathematics
School of Advanced Sciences
Vellore Institute of Technology Chennai
Tamil Nadu - 600127
jeeva.2023@vitstudent.ac.in

Abstract— Deep learning has enabled the development of sophisticated systems for medical diagnosis, particularly in detecting critical brain conditions using imaging data. This paper focuses on designing a deep learning-based model to automatically diagnose brain pathologies such as tumors, cancer, and aneurysms using CT scan images. The dataset for this project is sourced from Kaggle and contains CT images labeled with various brain pathologies, providing a comprehensive foundation for training machine learning models. The initial stages of the project involve collecting and preprocessing the dataset, including resizing images, normalization, and augmentation to enhance model generalization. Exploratory data analysis is conducted to gain insights into the dataset's distribution and the visual characteristics of different pathologies. Following this, several deep learning models, such as VGG16, ResNet50, and EfficientNetB7, are implemented and fine-tuned to classify the brain abnormalities accurately. Each model is trained and evaluated on separate training, validation, and test datasets to measure performance metrics such as accuracy, precision, recall, and F1-score. The ultimate goal of this project is to build a robust and reliable tool that can assist healthcare professionals in making faster and more accurate diagnoses of critical brain conditions, ultimately improving patient outcomes.

Index Terms— **Deep Learning; Medical Imaging; VGG16; ResNet50; EfficientNetB7**

I. INTRODUCTION

The increasing availability of medical imaging data, particularly computed tomography (CT) scans, has opened new avenues for improving diagnostic accuracy in brain pathologies through the use of deep learning. Brain conditions such as tumors, cancer, aneurysms, and other structural abnormalities present significant diagnostic challenges, often requiring specialized expertise and manual interpretation by radiologists.However, with the rapid progress in artificial intelligence, deep learning has emerged as a powerful tool for medical image analysis. These techniques enable the development of systems that can independently analyze CT scans and identify brain abnormalities without the need for human intervention. This shift not only accelerates the diagnostic process but also enhances the consistency and reliability of medical evaluations.

The integration of deep learning into the medical field not only enhances diagnostic precision but also addresses the growing demand for efficient healthcare solutions. Automated systems can assist radiologists by flagging potential areas of concern, reducing the time required for analysis, and minimizing human error. Moreover, these systems can operate at scale, making high-quality diagnostic tools accessible even in resource-limited settings. This project explores the application of deep learning techniques in the automated diagnosis of brain pathologies using CT images, aiming to develop models that can accurately identify and classify various brain conditions, ultimately contributing to more timely and effective treatments.

II. OBJECTIVE

The objective of this project is to develop and apply deep learning algorithms for the automated diagnosis of brain pathologies, specifically tumors, cancer, and aneurysms, using CT scan images. The goal is to enhance the accuracy and efficiency of medical diagnoses to assist healthcare professionals in patient treatment.

III. REVIEW LITERATURE

Seifedine Kadry, Yunyoung Nam, Hafiz Tayyab Rauf, Venkatesan Rajinikanth, Isah A. Lawal (2021) significantly impacted medical image classification, particularly in brain MRI analysis. Studies have explored the use of pre-trained deep learning models like VGG16, VGG19, and ResNet50, known for their ability to extract intricate features from images. These models have been widely adopted for classifying various brain conditions, such as normal, stroke, Low-Grade Glioma (LGG), and High-Grade Glioma (HGG), by resizing MRI slices to standard dimensions for input consistency. Research shows that combining these models with classifiers like SoftMax, SVM-RBF, and SVM-Cubic enhances classification accuracy. SoftMax remains a default choice, while SVM variants often provide superior performance in complex multi-class problems. This combination of deep learning architectures and advanced classifiers has shown promising results in improving diagnostic accuracy, helping automate brain pathology detection and classification[1].

Puttagunta, M.,& Ravi, S. (2021) deep learning approaches in medical image analysis, numerous studies have demonstrated the growing significance of deep learning (DL) techniques, particularly convolutional neural networks (CNN), for various medical imaging tasks such as classification, detection, and segmentation. The integration of deep learning has enabled advancements in automated diagnosis, reducing the workload of radiologists while improving diagnostic accuracy. The field has evolved from early machine learning methods requiring manual feature extraction to advanced DL models capable of learning features directly from raw data. Applications of these techniques span across different imaging modalities including X-rays, CT scans, mammograms, and histopathology images. Research highlights the effectiveness of DL in detecting diseases such as tuberculosis, lung cancer, and breast cancer, with CNNs being particularly useful in pattern recognition and image classification. The review literature emphasizes the need for continuous exploration of deep learning techniques to address challenges like data scarcity and interpretation complexity in medical imaging[2].

Khumancha, M. B., Barai, A., & Rao, C. R. (2019) Lung cancer detection using low-dose CT scans has been an area of significant research, with particular focus on detecting pulmonary nodules, which are potential indicators of cancer. The LUNA16 dataset, comprising 888 CT scans with annotated nodules, has been widely utilized to de- velop detection algorithms. Researchers typically extract 32 × 32 × 32 voxel cubes centered on these nodules to train a 3D Convolutional Neural Network(CNN) for nodule detection. Following this, datasets like the Data Science Bowl 2017 from Kaggle, which contains 1,595 CT scans, are used to further assess the presence of cancer. The predicted nodule locations are used to generate similarly sized cubes, and a second 3D CNN is employed to classify cancer. Studies implementing this two-step detection and classification method, combined with image processing techniques for lung ROI extraction, have reported promising results, achieving precision and recall of 89.24 percent and 82.17 perce, respectively[3].

min, J., Sharif, M., Haldorai, A., Yasmin, M., & Nayak, R. S. (2022) brain tumor detection and classification highlights the rapid advancements in machine learning, specifically deep learning techniques, which have significantly improved the accuracy and efficiency of medical image analysis. Various approaches, such as convolutional neural networks (CNN), transfer learning, and hybrid models, have been applied to detect and classify brain tumors from MRI scans. These models have shown promising results, particularly in challenging cases involving tumors of varying sizes and locations. Techniques like image preprocessing, segmentation, feature extraction, and classification have been essential for enhancing performance. However, there are limitations, such as dealing with noise in MRI scans and segmenting complex tumor structures likegliomas. Furthermore, the accuracy of deep learning models is often dependent on the quality of the training datasets, which vary in terms of resolution and intensity. As a result, future research should focus on lightweight, high-accuracy models, the fusion of handcrafted and deep learning features, and the use of real patient data to improve robustness[4]. Shakeel, P. M., Tobely, T. E. E., Al-Feel, H., Manogaran, G., & Baskar, S. (2019) Various meth- ods for brain tumor detection have been explored, including the use of neural networks and infrared imaging sensors.The incorporation of feature extraction techniques like Gray Level covariance Matrix (GLCM) and Principal Component Analysis (PCA) has shown promising results in improving classification accuracy.Researchers have focused on enhancing the division process of MRI brain images by considering anatomical structures and utilizing advanced algorithms like Convolutional Neural Networks (CNN) for core segmentation.The utilization of multifractal features and Ada Boosting algorithm has been proposed for tumor division, showing improved results in terms of accuracy and stability.The integration of wireless infrared imaging sensors with biosensors has been highlighted as essential for transmitting tumor thermal information for remote access and monitoring.This review highlights the advancements in brain tumor detection methods, emphasizing the importance of feature extraction techniques, neural networks, and infrared imaging sensors in improving classification accuracy and segmentation processes[5].

Munir, K., Elahi, H., Ayub, A., Frezza, F., & Rizzi, A. (2019) cancer diagnosis utilizing deep learning emphasizes the significant advancements in computational techniques that enhance the accuracy and efficiency of cancer detection and classification. It discusses traditional diagnostic methods, such as the ABCD and seven-point detection methods, highlighting their limitations and the growing need for more effective solutions. The paper underscores the critical role of convolutional neural networks (CNNs) in improving diagnostic performance, while also addressing challenges such as data imbalance in training datasets, which can lead to biased results. To counteract these issues, the review advocates for equal representation of positive and negative cases in training data and suggests techniques like multicrop pooling to handle variations in target object sizes within images.

Furthermore, it compiles various deep learning architectures and their applications across different cancer types, including breast, lung, brain, and skin cancers, providing insights into successful studies and the importance of accessible datasets for research. Overall, this literature review serves as a comprehensive resource for researchers and practitioners, offering foundational knowledge and practical guidance for implementing deep learning in cancer diagnosis[6].

Sinnaswamy, R. A., Palanisamy, N., Subramaniam, K., Muthusamy, S., Lamba, R., & Sekaran,S. (2023) Aneurysm detection, particularly for cerebral, abdominal aortic, and thoracic aortic aneurysms, is critical due to the high fatality rates associated with aneurysm rupture. Recent advancements in artificial intelligence, especially machine learning and deep learning, have shown promise in early prediction and risk reduction. A comprehensive review of literature from 2007 to 2022 high- lights a wide range of methodologies and models for aneurysm prediction, focusing on image acquisition techniques, dataset sizes, and performance metrics. The review encompasses algorithms from non-linear kernel support vector regression to advanced 3D U-Net architectures, analyzing their effectiveness with CT scan images. Performance metrics such as sensitivity, specificity, and area under the receiver operating characteristic curve range from 0.7 to 1 for abdominal and intracranial aneurysms. However, thoracic aortic aneurysms have received less attention, suggesting a need for further research using both machine learning and deep learning models to enhance prediction capabilities for this aneurysm type[7].

Yang, H., Cho, K. C., Kim, J. J., Kim, J. H., Kim, Y. B., & Oh, J. H. (2023) Accurate prediction of rupture risk in cerebral aneurysms is cru- cial for preventing severe disability from aneurysm ruptures. Recent advances propose a novel deep learning approach incorporating hemodynamic parameters to enhance prediction accuracy. A new convolutional neural network (CNN) model was developed and retrospectively evaluated on 123 aneurysm cases, integrating hemodynamic factors such as wall shear stress (WSS) and strain. These parameters, initially computed via computational fluid dynamics and fluid–structure interaction, were converted into images for CNN training. Innovative data augmentation techniques generated a substantial dataset of 53,136 images. The CNN models demonstrated varying performance, with area under the receiver operating characteristic curve (AUC) values of 0.716 for WSS images, 0.741 for strain images, and 0.883 for combined images. While individual models for WSS or strain alone showed moderate predictive capabilities, the combined model achieved a sensitivity and specificity of 0.81 and 0.82, respectively, indicating strong potential foraccurate rupture risk assessment using deep learningmethods[8].

Zhang, Q., Bai, C., Liu, Z., Yang, L. T., Yu, H., Zhao, J., & Yuan, H. (2020) Recent advancements in high-performance computing, particularly through the use of GPUs, have significantly enhanced deep learning applications in medical image analysis. Despite the success of convolutional neural networks (CNNs) such as ResNet and VGG-16 in classifying medical images, challenges persist in cases involving multiple diseases within a single X-ray image. Addressing this, recent research introduces a modified ResNet model that replaces global average pooling with adaptive dropout, aiming to improve multi-label classification accuracy. By converting multi-label classification tasks into N binary classifications and employing novel data augmentation techniques, the proposed model effectively handles complex diagnostic scenarios. Evaluations conducted on GPU clusters with datasets including Montgomery County, Shenzhen, and NIH chest X-ray sets demonstrate notable improvementsin classification performance, achieving higher accuracy without significant reductions in efficiency compared to traditional architectures. This approach highlights the potential for advanced deep learning models to enhance diagnostic accuracy and efficiency in smart medicine[9].

Mascarenhas, S., & Agarwal, M. (2021) the field of artificial intelligence (AI) has seen remarkable progress over the past two decades, with machine learning (ML) emerging as a key subfield. ML allows computers to learn from data and improve their performance without explicit programming. A major evolution within ML is deep learning, which employs artificial neural networks to enable systems to learn from large datasets and make intuitive decisions. One of the most significant applications of deep learning is image classification, which automates the sorting of images into predefined categories. This technology has demonstrated a substantial impact on business productivity by reducing time and manpower for tasks such as product classification. Various Convolutional Neural Network (CNN) architectures, such as VGG16, VGG19, and ResNet50, have been developed for this purpose, offering different levels of accuracy. Studies comparing these architectures for image classification have shown that ResNet50 outperforms VGG16 and VGG19, achieving higher accuracy rates, making it a superior choice for tasks like product classificationin retail[10].

Maur´ıcio, J., Domingues, I., & Bernardino, J. (2023)examines the comparative performance of these two prominent architectures in the field of image processing. It highlights that while CNNs have been the traditional choice for image classification tasks, ViTs have emerged as a strong alternative, particularly in scenarios involving smaller datasets due to their self-attention mechanisms that enhance relational understanding between images. The review synthesizes findings from various studies, noting that ViTs can outperform CNNs in certain contexts, especially when dealing with noise and perturbations in images. However, it also points out limitations of CNNs, such as their shift-invariance property, which can hinder performance under specific conditions. The review emphasizes the importance of understanding the characteristics and transferability of these architectures, suggesting that hybrid models combining both approaches may yield improved accuracy[11].

Hastomo, W., Karno, A. S. B., Sestri, E., Terisia, V., Yusuf, D., Arman, S. A., & Arif, D. (2024) discusses various methodologies and algorithms previously employed for brain tumor classification, highlighting the evolution of techniques from traditional machine learning approaches to advanced deep learning models. It references several studies that utilized different neural network architectures, such as CNN-pretrained ResNet-50, Inception-v3, and VGG-16, as well as optimization techniques like particle swarm optimization and genetic algorithms. The review emphasizes the challenges faced in accurately detecting tumors due to their diverse shapes and locations, which complicate manual analysis.

The authors argue that the EfficientNet architecture, particularly versions B1 and B2, offers a promising solution by leveraging its efficient design to improve classification accuracy and reduce computational costs, thereby enhancing the overall effectiveness of medical image analysis in diagnosing brain tumors[12].

Willemink, M. J., Koszek, W. A., Hardell, C., Wu, J., Fleischmann, D., Harvey, H., ... & Lungren, M. P. (2020)medical imaging highlights several critical challenges and advancements. A significant hurdle is the availability of image data, which is essential for implementing AI in clinical settings. Researchers emphasize the importance of understanding data sources and potential biases that can affect the generalizability of AI algorithms New methodologies, such as federated learning and interactive reporting, are being explored to improve data accessibility, although issues related to data curation and computational demands remain substantial barriers Additionally, the literature discusses the necessity of proper de-identification of sensitive information in compliance with regulations like HIPAA and GDPR, as identifiable information can often be found in DICOM metadata and embedded within images. The process of anonymizing medical data is complex, requiring advanced techniques such as optical character recognition and human review to ensure patient identities are protected Furthermore, the literature addresses the challenges of labeling and defining ground truth in AI applications, noting that while expert labeling is feasible for small datasets, it becomes impractical for larger populations dueto time and cost constraints Overall, the literature underscores the need for innovative solutions to enhance data sharing and improve the robustness of AI applications in medical imaging[13].

Sarvamangala, D. R., & Kulkarni, R. V. (2022)Convolutional Neural Networks (CNNs) in medical image understanding reveals a significant focus on their application across various medical domains, including brain, breast, and lung imaging. Researchers have highlighted the effectiveness of CNNs in overcoming challenges traditionally faced by human experts, such as fatigue and oversight, which can hinder accurate diagnosis and prognosis .The review indicates that CNNs excel in feature extraction, learning low, mid, and high-level features from medical images, thus enhancing diagnostic accuracy .However, the literature also points out that the efficiency of feature extraction methods remains a concern, necessitating the development of intelligent healthcare systems capable of automating this process .Additionally, the diversity of CNN architectures complicates the selection of the most suitable model for specific tasks, and issues such as class imbalance and the need for extensive training data are prevalent challenges .The review underscores the importance of augmentation techniques and transfer learning to improve model performance, especially in scenarios with limited data .Overall, while CNNs have shown promising results, the literature calls for further exploration into efficient feature extraction methods and the integration of contextual information, such as patient history, to enhance medical image understanding[14].

Singh, A., Sengupta, S., & Lakshminarayanan, V. (2020) explainable deep learning models in medical image analysis highlights the significant advancements and challenges in the field. Deep learning has shown remarkable effectiveness in various medical diagnostic tasks, often surpassing human experts; however, its black-box nature limits clinical adoption due to a lack of transparency and interpretability. Recent studies focus on elucidating the features that most influence model decisions, emphasizing the need for explainability to foster trust among healthcare professionals and patients. The literature also discusses various approaches to explainability, including attribution methods and perturbation- based techniques, which aim to clarify how models arrive at their conclusions. Furthermore, the integration of expert knowledge into model design has been explored, demonstrating that domain-specific insights can enhance both the performance and interpretability of deep learning systems . Despite these advancements, the review identifies ongoing challenges, such as the need for more robust methods that can provide both positive and negative evidence in explanations, and the necessity for further research to address the ethical implications of deploying AI in healthcare. Overall, the literature underscores the critical role of explainability in the successful implementation of AI technologiesin medical imaging, advocating for continued exploration of methods that bridge the gap between complex algorithms and user understanding[15].

Ganaie, M. A., Hu, M., Malik, A. K., Tanveer, M., & Suganthan, P. N. (2022) Ensemble learning combines multiple models to improve generalization and performance, a concept that has been further enhanced by integrating deep learning architectures. Deep ensemble learning leverages both the powerful feature extraction capabilities of deep neural networks and the diversity of ensemble methods to create more robust models. This approach has been successfully applied across a range of tasks, such as classification, regression, and clustering, with applications spanning domains like healthcare, speech processing, and image classification. Various ensemble strategies, including bagging, boosting,and stacking, as well as novel methods like negativecorrelation and decision fusion, have been developed to improve model performance. Despite the computational challenges, deep ensemble models continue to show promise in delivering higher accuracy and generalization, outperforming traditional models in many cases. However, further research is needed to address challenges such as training time, model complexity, and the optimal fusion of complementary algorithms[16].

Yang, S., Xiao, W., Zhang, M., Guo, S., Zhao, J., & Shen, F. (2022) image data augmentation highlights its critical role in enhancing the performance of deep learning models, particularly in computer vision tasks. Various studies have explored different augmentation techniques, emphasizing the need for diverse and sufficient training data to prevent overfitting in deep neural networks. For instance, while some works focus on traditional transformations and generative adversarial networks (GANs) for image classification, others have reviewed augmentation methods specifically for face recognition tasks or based on data warping and oversampling. However, many of these studies lack a systematic review across multiple computer vision tasks, such as object detection, semantic segmentation, and image classification . The existing research also reveals challenges, including the inefficiency of task-independent augmentation methods and the absence of theoretical frameworks to determine the optimal size of training datasets . Recent advancements have introduced novel techniques like Mixup,which creates new images by averaging pairs from the training set, and methods that utilize saliency maps to enhance the quality of augmented data. Additionally, approaches like Cutout and GridMask have been proposed to improve model robustness by randomly masking regions of input images during training. Overall, the literature indicates a growing recognition of the importance of data augmentation in deep learning, while also identifying gaps that necessitate further exploration and systematic evaluation of various methods[17].

Liu, Y., Yang, G., Qiao, S., Liu, M., Qu, L., Han, N., ... & Peng, Y. (2023) deep learning models have made significant advancements in computer vision by relying on large-scale, class-balanced datasets, yet they often struggle with class-imbalanced data, a common issue in real-world applications. Imbalanced datasets, characterized by the dominance of majority classes, lead to degraded model performance, particularly in predicting minority classes. While traditional classification methods focus on class-balanced data, few studies address the challenges posed by class-imbalance, especially in real- time data scenarios generated from mobile devices. To tackle this, researchers have explored techniques such as active sampling, data augmentation, and transfer learning. DenseNet-based models, enhanced with transfer learning, have been proposed to improve performance on imbalanced datasets. Prior studies have examined approaches like bagging, boosting, and other ensemble methods to mitigate this issue, but research continues to evolve, as methods that dynamically adjust sample sizes, improve memory efficiency, and leverage real-time data show promising results in handling class-imbalance and enhancing classification performance[18].

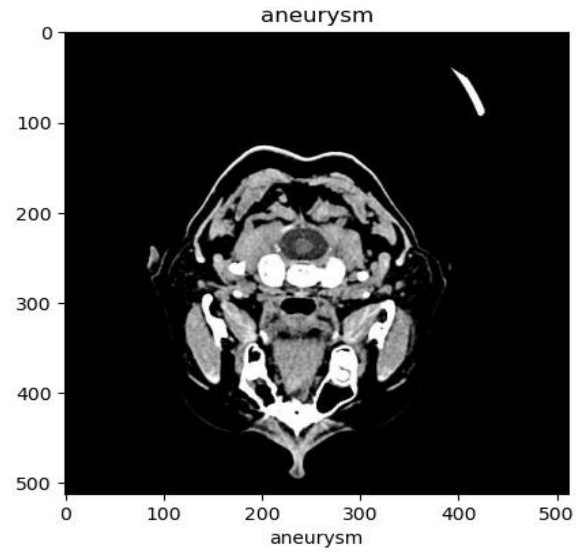
Algan, G., & Ulusoy, I. (2021) label noise in deep neural networks has become a critical area of study due to its prevalence in real-world datasets, where obtaining correctly labeled data is often challenging, expensive, or even prone to errors. Various approaches have been developed to tackle this issue,ranging from noise model-based methods, which estimate and leverage the noise distribution, to noise model-free methods that aim to make models inherently robust against noise by using techniques like robust loss functions, regularization, and alternative learning paradigms. Although deep neural networks possess remarkable representation learning abilities, their capacity to overfit noisy labels is well-documented, making the development of noise- robust algorithms vital for applications like automated medical diagnosis. Despite the extensive research on handling label noise in traditional machine learning, the literature on deep learning-specific solutions remains less comprehensive. Therefore, this paper aims to bridge that gap by categorizing and summarizing key deep learning methodologies for managing noisy labels[19].

Maharana, K., Mondal, S., & Nemade, B. (2022) data pre-processing in machine learning highlights its critical role in enhancing data quality and model performance. It emphasizes that raw data often contains noise, inconsistencies, and missing values, necessitating thorough cleaning and transformation processes to prepare it for analysis. Various methodologies, such as data integration, data augmentation, and cleaning techniques, are discussed to ensure that the data is coherent and relevant for mining. The review also points out that data pre-processing can consume a significant portion of the overall classification process, with estimates suggesting it can take up to 80 percent of the time. Furthermore, it addresses the importance of understanding the features of the data and adhering to specific business rulesto maintain data integrity. The review concludes that effective data pre-processing is essential for improving the accuracy of machine learning models and reducing the risk of false predictions[20].

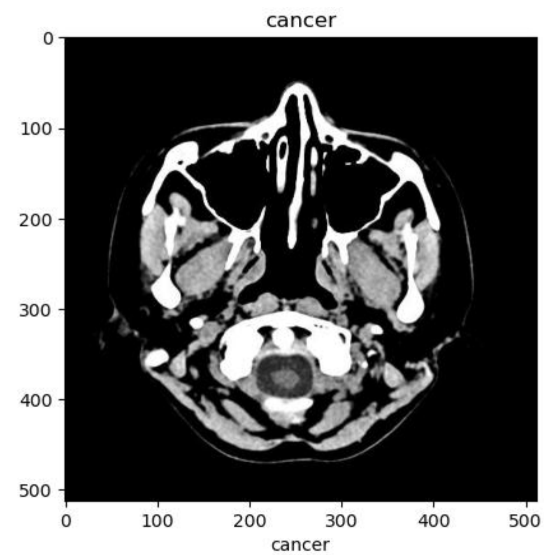
IV. METHDOLOGY

A. Data collection

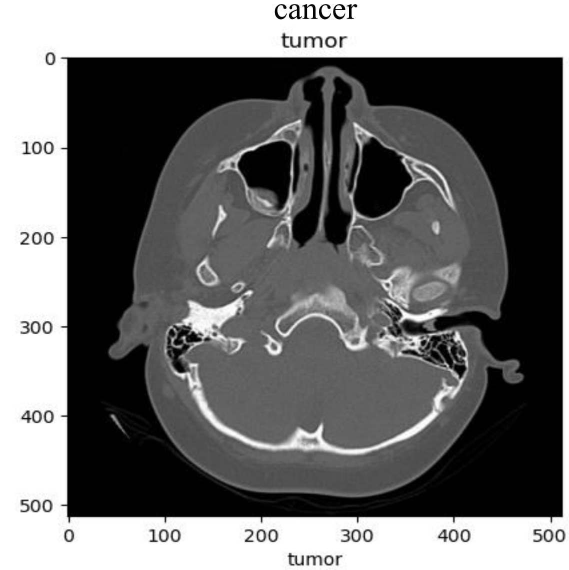
The dataset includes labeled CT images representing three brain pathologies: aneurysms, cancer, and tumors. These labels provide the necessary ground truth for training machine learning models to detect each pathology. The diversity of the images across classes enhances the model's ability to generalize across varied cases.



Fig(1)
aneurysm

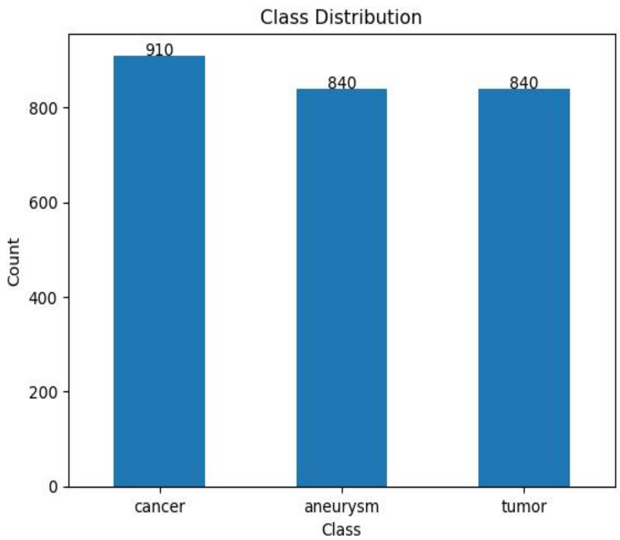


Fig(2)
cancer



Fig(3)
tumor

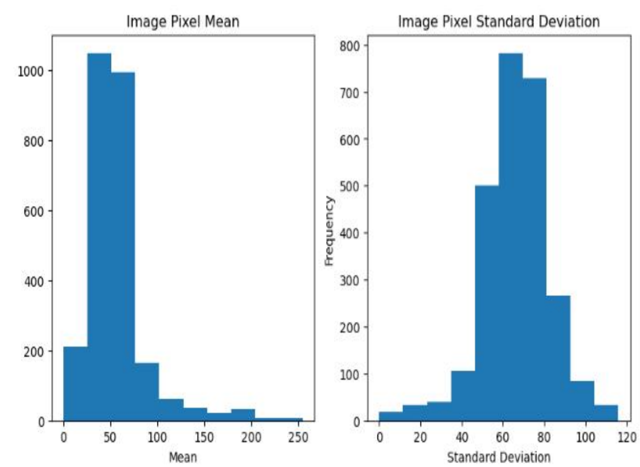
B. EDA



Fig(4)

Class distribution

Inference: Fig(4) represents the Class Distribution Chart shows a relatively balanced distribution across three medical conditions, with cancer cases (910) slightly higher than aneurysm and tumor cases (840 each). This near-balanced distribution suggests the dataset is well-suited for training machine learning models without significant class imbalance concerns.



Fig(5)

Image pixel mean and standard deviation

Inference: Fig(5) represents Pixel Mean: Most images have relatively low mean pixel values (concentrated between 0-50), suggesting darker images overall which is typical for medical CT scans. Pixel Standard Deviation: The standard deviation peaks between 60-80, indicating consistent contrast levels across most images, which is important for reliable medical image analysis.

C. Train-Test Split

To evaluate model performance, the dataset is split into training, validation, and test sets. This division allows for training the model on one portion of the data, validating it on another to adjust hyperparameters, and finally testing its generalization on unseen data.

Training Set: Contains 1,554 entries for training the model. This set is used to optimize the model parameters.

Validation Set: Contains 129 entries to fine-tune model hyperparameters and help prevent overfitting.

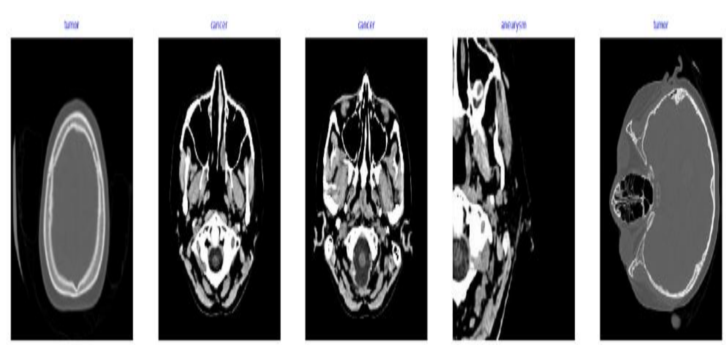
Test Set: Contains 907 entries for final evaluation, measuring model generalization on unseen data.

D. Data Preprocessing

1. Image Resizing All images are resized to a fixed dimension of 224x224 pixels. This resizing ensures uniformity in input size, making it compatible with the model architecture and optimizing memory usage.

2. Image Normalization Each image's pixel values are scaled by dividing by 255, bringing pixel values into a range between 0 and 1. This normalization is crucial for stable and efficient model training.

3. Data Augmentation Random Horizontal Flipping:During training, images are randomly flipped horizontally, creating slight variations in orientation. This flip augmentation is implemented using the `ImageDataGenerator` with the parameter `horizontal_flip=True`. Horizontal flipping enhances the model's generalization ability by exposing it to a wider variety of image perspectives.



Fig(6)

Sample images after preprocessing

E. Model Selection

VGG-16

Model Architecture:

Model: "sequential"

Layer (type)	Output Shape	Param #
vgg16 (Functional)	?	14,714,688
dense (Dense)	?	0 (unbuilt)
batch_normalization (BatchNormalization)	?	0 (unbuilt)
dropout (Dropout)	?	0
dense_1 (Dense)	?	0 (unbuilt)
batch_normalization_1 (BatchNormalization)	?	0 (unbuilt)
dropout_1 (Dropout)	?	0
dense_2 (Dense)	?	0 (unbuilt)
batch_normalization_2 (BatchNormalization)	?	0 (unbuilt)
dropout_2 (Dropout)	?	0
dense_3 (Dense)	?	0 (unbuilt)

Total params: 14,714,688 (56.13 MB)

Trainable params: 14,714,688 (56.13 MB)

Non-trainable params: 0 (0.00 B)

Fig(7)

VGG-16 Model Architecture

Fig(7) represents:

1. The model starts with a pre-trained VGG16 base (14,714,688 parameters)

2. Followed by 3 dense layer blocks, each consisting of a Dense layer, BatchNormalization, and Dropout(0.45)

3. Final Dense layer with 3 outputs using softmax activation for classification 4. Total model size is 56.13 MB with all parameters being trainable

ResNet-50

Model Architecture:

Model: "sequential"

Layer (type)	Output Shape	Param #
resnet50 (Functional)	?	23,587,712
dense (Dense)	?	0 (unbuilt)
batch_normalization (BatchNormalization)	?	0 (unbuilt)
dropout (Dropout)	?	0
dense_1 (Dense)	?	0 (unbuilt)
batch_normalization_1 (BatchNormalization)	?	0 (unbuilt)
dropout_1 (Dropout)	?	0
dense_2 (Dense)	?	0 (unbuilt)
batch_normalization_2 (BatchNormalization)	?	0 (unbuilt)
dropout_2 (Dropout)	?	0
dense_3 (Dense)	?	0 (unbuilt)

Total params: 23,587,712 (89.98 MB)

Trainable params: 23,534,592 (89.78 MB)

Non-trainable params: 53,120 (207.50 KB)

Fig(8)

ResNet-50 Model Architecture

Fig(8) represents:

1. Uses pre-trained ResNet50 as the base model (23,587,712 parameters)

2. It has 3 dense layer blocks with BatchNormalization and Dropout

3. Total model size is 89.98 MB, with 23,534,592 trainable parameters and 53,120 non-trainable parameters 4. The non-trainable parameters (207.50 KB) indicate some layers are frozen, which is common in transfer learning

EfficientNet-B7

Model Architecture:

Model: "sequential"

Layer (type)	Output Shape	Param #
efficientnetb7 (Functional)	?	64,097,687
dense (Dense)	?	0 (unbuilt)
batch_normalization (BatchNormalization)	?	0 (unbuilt)
dropout (Dropout)	?	0
dense_1 (Dense)	?	0 (unbuilt)
batch_normalization_1 (BatchNormalization)	?	0 (unbuilt)
dropout_1 (Dropout)	?	0
dense_2 (Dense)	?	0 (unbuilt)
batch_normalization_2 (BatchNormalization)	?	0 (unbuilt)
dropout_2 (Dropout)	?	0
dense_3 (Dense)	?	0 (unbuilt)

Total params: 64,097,687 (244.51 MB)

Trainable params: 63,786,960 (243.33 MB)

Non-trainable params: 310,727 (1.19 MB)

Fig(9)

EfficientNet-B7 Model Architecture

Fig(9) represents:

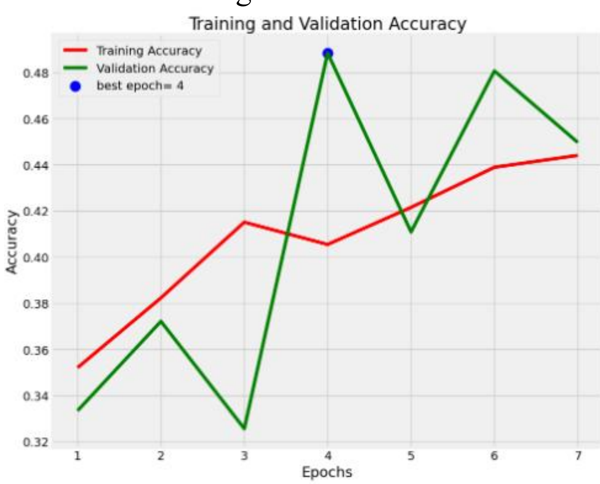
1. Uses pre-trained EfficientNetB7 as the base model (64,097,687 total parameters) Features the standard 3 dense layer blocks with BatchNormalization and Dropout pattern

2. Much larger model at 244.51 MB total size, with 63,786,960 trainable parameters and 310,727 non-trainable parameters (1.19 MB)

F. Evaluation Metrices

VGG-16

Fig(10)
Training and validation Loss



Fig(11)
Training and validation Loss

Fig(10) and Fig(11) represents the VGG16 model demonstrates unstable performance across 7 epochs. The validation loss (blue line) drops sharplyinitially from 8.0 to approximately 5.5 but fluctuates thereafter, while the training loss (red line) remains relatively flat around 5.5.

The accuracy plot shows inconsistent behavior, with significant fluctuations in validation accuracy (ranging from 32% to 48%), whereas training accuracy shows only modest improvement to about 44%, indicating potential challenges with model training and possible underfitting.

Confusion Matrix, Without Normalization

[[14 112 173]

[0 139 174]

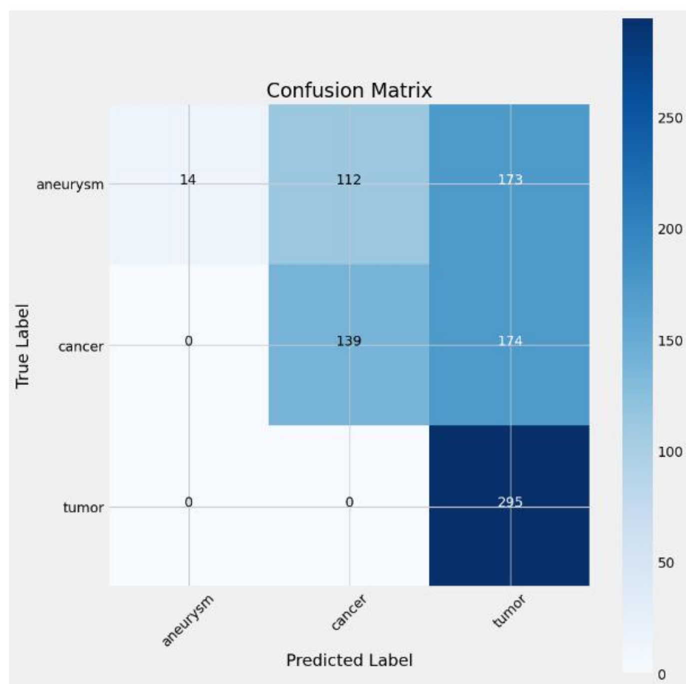
[0 0 295]]

	precision	recall	f1-score	support
aneurysm	1.00	0.05	0.09	299
cancer	0.55	0.44	0.49	313
tumor	0.46	1.00	0.63	295
accuracy			0.49	907
macro avg	0.67	0.50	0.40	907
weighted avg	0.67	0.49	0.40	907

TABLE I
CLASSIFICATION REPORT METRICS

The table I represents model shows significantly imbalanced performance across classes, with poor recall for aneurysm (0.05) despite perfect precision (1.00), and moderate performance for cancer and tumor detection. The overall accuracy of 0.49 on 907 test samples indicates suboptimal performance, with macro and weighted averages around 0.40-0.67 suggesting inconsistent classification abilities across classes. This indicates that VGG16 struggles with reliable disease classification in this medical imaging task compared to other architectures.

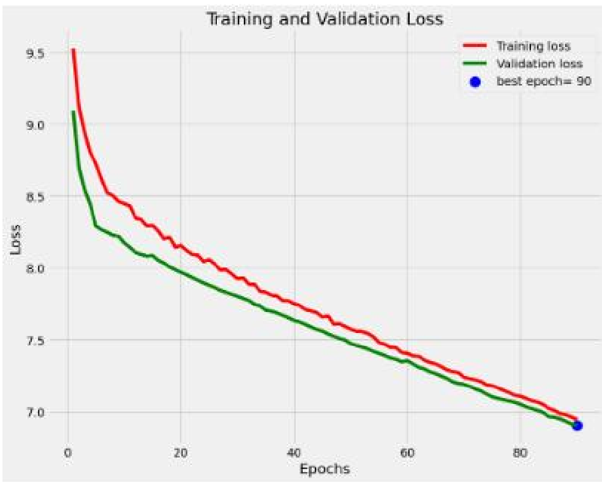
Confusion matrix, which shows many misclassifications, especially with aneurysm cases being incorrectlyclassified as cancer or tumor.



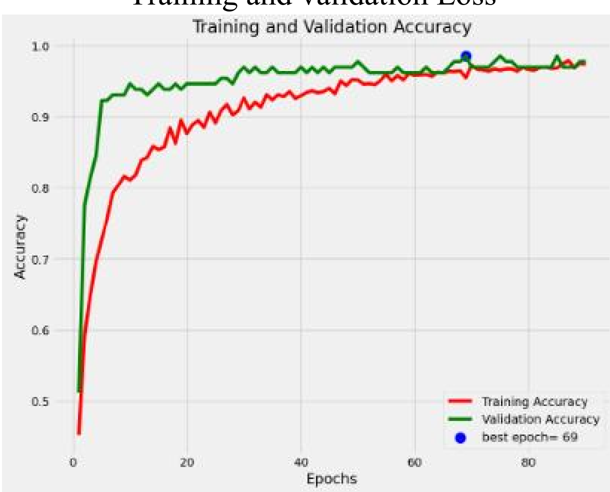
Fig(12)
Confusion matrix for VGG-16

Fig(12) represents the confusion matrix shows how a VGG16 model performs in classifying medical conditions. The model correctly identified 14 aneurysms, 139 cancer cases, and 295 tumors. However, it shows significant misclassification of aneurysms (112 as cancer and 173 as tumors) and some cancer cases (174 misclassified as tumors), indicating the model needs improvement in distinguishing between these conditions, particularly for aneurysm detection.

ResNet50



Fig(13)
Training and validation Loss



Fig(14)
Training and validation accuracy

Fig(13) and Fig(14) represents the ResNet50 model shows strong performance, with both training and validation metrics improving over 80 epochs. The loss curves (left) steadily decrease from around 9.5 to 7.0, indicating good model convergence.

The accuracy curves (right) demonstrate rapid improvement in the early epochs, reaching approximately 90% accuracy by epoch 20, and continuing to improve to nearly 95% by epoch 80. The minimal gap between training and validation accuracy suggests good generalization without overfitting.

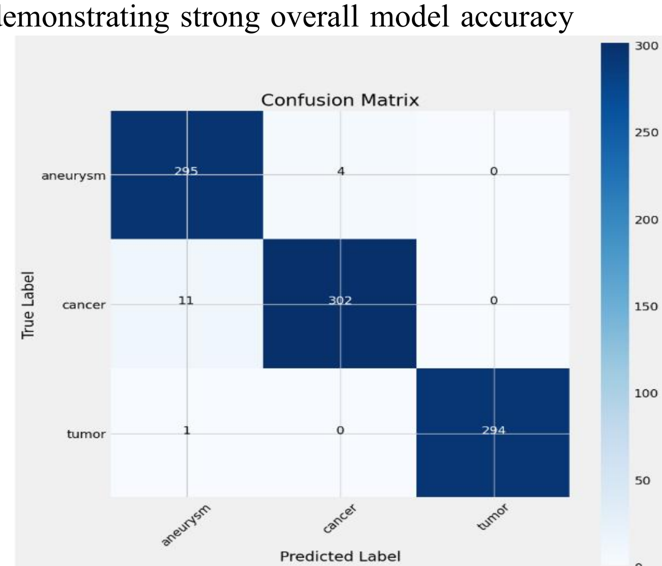
Confusion Matrix, Without Normalization
[[295 4 0]
[11 302 0]
[1 0 294]]

	precision	recall	f1-score	support
aneurysm	0.96	0.99	0.97	299
cancer	0.99	0.96	0.98	313
tumor	1.00	1.00	1.00	295
accuracy			0.98	907
macro avg	0.98	0.98	0.98	907
weighted avg	0.98	0.98	0.98	907

TABLE II
CLASSIFICATION METRICS

The table II represents model exhibits outstanding performance with high precision (0.96-1.00) and recall (0.96-1.00) across all classes, particularly achieving perfect scores for tumor detection (precision=1.00, recall=1.00, f1-score=1.00). The overall accuracy of 0.98 across 907 test samples demonstrates robust performance, and the consistent macro and weighted averages (both 0.98) indicate balanced performance across all classes without class imbalance issues.

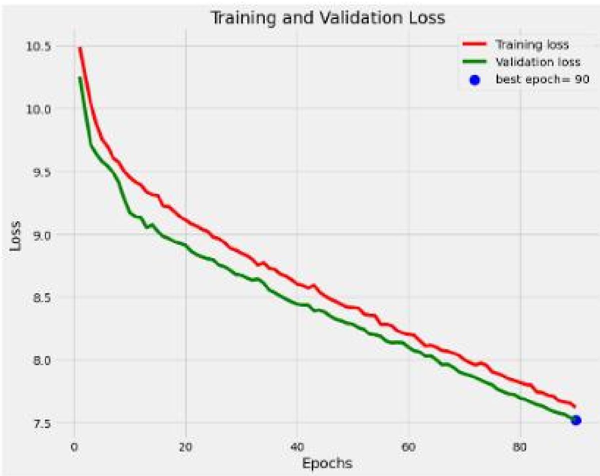
Confusion matrix for ResNet50 shows excellent classification performance across three classes. Out of 299 instances in the first class, 295 were correctly classified with only 4 misclassifications. The second class shows 302 correct predictions with 11 misclassifications, and the third class achieved 294 correct predictions with just 1 misclassification, demonstrating strong overall model accuracy



Fig(15) Confusion matrix for ResNet-5

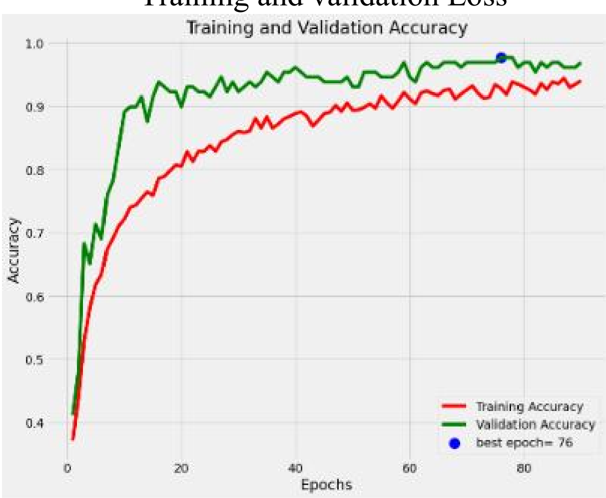
Fig(15) represents the ResNet-50 model demonstrates excellent performance in classifying medical conditions, with high accuracy across all three classes. It correctly identified 295 aneurysm cases, 302 cancer cases, and 294 tumor cases. The model shows minimal misclassification with only 4 aneurysms misclassified as cancer, 11 cancer cases misclassified as aneurysms, and just 1 tumor case misclassified as aneurysm, indicating strong reliability in medical image classification.

EfficientNet-B7



Fig(16)

Training and validation Loss



Fig(17)

Training and validation accuracy

Fig(16) and Fig(17) represents theEfficientNetB7 model shows strong learning progression over 90 epochs, with both training and validation loss steadily decreasing from 10.5 to around 7.5. The accuracy plot demonstrates impressive performance, with validation accuracy (green line) reaching above 90% early and consistentlyoutperforming training accuracy (red line), reaching approximately 95% by the end, suggesting excellent generalization without overfitting.

Confusion Matrix, Without Normalization

[[293 6 0]

[13 299 1]

[1 1 293]]

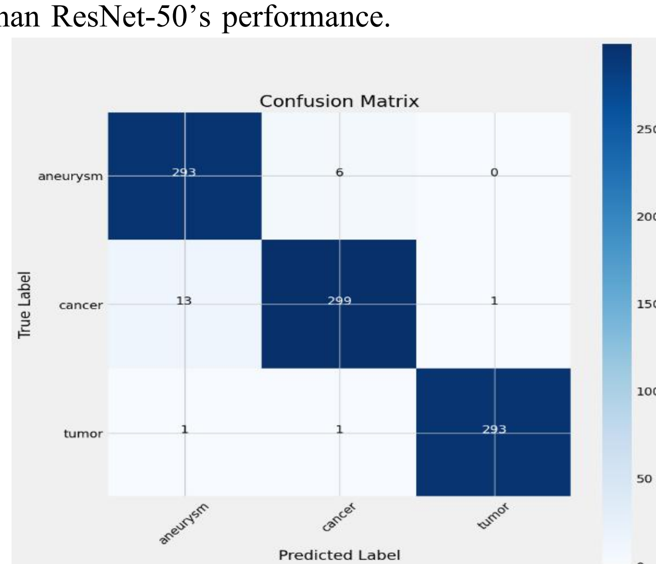
	precision	recall	f1-score	support
aneurysm	0.95	0.98	0.97	299
cancer	0.98	0.96	0.97	313
tumor	1.00	0.99	0.99	295
accuracy			0.98	907
macro avg	0.98	0.98	0.98	907
weighted avg	0.98	0.98	0.98	907

TABLE III

Classification Metrics

The table III represents model demonstrates excellent performance across all three classes (aneurysm, cancer, and tumor) with high precision (0.95-1.00) and recall (0.96-0.99) values. Notably, it achieves perfect precision (1.00) for tumor detection and shows consistent performance with an overall accuracy of 0.98 across 907 test samples. The balanced scores across macro and weighted averages (both 0.98) indicate that the model performs uniformly well across all classes without significant bias.

Confusion matrix for EfficientNet-B7 shows strong classification performance across three classes. The model correctly classified 293 aneurysm cases (with 6 misclassifications), 299 cancer cases (with 14 misclassifications), and 293 tumor cases (with only 2 misclassifications), demonstrating high accuracy though slightly lower than ResNet-50's performance.



Fig(18)

Confusion matrix for EfficientNet-B7

Fig(18) represents the EfficientNet-B7 model shows strong classification performance in medical image analysis, correctly identifying 293 aneurysm cases, 299 cancer cases, and 293 tumor cases. The model shows minimal misclassification with 6 aneurysms misclassified as cancer, 13 cancercases misclassified as aneurysms, 1 cancer case misclassified as tumor, and 2 tumor cases misclassified as aneurysm and cancer respectively, demonstratingreliable diagnostic capabilities.

V. RESULTS

VGG-16

Train Loss: 5.221975326538086 Train Accuracy: 0.4761904776096344

Validation Loss: 5.168546676635742 ValidationAccuracy: 0.4883720874786377

Test Loss: 5.197704792022705 Test Accuracy: 0.49393606185913086

Resnet5

Train Loss: 6.8499298095703125 Train Accuracy: 0.993565022945404

Validation Loss: 6.904529571533203 Validation Accuracy: 0.9767441749572754

Test Loss: 6.8907952308654785 Test Accuracy: 0.9823594093322754

EfficientNet-B7

Train Loss: 7.4707350730896 Train Accuracy: 0.9929214715957642

Validation Loss: 7.52120304107666 Validation Accuracy: 0.9689922332763672

Test Loss: 7.505218505859375 Test Accuracy: 0.9757441878318787

Comparing the three models, VGG16 shows the poorest performance with a test accuracy of only 49.39%, indicating significant room for improvement. ResNet50 demonstrates the best performance with exceptional test accuracy of 98.23%, slightly outperforming EfficientNet-B7 which achieved 97.57% test accuracy. While both ResNet50 and EfficientNet-B7 show strong training accuracy (99.35% and 99.29% respectively), ResNet50's higher validation and test accuracies suggest it has better generalization capability for this medical image classification task.

VI. FUTURE WORKS

1. Building a hybrid model by incorporating parts of ResNet or EfficientNet into VGG16 is an advanced strategy. You might try this approach if simple architectural modifications don't yield sufficientimprovements.

2.Deploying the Model in Real-World Settings: Integrating the model into clinical software for real-time analysis and assisting healthcare professionals in diagnostics.

3.Implementing Explainable AI (XAI) Techniques: Applying techniques such as Grad-CAM or LIME to provide visual explanations for model predictions. This would help clinicians understand which areas of the CT scan contributed to the model's decision, increasing trust and interpretability in clinical settings.

REFERENCES

[1] Algan, G., & Ulusoy, I. (2021). Image classification with deep learning in the presence of noisy labels: A survey. *Knowledge-Based Systems*, 215, 106771.

[2] Amin, J., Sharif, M., Haldorai, A., Yasmin, M., & Nayak,R. S. (2022). Brain tumor detection and classification us- ing machine learning: a comprehensive survey. *Complex & Intelligent Systems*, 8(4), 3161-3183.

[3] Ganaie, M. A., Hu, M., Malik, A. K., Tanveer, M., & Suganthan, P. N. (2022). Ensemble deep learning: A re- view. *Engineering Applications of Artificial Intelligence*, 115, 105151.

[4] Hastomo, W., Karno, A. S. B., Sestri, E., Terisia, V., Yusuf, D., Arman, S. A., & Arif, D. (2024). Classification of Brain Image Tumor using EfficientNet B1-B2 Deep Learning. *Semesta Teknika*, 27(1), 46-54.

[5] Kadry, S., Nam, Y., Rauf, H. T., Rajinikanth, V., & Lawal, I. A. (2021, March). Automated detection of brain abnormality using deep-learning-scheme: a study. In *2021 Seventh International Conference on Bio Signals, Images, and Instrumentation (ICBSII)* (pp. 1-5). IEEE.

[6] Khumancha, M. B., Barai, A., & Rao, C. R. (2019, July). Lung cancer detection from computed tomography (CT) scans using convolutional neural network. In *2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT)* (pp. 1-7). IEEE.

[7] Liu, Y., Yang, G., Qiao, S., Liu, M., Qu, L., Han, N., ... & Peng, Y. (2023). Imbalanced data classification: Using transfer learning and active sampling. *Engineering Applications of Artificial Intelligence*, 117, 105621.

[8] Maharana, K., Mondal, S., & Nemade, B. (2022). A review: Data pre-processing and data augmentation techniques. *Global Transitions Proceedings*, 3(1), 91-99.

[9] Mascarenhas, S., & Agarwal, M. (2021, November). A comparison between VGG16, VGG19 and ResNet50 architecture frameworks for image classification. In *2021 International Conference on Disruptive Technologies for Multi-Disciplinary Research and Applications (CENT-CON)* (Vol. 1, pp. 96-99). IEEE.

[10] Maur´ıcio, J., Domingues, I., & Bernardino, J. (2023). Comparing vision transformers and convolutional neural networks for image classification: A literature review. *Applied Sciences*, 13(9), 5521.

[11] Munir, K., Elahi, H., Ayub, A., Frezza, F., & Rizzi, A. (2019). Cancer diagnosis using deep learning: a bibliographic review. *Cancers*, 11(9), 1235.

[12] Puttagunta, M., & Ravi, S. (2021). Medical image analysis based on deep learning approach. *Multimedia Tools and Applications*, 80(16), 24365-24398.

[13] Sarvamangala, D. R., & Kulkarni, R. V. (2022). Convolu- tional neural networks in medical image understanding: a survey. *Evolutionary Intelligence*, 15(1), 1-22.

[14] Shakeel, P. M., Tobely, T. E. E., Al-Feel, H., Manogaran, G., & Baskar, S. (2019). Neural network based brain tumor detection using wireless infrared imaging sensor. *IEEE Access*, 7, 5577-5588.

[15] Singh, A., Sengupta, S., & Lakshminarayanan, V. (2020). Explainable deep learning models in medical image analysis. *Journal of Imaging*, 6(6), 52.

[16] Sinnaswamy, R. A., Palanisamy, N., Subramaniam, K.,Muthusamy, S., Lamba, R., & Sekaran, S. (2023). An extensive review on deep learning and machine learning intervention in prediction and classification of types of aneurysms. Wireless Personal Communications, 131(3), 2055-2080.

[17] Willemink, M. J., Koszek, W. A., Hardell, C., Wu, J., Fleischmann, D., Harvey, H., ... & Lungren, M. P. (2020). Preparing medical imaging data for machine learning. Radiology, 295(1), 4-15.

[18] Yang, H., Cho, K. C., Kim, J. J., Kim, J. H., Kim, Y. B., & Oh, J. H. (2023). Rupture risk prediction of cerebral aneurysms using a novel convolutional neural network-based deep learning model. Journal of NeuroInterventional Surgery, 15(2), 200-204.

[19] Yang, S., Xiao, W., Zhang, M., Guo, S., Zhao, J., & Shen,F. (2022). Image data augmentation for deep learning: A survey. arXiv preprint arXiv:2204.08610.

[20] Zhang, Q., Bai, C., Liu, Z., Yang, L. T., Yu, H., Zhao, J., & Yuan, H. (2020). A GPU-based residual network for medical image classification in smart medicine. Information Sciences, 536, 91-100.