FEATURE FUSION-BASED DETECTION OF SCHIZOPHRENIA USING fMRI DATASET

A Project Report

Submitted by

MOHAN V (CB.EN.U4CCE20033)
NITHIN RAJASEHARAN (CB.EN.U4CCE20039)
TAUSHIQ BALAMURUGAN (CB.EN.U4CCE20064)
VARUN S (CB.EN.U4CCE20069)

in partial fulfillment of the requirements for the award of

the degree of

Bachelor of Technology

in

Computer and Communication Engineering

under the supervision of

Ms. Suguna G

Department of Electronics and Communication Engineering,
Amrita School of Engineering,
Amrita Vishwa Vidyapeetham,
Coimbatore – 641112

June 2024

Dept of Elec. & Communication Engineering,
Amrita School of Engineering,
Amrita Vishwa Vidyapeetham,
Coimbatore – 641112

Certificate

This is to certify that this project titled **"Feature fusion-based detection of Schizophrenia using fMRI dataset"** submitted in partial fulfillment of the requirements for the award of the degree of **Bachelor of Technology in Computer and Communication Engineering**, by **Mohan V, Nithin Rajaseharan, Taushiq Balamurugan** and **Varun S**, is a bonafide record of work carried out by them, under my supervision and that it has not been submitted, to the best of my knowledge, in part or in full, for the award of any other degree or diploma.

Ms. Suguna G
(Advisor)
Department of Electronics and
Communication Engineering,
Amrita School of Engg.,
Coimbatore

Dr. Shanmugha Sundaram G A
(**Batch Coordinator**)
BTech 2020-2024

Dr. Madhu Mohan N
(Dept. Chair)

Date:

This project was evaluated by us on ----------------------------------.

Internal Examiner

External Examiner

Declaration

We do hereby declare that this project titled **"Feature fusion-based detection of Schizophrenia using fMRI dataset",** submitted in partial fulfillment of the requirements for the award of the degree of **Bachelor of Technology in Computer & Communication Engineering**, is a true record of work carried out by us, under the supervision of **Ms. Suguna G** and that all information contained herein, which do not arise directly from our work, have been properly acknowledged and cited, using acceptable international standards. Further, we declare that the contents of this report have not been submitted, in part or in full, for the award of any other degree or diploma.

Coimbatore – 641112

Date:

MOHAN V

NITHIN RAJASEHARAN

TAUSHIQ BALAMURUGAN

VARUN S

Dedicated to our teachers, friends and parents

Acknowledgements

We want to express our deepest gratitude to Amrita Vishwa Vidyapeetham for giving us the chance to work on this final-year project. This journey has been very enriching, and we owe our thanks to many people.

We would like to thank our mentor, Ms. Suguna G for her constant guidance, encouragement, and knowledge were vital in shaping this project. Her beneficial comments and support kept us motivated and focused.

We also thank the Head of the Department, Dr. Madhu Mohan N, for ensuring about the legalities behind our biomedical image processing project. His management and support have been important for the successful completion of our project.

We honestly admire all of the faculty members who supported us for the duration of this journey. Their expertise and assistance have been beneficial. Special thanks to Dr. Harikumar M E of the ECE department for validating our results and helping towards successful completion of the project.

We are very thankful to our panel members, Dr. Harish Ram D S, Dr. Natarajamani S, for their valuable time and helpful feedback. Their insights significantly improved our work.

Thank you to Amrita Vishwa Vidyapeetham for developing an instructional environment that supported our learning and improvement. This project has been a significant milestone in our academic journey, and we are thankful for the opportunity to have undertaken it.

Abstract

Schizophrenia is a severe mental illness that disrupts patients' perception of reality. Diagnosing it currently relies heavily on psychiatrists' subjective evaluation of symptoms and social functioning. This method is time-consuming, prone to delays due to manual MRI analysis, and lacks objectivity. The absence of a computer-aided diagnosis (CAD) system can lead to misdiagnoses.

A novel approach for schizophrenia detection through CAD system is using feature fusion with machine learning (ML) and deep learning (DL). The system leverages data from functional Magnetic Resonance Imaging (fMRI) scans. ML techniques excel at identifying patterns in data, while DL excels at uncovering complex relationships within it. By combining these two methods, the system can overcome limitations of traditional methods and enhance detection accuracy. Metrics like accuracy, sensitivity, and specificity will be used to ensure the system's reliability and efficiency.

This integration of ML and DL has significantly improved diagnostic accuracy to 95%. This, in turn, can facilitate earlier diagnoses for patients struggling with schizophrenia. Automating part of the evaluation process can alleviate the burden on psychiatrists and expedite diagnosis, leading to quicker treatment and improved outcomes.

Table of Contents

List of Abbreviations

CAD	Computer-Aided Diagnosis
ML	Machine Learning
DL	Deep Learning
MRI	Magnetic Resonance Imaging
DF	Deep Feature
HF	Handcrafted Feature
VGG	Visual Geometry Group
RFE	Recursive Feature Elimination
PCA	Principal Component Analysis
MLVSF	Multi-Layer Visual feature Fusion
ITK-SNAP	Insight Segmentation and Registration Toolkit for SNAP
PNG	Portable Network Graphic
CNN	Convolution Neural Network
Acc	Accuracy
Sens	Sensitivity
Spe	Specificity
GLCM	Gray Level Co-occurrence Matrix
LBP	Local Binary Pattern
HOG	Histogram of Oriented Gradients
AUC-ROC	Area Under the Receiver Operating Characteristic Curve
ViT	Vision Transformer

List of Tables

List of Figures

1. Introduction

Schizophrenia, a profound and debilitating mental disorder affecting approximately 20 million individuals globally (World Health Organization, 2020), encompasses the symptoms that distort the individual's perception of reality. Characterized by hallucinations, delusions, disorganized thinking, and emotional disturbances, schizophrenia disrupts the delicate tapestry of human experience, posing significant challenges to daily life.

Neuroimaging studies reveal structural and functional anomalies within the brain, including reduced gray matter volume, enlarged ventricles, and white matter abnormalities. These alterations suggest a possible dysregulation of neural networks, impeding the seamless flow of information and contributing to the characteristic symptomatology.

1.1 Motivation

The diagnostic method for schizophrenia remains reliant on the subjective evaluation of clinical symptoms and social functioning by psychiatrists, as the elusive quest for definitive organic indicators persists. This traditional approach, while fundamental, is marred by its time-consuming nature, compounded by the manual scrutiny of MRI scans, which introduces significant delays in the diagnostic process. The absence of a dedicated Computer-Aided Diagnosis (CAD) system further underscores the critical need for a more efficient diagnostic tool.

1.2 Anatomy of Brain

Schizophrenia is characterized by notable brain abnormalities, including reduced gray matter volume, enlarged ventricles, and white matter abnormalities. Neuroimaging studies consistently reveal these structural alterations, indicating a pervasive impact on the brain's anatomy. Reduced gray matter volume suggests neuronal loss or dysfunction, while enlarged ventricles signify abnormal cerebrospinal fluid distribution. White matter abnormalities point to disruptions in neural connectivity. These distinct characteristics underline the neurobiological complexity of schizophrenia, offering valuable insights into its pathophysiology. Understanding these features is crucial for advancing diagnostic precision and developing targeted interventions to address the intricate challenges posed by this severe mental disorder.

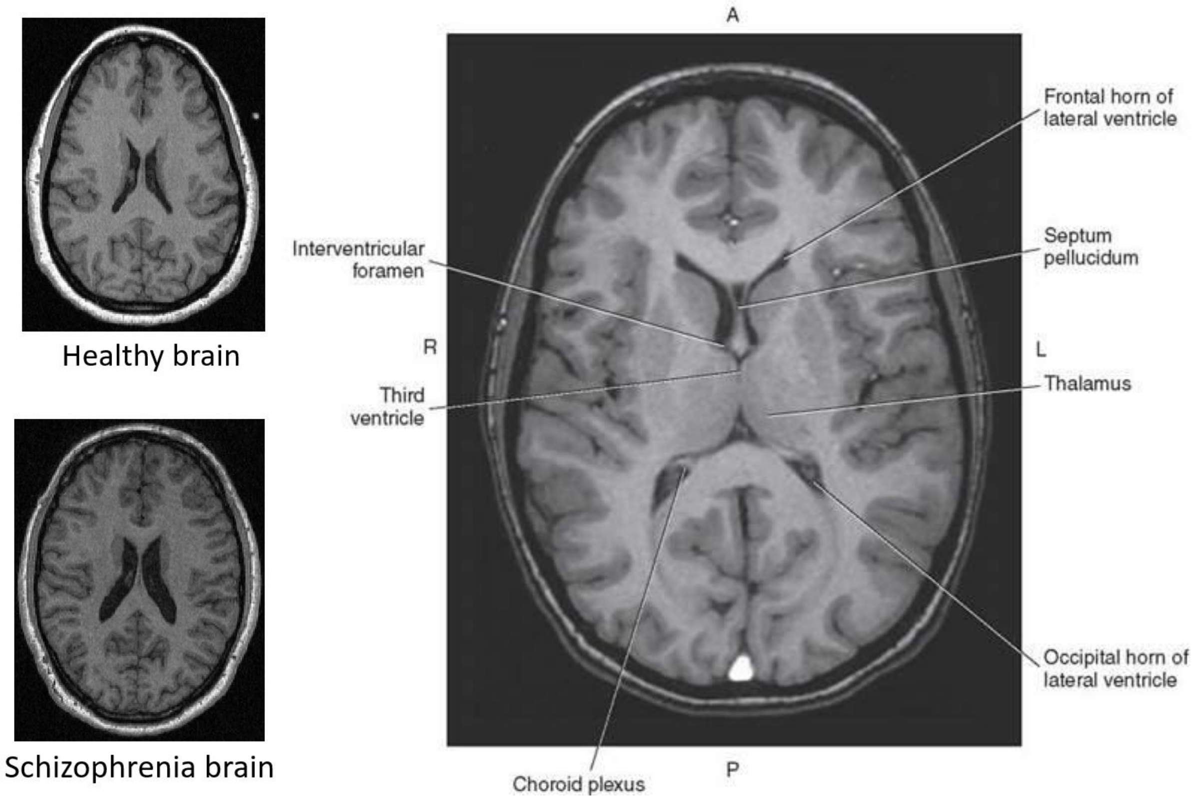

Fig. 1.1 Anatomy of brain along with Axial plane of Healthy and Schizophrenia Brain

Schizophrenia, a profound and debilitating mental disorder, challenges the realms of psychiatric diagnosis and treatment due to its inherent complexity. Characterized by a distorted interpretation of reality, individuals afflicted by schizophrenia exhibit a range of symptoms, including hallucinations, delusions, disorganized thinking, and emotional disturbances.

2. Literature Survey

In this chapter, the literature helpful in guiding this project is discussed. Let's delve into the collection of notable contributions from recent publications of standard journals.

Schizophrenia is a complex mental disorder that affects millions of individuals worldwide. People with schizophrenia are characterized by hallucinations, delusions, and cognitive impairments. Researchers have been actively studying schizophrenia and ways to diagnose this at an early stage [1]. An early and accurate diagnosis of schizophrenia is crucial for treatment. Several brain structures are affected in schizophrenia than expected from the overall reduction in volume [2]. Medical imaging systems aid healthcare workers and researchers improve diagnosis, while deep learning (DL) approaches assist the researchers in analyzing the neuroimaging data [3]. Modalities like Magnetic Resonance Imaging (MRI), and Electroencephalogram (EEG) have been used for diagnosis, while the ability of machine learning and deep learning models to distinguish between cases is higher in functional MRI (fMRI) modality than EEG modality [4]. MRI is an important technique in the diagnosis of schizophrenia. Various regions in the resting-state fMRI have shown 75% accuracy in distinguishing between control and schizophrenia subjects [5]. Traditional diagnostic methods, relying on clinical symptom assessment by psychiatrists, lack the precision demanded by the intricate neurobiological landscape of schizophrenia. Neuroimaging studies have consistently revealed structural abnormalities, emphasizing the need for advanced diagnostic tools.

The potential of machine learning is high in revolutionizing psychiatric diagnostics, with an increasing focus on their application in the realm of schizophrenia detection. Researchers have explored diverse datasets, including functional Magnetic Resonance Imaging (fMRI), to identify neurobiological markers indicative of schizophrenia. Chen et al. [6] used a ML model with two-sample t-tests to find, then used a different elimination method called recursive feature elimination (RFE) to eliminate redundant features, and SVM classifier to decide with gray matter and white matter features. It had the best classification accuracy of up to 85%. Mateos-Pérez et al. [7] presented that the reviewed studies have used fMRI dataset as well as sMRI dataset, mostly SVM classifier and Random Forests classifier for classification, reported a good range of accuracies from 61.8% to 95%. De Filippis et al. [8] includes 35 papers, out of which 8 work used sMRI dataset, 26 work fMRI, and one that utilized both, all had a minimum accuracy of 60%. Accuracy, Recall (Sensitivity) and specificity were considered as important metrics.

The DL model learns the features on its own and recorded a greater accuracy then ML work. Zheng J. el. at [9] used fMRI data that is preprocessed, the architecture used is VGG16 net to classify with transfer learning. The classification accuracy is 84.3%. Oh J et al. [10] trained a DL model on sMRI. The AUC score ranges from 0.71 to 0.90, the model had the capability to identify MRI images from schizophrenia in an unidentified MRI data set. Zhang et al. [11] hypothesized that 3D CNN architecture works better from T1-weighted MRI scans. The algorithm accurately distinguishes schizophrenia patients from healthy patients with area under the ROC curve of 0.987. M. Hu et al. [12] experimented with both ML and DL method. In DL method, 3D CNN was used, while in ML method Voxel-Based Morphometry (VBM) feature is given to SVM classifier. The 3D CNN model performed better than ML model.

Though ML and DL work showed good accuracy, in medical diagnosis, better accuracy is always needed. Liu et al. [13] presented a new feature fusion approach named as multi-layer visual feature fusion (MLVSF). It has been created to find qualities by merging low-level features, mid-level features, and deep-learning features. MLVSF is very useful to enhance feature discrimination for medical dataset. Manic et al. [14] proposed a scheme that includes distinct phases, such as MRI slice collection and pre-processing, VGG16 implementation for deep feature (DF) extraction, handcrafted feature (HF) collection, optimal feature selection using the mayfly algorithm, where both features are concatenated and sent to binary classifier. Independent assessment of the system, which uses individual and fusion method, shows that screening accuracy for schizophrenia is significantly higher than other techniques. The research, which examined 40 brain MRI images, found that DF achieved more than 90% accuracy, HF above 85%, and concatenated DF+HF above 95%. This system is relevant and provides a potential path for evaluating actual patients' brain MRI slices in future applications.

The CNN models rely heavily on the spatial relationship between pixels and struggles with long-range dependencies. While CNNs have been the traditional choice due to their ability to effectively capture spatial features in images through convolutional filters, transformers offer a novel approach that emphasizes capturing long-range dependencies and contextual information. The key aspect in transformers is the replacement of convolutional layers with transformer encoders. Transformers were originally designed for natural language processing tasks which can capture long-range information, and are later used for image classification tasks [15]. ViT processes the entire image by splitting it into multiple patches. Each patch, with its position embedding, is fed into a transformer encoder, a powerful module originally designed for natural language processing (NLP) tasks. Though Vision Transformer was trained on ImageNet, it can be used for medical images [16]. Vision Transformer (ViT), a novel deep learning architecture introduced in 2020, offers a compelling alternative to CNNs for image classification tasks. Though the convolutional filter in VGG19 excels at mapping the inputs and extracting prominent features, the ViT introduces a novel approach with its self-attention mechanism. With this self-attention mechanism and transformer encoder, ViT has exceeded the accuracy of VGG19 in brain tumor classification [17].

Tyagi A et al. [18] discovered that researchers have been using resting-state functional MRI more often and proposed that using ML and DL methods and medical databases can improve the detection of schizophrenia.

This literature survey encapsulates the evolving landscape of schizophrenia diagnostics, highlighting the necessity and potential of innovative methodologies, such as the proposed feature fusion-based approach, to address existing limitations and pave the way for more effective early interventions in schizophrenia detection.

3. Methodology

3.1 Dataset

A publicly available dataset was obtained from the Northwestern University Schizophrenia Data and Software Tool Federation via the BIRN Infrastructure, also known as NUSDAST [19]. This extensive dataset includes 125 MRI scans from schizophrenia patients and 125 MRI scans from control people. Notably, the MRI images in this dataset are high-resolution T1-weighted functional MRI and created using the Fast Low Angle Shot MRI (FLASH MRI) technique. This resource-rich dataset enables rigorous analysis, allowing for a thorough examination of the neuroanatomical features associated with schizophrenia and providing useful insights into potential diagnostic indicators within the context of high-quality anatomical imaging.

3.2 Data preprocessing

ITK-SNAP (Insight Segmentation and Registration Toolkit for SNAP) is a flexible medical image analysis program known for its user-friendly interface and robust features. ITK-SNAP, which was designed to segment and visualize three-dimensional biomedical pictures, allows for the accurate identification of structures inside medical scans The program supports a variety of image file types and excels at working with MRI images.

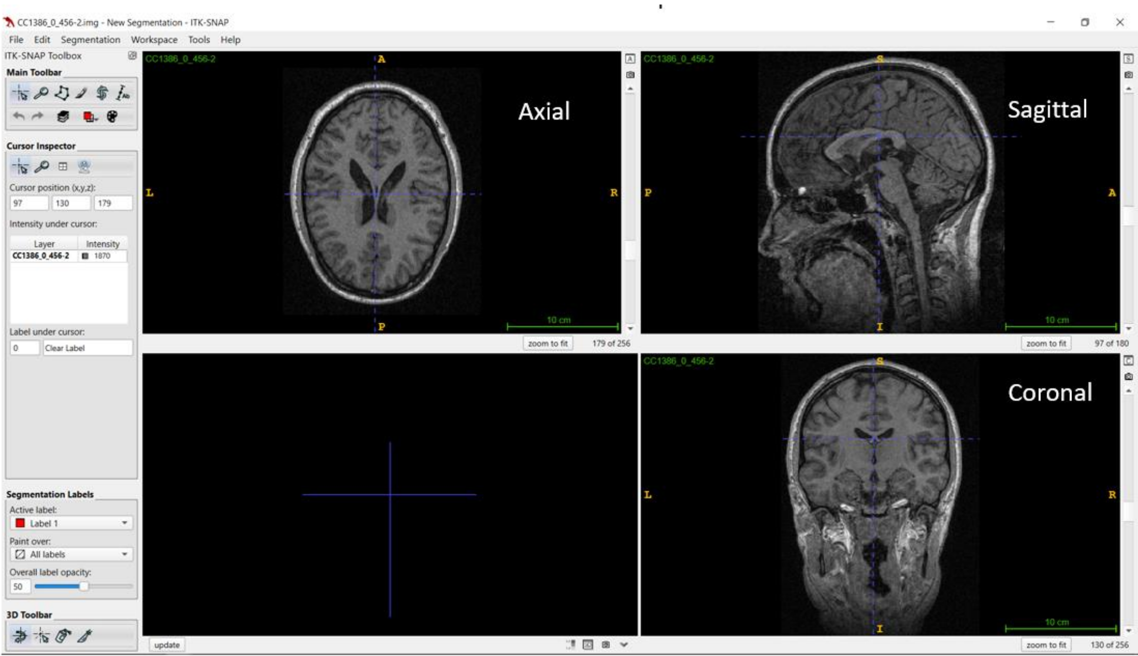

Fig. 3.1 ITK-Snap software with all planes

The dataset, which was originally stored in .img format, was smoothly displayed with the ITK-SNAP program, allowing for thorough analysis. To simplify model training, the axial plane of while MRI data was converted to.png image format, with a consistent size of 224*224 pixels. This change maintained the homogeneity and compatibility of following computing operations. Out of 3 planes, the axial plane, a cross-sectional view of the brain from a perspective perpendicular to the axis of the body is where the grey matter and white matter can be obtained. All the 125 schizophrenia and 125 healthy control patients’ scans have been converted into images to create a dataset of size 250. This dataset is used for training and testing of the model.

3.3 Machine Learning method

This approach captures and analyze intricate textures in brain imaging. Gray-Level Co-occurrence Matrices (GLCM) is employed meticulously to delineate the textural intricacies within the images. To enhance the depth of analysis, the discriminative power of Local Binary Pattern (LBP) and the informative gradients encapsulated in the Histogram of Gradients (HOG) is used. These advanced texture extraction techniques allow to unveil original textural nuances inherent in the images. As a crucial step in feature selection, the Recursive Feature Elimination (RFE) process is employed, strategically identifying and retaining the most salient features for our subsequent modelling endeavors.

The selected features, representing a synthesis of GLCM, LBP and HOG extracted nuances, serve as the foundation for binary classification model. This model is meticulously crafted to distinguish between Healthy and schizophrenia-afflicted brains, leveraging the amalgamated insights from the various imaging modalities. The binary classifier, equipped with the diverse features, stands poised to make nuanced and accurate distinctions between the intricate patterns associated with mental health and those indicative of schizophrenia.

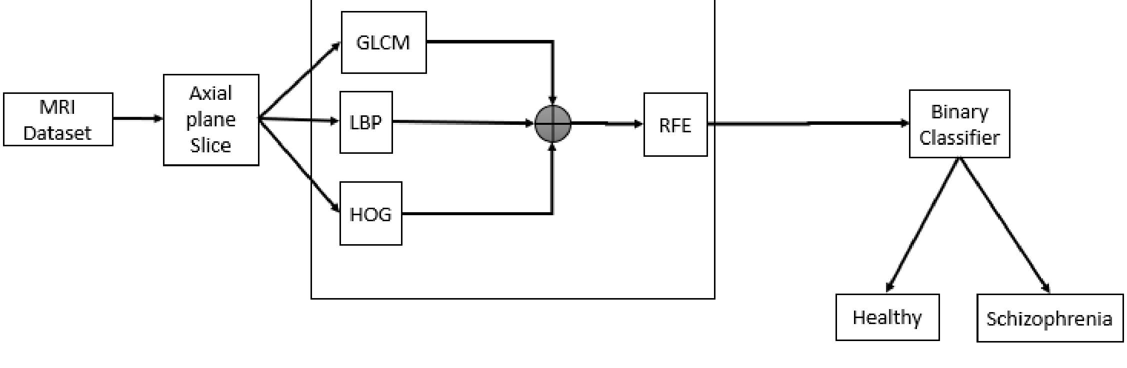

Fig. 3.2 Work flow of Machine Learning method

3.4 Deep Learning method

Various CNN models like VGG16, VGG19, ResNet152V2, InceptionV3, MobileNet and DenseNet121 are fine-tuned for this diagnostic task because of the smaller dataset that is available. All these models are trained on ImageNet dataset which has 14 million images with 1000 classes.

The pre-trained CNN network VGG16 and VGG19, created by Visual Geometry Group of Oxford University in 2014. VGG16 model is composed of 16 layers, including 13 convolutional layers and 3 fully connected layers. Each convolutional layer is followed by a max-pooling layer to reduce spatial dimensions. VGG19 model is composed of 16 layers, including 16 convolutional layers and 3 fully connected layers. Both consist of approximately 143 million parameters. As the input passes through the layers, it captures abstract and complex image features, which can be useful for downstream tasks. Though the network is trained on 1000 class general images, to utilize the capability of model to extract image features, transfer learning can be applied to the network for schizophrenia MRI image classification.

The Fine-tuning process involves unfreezing some convolutional layers and allow it to change during the training process with MRI images. The last 15 layers which consist of 3 set of 4 convolutional layers followed by 1 max-pooling layer and 3 fully connected layers which are followed by 1 SoftMax dense layer for final binary class prediction “healthy” or “schizophrenia”. In this way, other CNN models are also fine-tuned and their performance metrices for schizophrenia is analyzed.

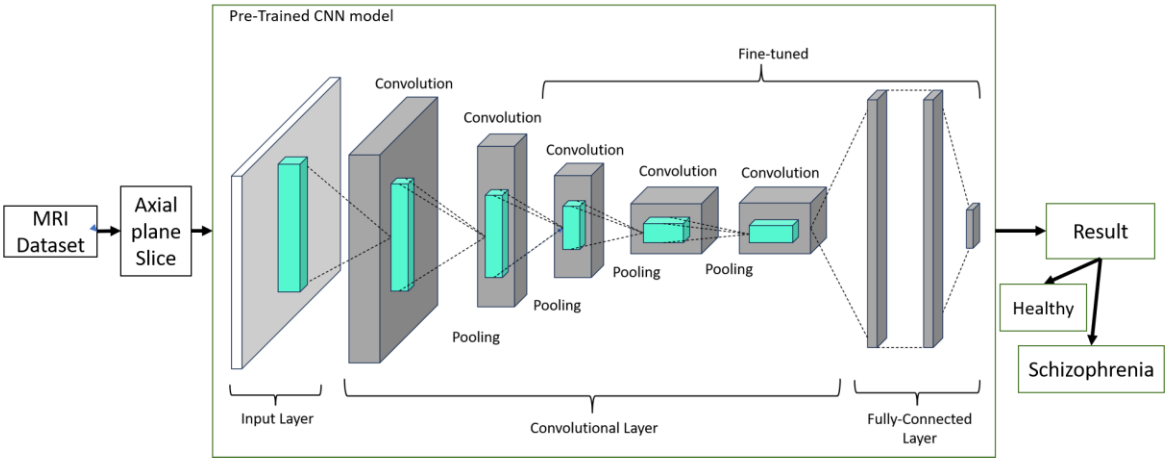

Fig. 3.3 Work flow of Deep Learning method

3.5 ML and DL method

One approach utilizes a pre-trained deep learning model known as VGG19 for feature extraction. This technique, called transfer learning, repurposes an existing model for a new task. Here, the VGG19, trained on a massive image dataset, extracts feature from the MRI scans.

The other approach employs three hand-crafted feature extraction techniques: Gray Level Co-occurrence Matrix (GLCM), Local Binary Pattern (LBP), and Histogram of Oriented Gradients (HOG). These techniques target specific image features potentially relevant for classification. For instance, GLCM extracts texture features, while LBP focuses on edges. The extracted features are then combined and used to train another machine learning model.

The ML method extracts low-level and mid-level features and DL method extracts deep features; both are complementary to each other. Combining these two models will yield a good accuracy model for schizophrenia detection. This fusion system leverages two main approaches for extracting features from the scans.

By combining deep learning and machine learning techniques for feature extraction, followed by a binary classifier for final categorization, this system offers a promising avenue for utilizing machine learning in medical diagnosis.

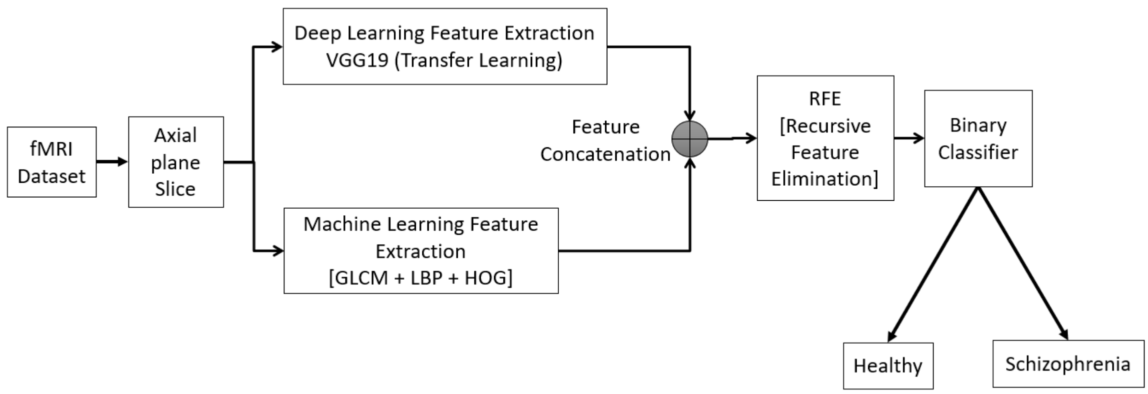

Fig. 3.4 Work flow of Machine Learning and Deep Learning method

3.6 Deployment using ML and DL method

The model created by combining ML and DL method is deployed in a website, which can be used for real-time schizophrenia diagnosis based on the fMRI scans. The most used fMRI scan format is ".img" or ".nii" format. Once that is fed into the CAD system, the correct axial plane slice will be saved in ".png" format.

The axial plane slice is sent to ML and DL model to extract low-level and deep features from the image. The "dense_1" layer of size 1024 is taken from the VGG19 model for the given input image. The ML features extracted from image are sent to RFE to choose top 512 features. These two features are combined into 1536 features through feature concatenation. Those 1536 features are further reduced to best 1024 features, which is given to binary classifier which is already trained and saved as ".pkl" file.

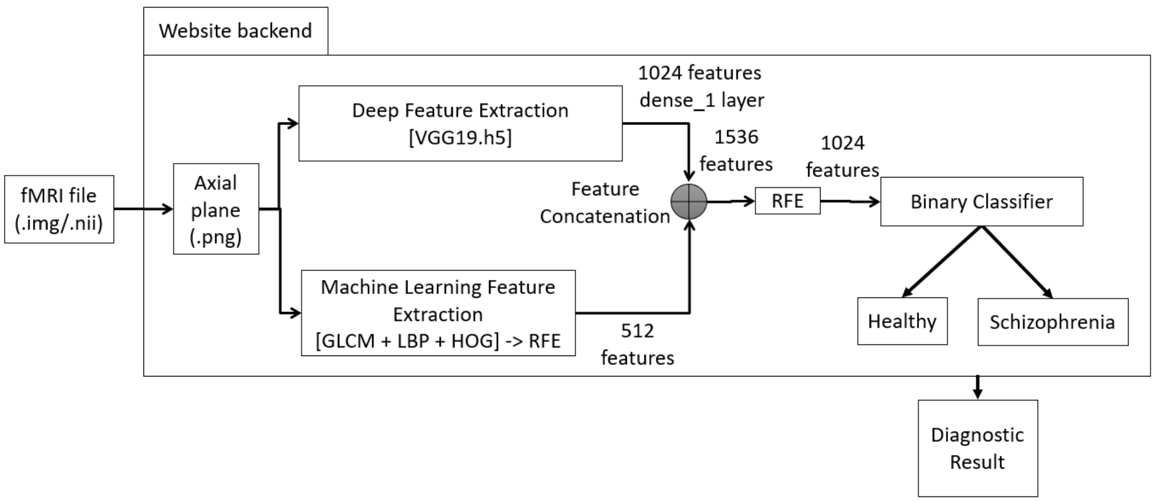

Fig. 3.5 Work flow of CAD system using ML and DL method

The final diagnosis, displayed as "Healthy" or "Schizophrenia" on the website, reflects the culmination of automated analysis based on ML and DL techniques. By automating the manual examination of fMRI scans for anomalies in the axial plane, the system expedites the diagnostic process, enabling timely interventions and treatment planning. Furthermore, the fusion of ML and DL techniques enhances the reliability and accuracy of the final diagnosis, providing healthcare professionals with invaluable insights for clinical decision-making.

3.7 Vision Transformer Method

The attention mechanism was first proposed in 'Attention is All You Need' [20]. The input image is sliced and linearly embedded with additional positional embedding to form sequence of vectors. This sequence of vectors is feed into the transformer encoder. Inside the encoder, the sequence is then normalized and given to a multi-head self-attention layer. The self-attention layer calculates attention weights of each pixel based on neighboring pixels, then the feed-forward layer implements transformation to the values. Thus, the multi-head self-attention layer process the different parts of image simultaneously.

Vision Transformer (ViT), a novel deep learning architecture introduced in 2020, offers a compelling alternative to CNNs for image classification tasks. Though the convolutional filter in VGG19 excels at mapping the inputs and extracting prominent features, the ViT introduces a novel method with its self-attention mechanism.

The self-attention in vision transformer is used to capture the contextual information in the input image. It calculates a weighted sum of the input data, with weights dependent on the relationship of the input characteristics. These weights prioritize relevant input features, resulting in more useful representations of the input data.

The output of feed-forward layer is normalized and given to multi-layer perceptron, which is feed to last transformer block. The multi-layer perceptron classifier takes up transformer block output. The final class prediction with the highest probability is the class of input image. Thus, the input image is classified as "healthy" or "schizophrenia".

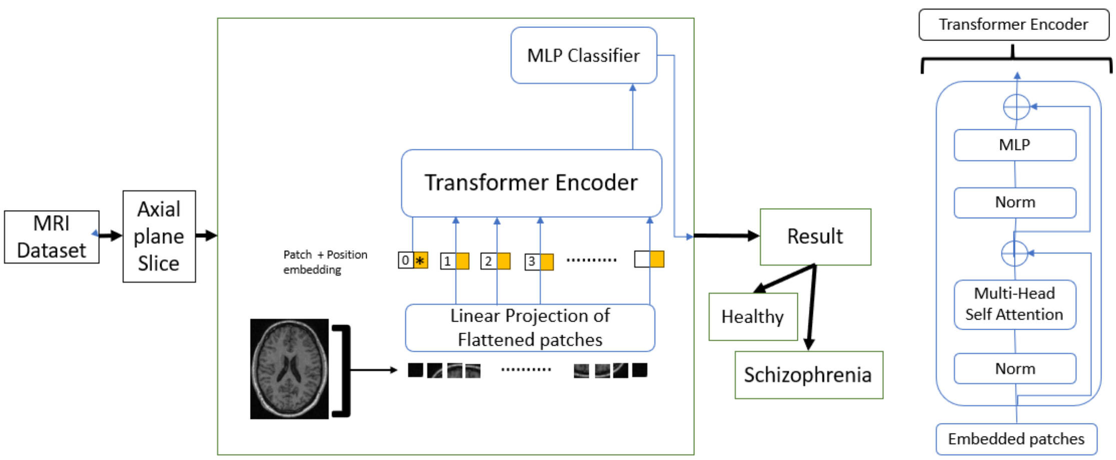

Fig. 3.6 Work flow of Vision transformer method

3.8 ML and ViT Method

The ML and DL method used the VGG19 model to extract the deep features. When the vision transformer method is applied to fMRI scans, the self-attention mechanism captures the contextual long-term dependencies in the axial plane. While the CNN model only holds the information in its convolutional layers, the image patches is fed into the multi-head self-attention mechanism in vision transformer. It allows the model to not only analyze individual patches but also understand the relationships between them. By attending to information from other patches, the self-attention mechanism captures the long-term dependencies across different brain regions within the axial plane slices of the fMRI scan. This provides crucial insights into how different brain areas communicate and coordinate with each other.

The machine learning method capture the short-term visual information which complements the contextual information from vision transformer. The axial plane slice is sent through the self-attention heads which holds the gray matter, white matter and ventricle information. It is concatenated with the machine learning features to hold the overall information.

RFE acts as a filter, eliminating redundant or irrelevant features from the combined pool. By selecting the most informative features, RFE ensures that the classification stage focuses on the most impactful information within the brain scan. These best features are fed to various binary classifiers to train them for those particular features. The best classifier is saved as ".pkl" file. The binary class prediction "healthy" or "schizophrenia" from this method is reliable as the self-attention mechanism for images captures information better than other CNN models.

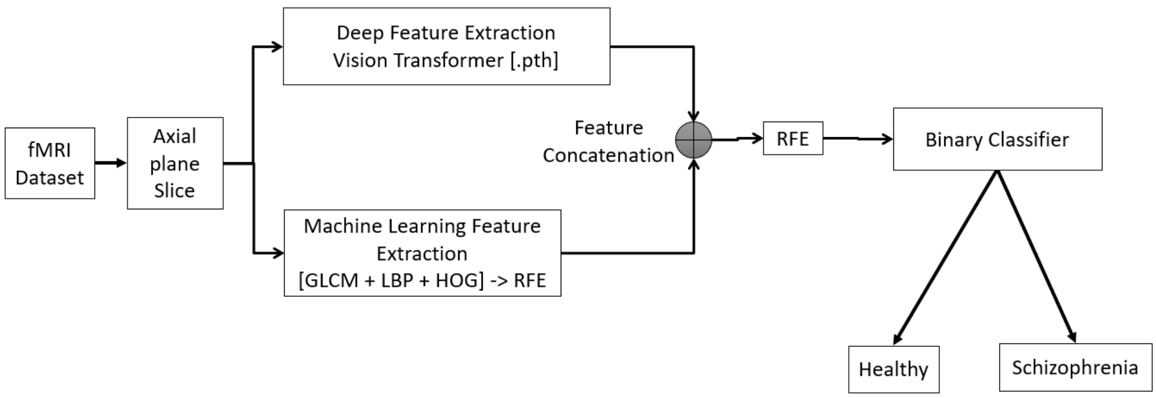

Fig. 3.7 Work flow of ML and ViT method

3.9 Deployment using ML and ViT method

The CAD (Computer-Aided Diagnosis) system represents a significant advancement in real-time schizophrenia diagnosis using fMRI scans. By integrating machine learning and ViT techniques within a web-based platform, the system streamlines the diagnostic process and enhances reliability.

The system begins by converting the fMRI scans into correct axial plane slice is extracted and saved as a ".png" image. This image is then passed through both an ML model and a Vision Transformer (ViT) model to extract features.

In particular, the ViT model extracts deep features from the image, focusing on the 12th layer of size 3072 from the VGG19 model. Concurrently, the ML model extracts low-level features from the image, which are then refined using Recursive Feature Elimination (RFE) to select the top 512 features. These features are combined through feature concatenation, resulting in a total of 3584 features.

To reduce dimensionality and enhance computational efficiency, the feature set is further refined to the best 1536 features. These features are then input into a binary classifier, previously trained and saved as a ".pkl" file, which provides the final diagnosis of "Healthy" or "Schizophrenia."

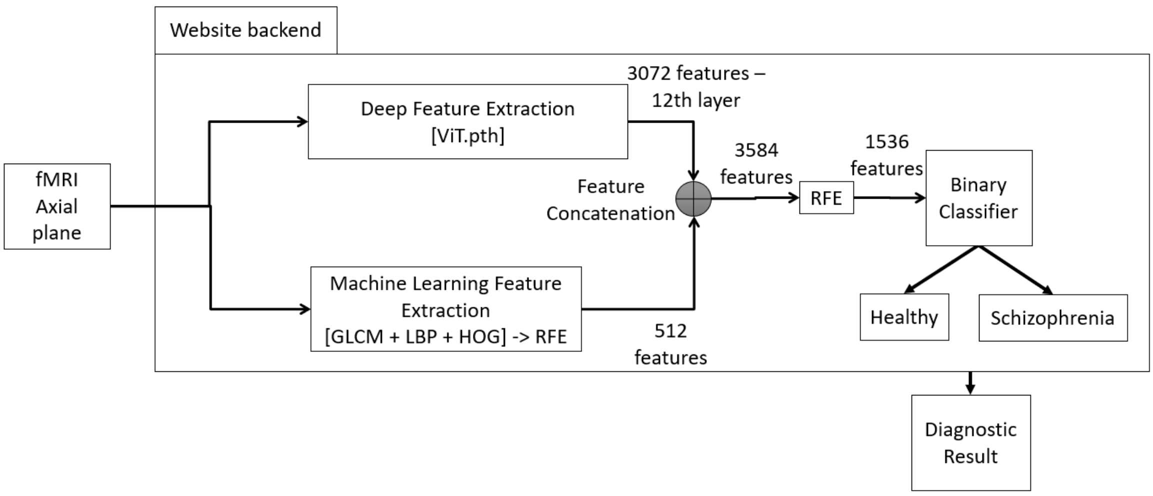

Fig. 3.8 Work flow of CAD system using ML and ViT method

The automated examination of fMRI scans in the axial plane significantly improves efficiency and reduces the reliance on manual inspection. Overall, the CAD system represents a powerful tool for real-time schizophrenia diagnosis, leveraging the capabilities of ML and DL to streamline the diagnostic process and improve patient care.

4. Results

4.1 Machine Learning Model

The machine learning model architecture adopts a multifaceted method by integrating texture features such as GLCM, LBP, HOG. These features undergo Recursive Feature Elimination (RFE) for efficient selection, a critical step in mitigating dimensionality challenges. The RFE process aids in identifying the most discriminative features, optimizing the model's ability to capture essential information from the input data.

This architecture introduces a nuanced strategy by exploring various combinations of feature sets, encompassing single, double, and all features. Additionally, the model undergoes evaluation with Principal Component Analysis (PCA) and RFE individually, offering a comprehensive examination of their impact on classification performance. The binary classifier at the model's core distills these diverse feature sets into a final diagnostic outcome, discerning between "healthy" and "schizophrenia."

This comprehensive method not only leverages the power of texture features and RFE for effective feature selection but also systematically explores the influence of different feature combinations and dimensionality reduction techniques, contributing to the model's robustness and diagnostic accuracy.

Table 4-1 Performance analysis of single feature with linear classifiers

Single Feature	**Logistic Regression**			**Random Forest**			**SVM**		
Metrics	Acc	Sens	Spe	Acc	Sens	Spe	Acc	Sens	Spe
GLCM	0.70	0.70	0.70	0.62	0.52	0.62	0.64	0.65	0.63
LBP	0.64	0.65	0.63	0.76	0.74	0.78	0.72	0.74	0.70
HOG	0.84	0.78	0.89	0.82	0.78	0.85	0.78	0.83	0.74

[a.] Acc- Accuracy, Sens- Sensitivity, Spe – Specificity

Table 4-2 Performance analysis of single feature with tree-based classifiers

Single Feature	**XGBoost**			**Adaboost**			**Histogram-based Gradient Boosting**		
Metrics	Acc	Sens	Spe	Acc	Sens	Spe	Acc	Sens	Spe
GLCM	0.58	0.57	0.56	0.58	0.61	0.56	0.58	0.61	0.56
LBP	0.78	0.78	0.62	0.62	0.57	0.67	0.72	0.74	0.70
HOG	0.80	0.78	0.80	0.76	0.83	0.70	**0.86**	**0.83**	**0.89**

[b.] Acc- Accuracy, Sens- Sensitivity, Spe – Specificity

Table 4-3 Performance analysis of double feature with linear classifiers

Double Feature	**Logistic Regression**			**Random Forest**			**SVM**		
Metrics	Acc	Sens	Spe	Acc	Sens	Spe	Acc	Sens	Spe
GLCM + LBP	0.66	0.65	0.67	0.70	0.65	0.74	0.72	0.70	0.74
LBP + HOG	0.84	0.83	0.86	0.82	0.78	0.84	**0.85**	**0.87**	**0.88**
LBP + HOG	0.78	0.78	0.81	0.78	0.74	0.81	0.76	0.83	0.78

[c.] Acc- Accuracy, Sens- Sensitivity, Spe – Specificity

Table 4-4 Performance analysis of double feature with tree-based classifiers

Double Feature	**Extra Tree**			**Adaboost**			**Histogram-based Gradient Boosting**		
Metrics	Acc	Sens	Spe	Acc	Sens	Spe	Acc	Sens	Spe
GLCM + LBP	0.72	0.65	0.78	0.62	0.61	0.63	0.70	0.65	0.78
LBP + HOG	0.76	0.70	0.76	0.84	0.87	0.84	0.84	0.78	0.89
LBP + HOG	0.82	0.70	0.85	0.74	0.78	0.70	**0.84**	**0.78**	**0.86**

[d.] Acc- Accuracy, Sens- Sensitivity, Spe – Specificity

Table 4-5 Performance analysis of all features with PCA/RFE and linear classifiers

All Feature	**Logistic Regression**			**Random Forest**			**SVM**		
Metrics	Acc	Sens	Spe	Acc	Sens	Spe	Acc	Sens	Spe
GLCM + LBP + HOG	0.65	0.64	0.65	0.76	0.62	0.81	0.80	0.82	0.78
GLCM + LBP + HOG [PCA]	0.84	0.89	0.79	0.76	0.84	0.68	0.85	0.89	0.82
GLCM + LBP + HOG [RFE]	0.87	0.89	0.84	0.88	0.92	0.84	0.87	0.82	0.82

e. Acc- Accuracy, Sens- Sensitivity, Spe – Specificity

Table 4-6 Performance analysis of all features with PCA/RFE and tree-based classifiers

All Feature	**XGBoost**			**Adaboost**			**Histogram-based Gradient Boosting**		
Metrics	Acc	Sens	Spe	Acc	Sens	Spe	Acc	Sens	Spe
GLCM + LBP + HOG	0.76	0.70	0.76	0.77	0.69	0.86	**0.87**	**0.85**	**0.89**
GLCM + LBP + HOG [PCA]	0.73	0.62	0.84	0.60	0.62	0.76	0.73	0.70	0.76
GLCM + LBP + HOG [RFE]	0.79	0.73	0.84	0.84	0.81	0.87	**0.89**	**0.89**	**0.89**

f. Acc- Accuracy, Sens- Sensitivity, Spe - Specificity

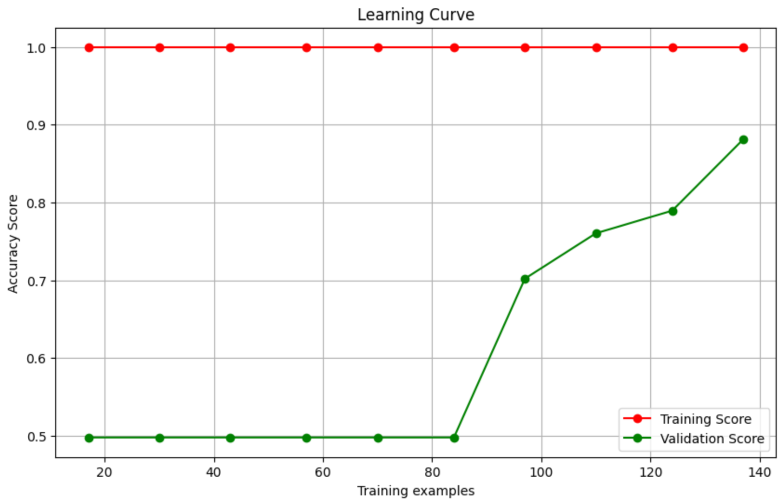

Fig. 4.1 AUC-ROC plot of all features with RFE and Histogram-based Grading Boosting Classifier

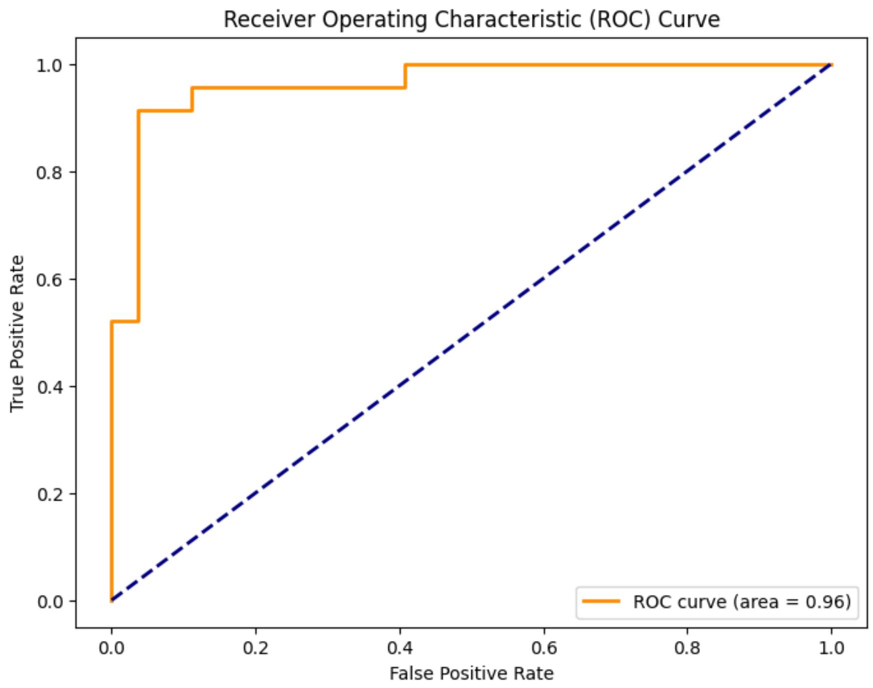

Fig. 4.2 AUC-ROC plot of all feature with RFE and Histogram-based Grading Boosting Classifier

The proposed method integrates diverse feature extraction techniques, combining GLCM, LBP, HOG. Feature selection employs Recursive Feature Elimination (RFE), optimizing the model's efficiency by identifying and retaining the most informative features. The dataset is partitioned into a training set (70%) and a test set (30%), ensuring robust evaluation of the developed model.

The classification task is executed using a Histogram-based Gradient Boosting Classifier, showcasing its prowess in handling complex relationships within the feature space. The achieved performance metrics underscore the model's proficiency, with an accuracy, precision, F1 score, recall and specificity at 89%. The confusion matrix provides a comprehensive insight into the model's classification outcomes. Moreover, the Area Under the Curve (AUC-ROC) Score stands impressively high at 0.96, affirming the model's robust discriminatory capabilities and its ability to distinguish between classes.

The combination of GLCM, LBP, and HOG features, coupled with RFE for optimal feature selection, and a Histogram-based Gradient Boosting Classifier, results in a well-performing model, demonstrated by its high accuracy, precision, recall, specificity, and AUC-ROC Score.

4.2 Deep Learning Model

With the use of 15 GB T4 GPU RAM in Google Colab, various CNN models have been implemented and all have showed a good accuracy, sensitivity and specificity.

For pretrained CNN models, 175 training and 75 testing images from the available dataset was used as for the train-test split of 70-30 and the images were resized to the same as the ImageNet dataset's (224,224,3). The convolutional layers of CNN model are imported without the fully connected layers at the top. Specifically, the weights of the few convolutional layers are unfrozen for training. The base model is connected with three dense layers of size 1024, 2048 and 4096. The final dense layer of size 2 with "SoftMax" activation function is connected to binary classifier. The dropout rate has been set to 0.5 to generalize the model with other hyper-parameters of pretrained models.

Table 4-7 Performance analysis of all CNN models

CNN Model	**Performance metrics**					
	Accuracy	Sensitivity	Specificity	Precision	F1 score	AUC-ROC Score
VGG16	86.6%	86.8%	86.4%	87%	0.868	0.92
VGG19	**90.6%**	88.5%	92.5%	91%	0.898	**0.954**
ResNet152V2	86.6%	81.5%	91.8%	91%	0.861	0.911
InceptionV3	89.3%	89%	92%	91%	0.904	0.95
MobileNet	87.9%	83.7%	92%	91%	0.873	0.96
DenseNet121	85.3%	89%	81%	82%	0.857	0.932

In assessing the test accuracy, VGG19 emerges as the top performer among the models evaluated, achieving an impressive accuracy of 90.6%. This demonstrates the robustness of VGG19 in accurately classifying MRI images into their respective categories. However, all models exhibit commendable accuracy levels, ranging from 85.3% to 90.6%, showcasing their effectiveness in capturing relevant features for classification tasks.

Regarding precision, VGG19 and InceptionV3 stand out with precision values of 91.1% and 91.6%, respectively. Precision is a crucial metric in medical image classification, as it indicates the model's ability to avoid false positives. The high precision values attained by these models underscore their capability to minimize misclassifications, which is paramount in medical diagnosis scenarios.

Sensitivity, a measure of the model's ability to correctly identify positive cases, is particularly crucial in medical image classification, where missing a positive diagnosis can have severe consequences. Here, InceptionV3 and DenseNet121 achieves the highest sensitivity of 89% followed by VGG19 with 88.5%. These models demonstrate a strong capability to detect instances of schizophrenia in MRI images.

Specificity, on the other hand, indicates the model's ability to correctly identify negative cases. VGG19 exhibits the highest specificity among the models, with a value of 92.5%. This signifies the model's proficiency in distinguishing healthy cases from those with schizophrenia, thereby reducing false alarms and unnecessary interventions.

The AUC-ROC score provides a comprehensive assessment of a model's ability to distinguish between classes across various thresholds. In this regard, VGG19 and DenseNet121 exhibit good performance, with AUC-ROC scores of 0.954 and 0.946, respectively. These high scores highlight the models' effectiveness in capturing the underlying patterns in MRI images and making accurate classifications.

Each model demonstrates impressive capabilities across various performance metrics, VGG19 stands out as the optimal choice for schizophrenia image classification among the CNN models. Due to high accuracy, precision, and specificity, VGG19 showcases robust performance in distinguishing between healthy and schizophrenia MRI images. Its superior performance across multiple metrics underscores its effectiveness and reliability, making it the preferred model for accurate and precise diagnosis.

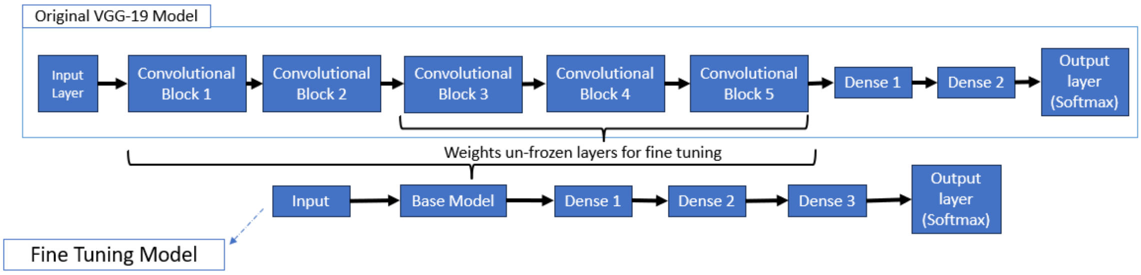

Fig. 4.3 Transfer Learning VGG-19 model

The model undergoes meticulous hyperparameter tuning, a critical optimization process that involves adjusting parameters such as learning rate, loss function, epoch, and batch size. These hyperparameter adjustments fine-tune the model's training dynamics, enhancing its ability to capture complex patterns within the schizophrenia-related data.

The incorporation of new fully connected layers facilitates the adaptation of the pre-trained model to the specific task of schizophrenia detection. This hybrid architecture, blending pre-trained convolutional layers with task-specific fully connected layers, strikes a balance between leveraging existing knowledge and tailoring the model to the intricacies of the medical image classification problem. The culmination of these design choices and optimization efforts contributes to a robust and effective deep learning model for schizophrenia detection.

Table 4-8 Output shape and trainable parameters of VGG19 model

Layers	**Output Shape**	**Parameter**
VGG19(base model)	(7,7,512)	20024384
Flatten	(25088)	0
dense_1 (Dropout=0.5)	(1024)	25691136
dense_2 (Dropout=0.5)	(2048)	2099200
dense_3	(4096)	8392704
dense_4 (SoftMax)	(2)	8194

The deep learning model, based on a pre-trained VGG19 architecture with frozen base layers and augmented with 117,200 images, demonstrates impressive performance in schizophrenia detection. The test results accuracy is 90.6%, affirming the model's ability to effectively discern between healthy and schizophrenia cases.

Precision, a measure of correctness in predicting positive instances, stands at 91%, indicating a high level of accuracy in identifying actual schizophrenia cases among the predicted positive results. Sensitivity, also known as recall, at 88.5%, reflects the DL model's capability to correctly identify the majority of actual positive cases. Specificity, with a commendable value of 92.5%, underscores the DL model's proficiency in correctly classifying true negative instances.

The F1 score, which balances precision and recall, is calculated at 0.89, illustrating a harmonious blend of accuracy in both positive and negative predictions. The confusion matrix further delineates the model's performance, revealing 37 true negatives, 31 true positives, 3 false positives, and 4 false negatives.

The deep learning model, fine-tuned through hyperparameter optimization and augmented with an extensive dataset, demonstrates robust performance with high accuracy, precision, sensitivity, and specificity in the challenging task of schizophrenia detection. The detailed results provide a comprehensive evaluation of the model's effectiveness in both positive and negative predictions, affirming its potential for clinical applications.

The analysis of the training and validation graphs indicates the model successfully avoided overfitting. This means the model learned from the training data without memorizing it too specifically. As a result, the model should be able to perform well on unseen data, which is a key sign of good generalization.

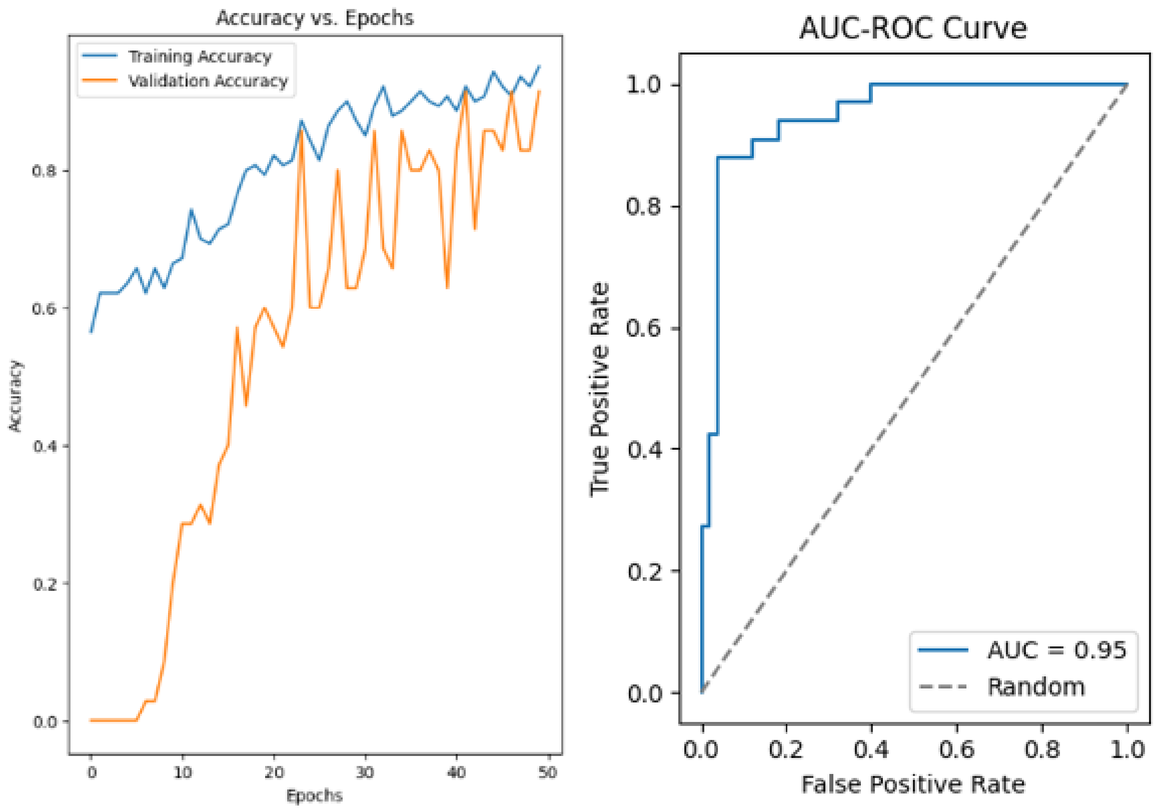

Fig. 4.4 Training and validation accuracy plot and AUC-ROC plot of VGG19 model

4.3 Feature Fusion of ML and DL Model

The Machine Learning model with GLCM, LBP, HOG have shown an accuracy of 89% with the visual features. GLCM examines the spatial relationship between pixel pairs, capturing texture information crucial for identifying abnormalities in brain scans. LBP focuses on micro-patterns by comparing each pixel with its neighbors, encoding this relationship into a binary pattern, which helps in distinguishing subtle textural differences. HOG, on the other hand, captures the distribution of gradient orientations, emphasizing edge structures to highlight the shape and layout of brain structures in MRI images.

The Deep Learning model VGG19 extracted the spatial relationship between pixels through its convolutional filters have shown an accuracy of 91%. This model extracts complex patterns and structures within the MRI scans that might not be discernible using traditional feature extraction techniques alone. The deep features from VGG19 thus provide a robust understanding of the spatial arrangements and intricate details of brain tissues, enhancing the diagnostic accuracy. These two features are complimentary to each other and holds different information about the MRI scans. Thus, integration of both models and selection of the best features using RFE is done.

Table 4-9 Performance analysis of ML and DL features with RFE

ML + DL Feature	**ML (GLCM+LBP+HOG) + DL (VGG-19) ->[RFE]**					
Classifier	Logistic Regression	Random Forest	SVM	XGBoost	Adaboost	HistGrading Boosting
Accuracy	0.89	**0.94**	0.80	0.90	0.91	0.93
Sensitivity	0.89	**0.94**	0.83	0.86	0.89	0.91
Specificity	0.89	**0.94**	0.77	0.94	0.94	0.94

Two linear binary classifiers and four tree-based binary classifies are trained on the concatenated features to detect whether the given input scan belongs to “Healthy” or “Schizophrenia”. Out of all classifier, the random forest classifier has shown an accuracy of 94% which is higher than individual ML model and individual DL model. The integration of diverse feature types allowed the model to leverage both texture and deep spatial information, significantly improving the classification accuracy for detecting schizophrenia in MRI scans.

4.4 Website using ML and DL Model

The deployment of machine learning (ML) and deep learning (DL) models within a website framework using Flask for medical diagnosis is a significant step towards enhancing the efficiency and accessibility of psychiatric evaluations. By integrating technologies such as ML and DL with web-based platforms, healthcare professionals can leverage automated analysis to assist in diagnosing conditions based on fMRI scans.

In this system, the VGG19 model serves as a powerful deep learning architecture for feature extraction from fMRI images. VGG19, known for its deep architecture and excellent performance in image recognition tasks, provides a robust foundation for extracting relevant features from the scans. These features are crucial for subsequent analysis and diagnosis.

The ML model complements the DL model by providing a binary classification capability. It takes the extracted features from the chosen axial slice of the fMRI scan and predicts the likelihood of a specific diagnosis. This binary classifier enables rapid decision-making by providing psychiatric professionals with actionable insights based on the processed data.

A key challenge in this system is ensuring that the appropriate axial slice is chosen for analysis. The accuracy and reliability of the diagnostic results heavily depend on selecting the slice that best represents the condition of interest. This requirement underscores the importance of thorough training and validation of the ML and DL models on a diverse range of fMRI scans to ensure robust performance across different patient cases.

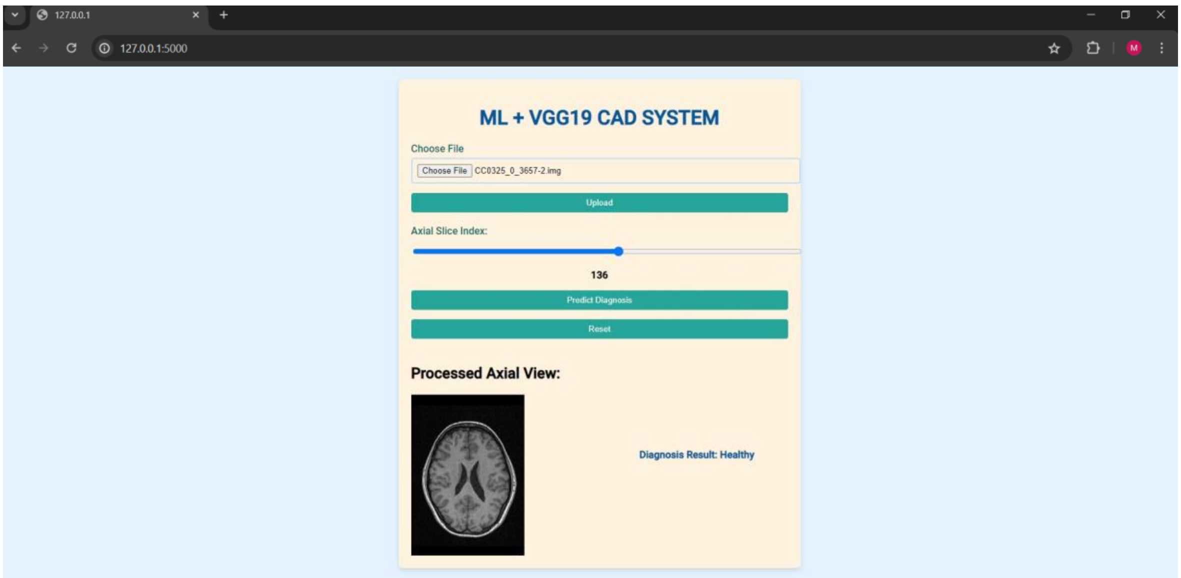

Fig. 4.5 UI of Website using ML and VGG19 with Healthy diagnostic result

Upon selecting the axial slice and initiating the diagnosis process by clicking the "Predict diagnosis" button, the system sends the corresponding image file to both the ML and DL models for feature extraction. The processing time for generating diagnostic results typically ranges from 3 to 5 seconds, providing a timely response to healthcare professionals.

The integration of these technologies into a web-based platform offers several advantages. It enhances the accessibility of psychiatric evaluations by enabling remote access to diagnostic tools, thereby facilitating timely interventions and improving patient outcomes. Additionally, the real-time processing capability empowers healthcare professionals with actionable insights during clinical assessments, ultimately contributing to more effective and efficient patient care.

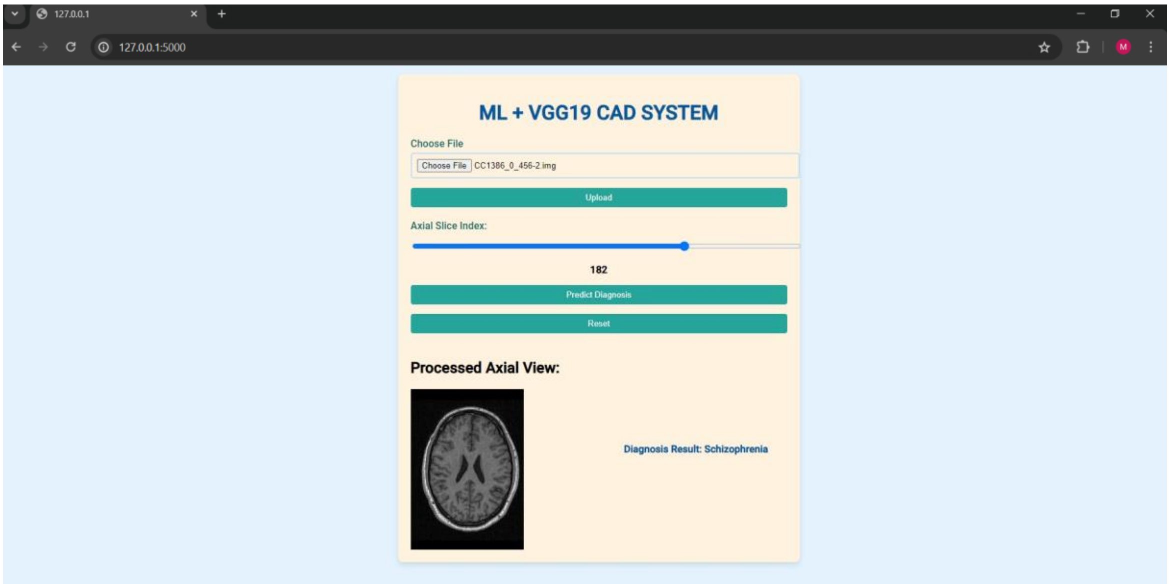

Fig. 4.6 UI of Website using ML and VGG19 with Schizophrenia diagnostic result

The deployment of ML and DL models within a Flask-based website framework for psychiatric diagnosis represents a significant advancement in healthcare technology. By leveraging these technologies, healthcare professionals can access automated diagnostic tools that enhance the accuracy, efficiency, and accessibility of psychiatric evaluations based on fMRI scans.

4.5 Vision Transformer Model

Fine-tuning is a pivotal step in leveraging pre-trained models like Vision Transformer (ViT) for specific tasks. In this scenario, the dataset comprising 150 training images, 30 validation images, and 70 testing images undergoes preprocessing to conform to the input shape required by the transformer architecture. Each transformer block comprises multiple layers, including self-attention mechanisms and feed-forward neural networks. During fine-tuning, the parameters of these transformer blocks are adjusted to adapt the model's representations to the specifics of the target task.

The test accuracy stands at an impressive 95.7. Precision, a measure reflecting the model's overall performance, is reported at a commendable 94.2%. The model's exceptional sensitivity score of 97.1%, underscoring its proficiency in accurately diagnosing conditions.

Furthermore, the specificity score is reported at 94.2%, affirming the model's capability to discern accurately between healthy control and schizophrenia MRI images. Fig. 6 shows the AUC-ROC score of 0.957 is particularly noteworthy, highlighting the model's remarkable ability to leverage the self-attention mechanism within the Vision Transformer architecture. This mechanism allows the model to focus on relevant features across different regions of the input images simultaneously, enabling it to extract meaningful representations effectively.

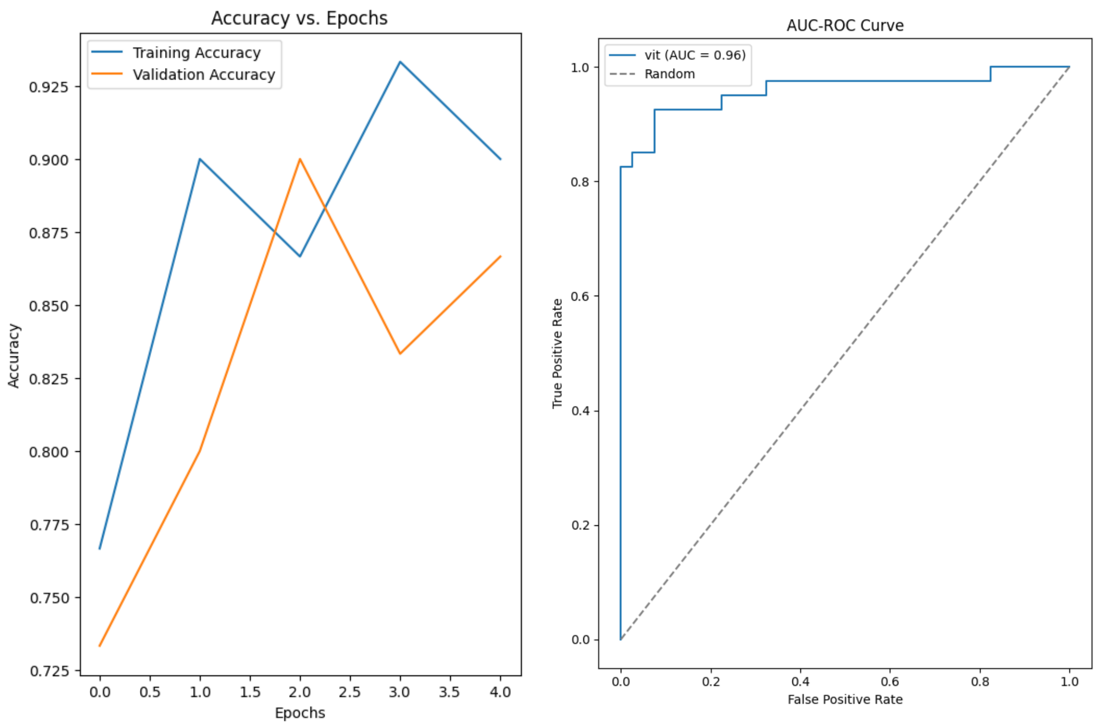

Fig. 4.7 Training and validation accuracy plot and AUC-ROC plot of ViT model

The pre-trained ViT model and fine-tuned image-net's VGG19 have performed well. But when in real-time applications, high accuracy, sensitivity, and specificity is crucial which ensures the reliability on the tool.

Table 4-10 Performance analysis of VGG19 and ViT

Model	**Performance metrics**					
	Acc	***Sens***	***Spec***	**Prec**	***F1 score***	***AUC-ROC Score***
VGG19	90.6%	88.5%	92.5%	91%	0.898	0.954
ViT	**95.7%**	97%	94%	94%	0.957	**0.957**

CNN models and ViT model have the ability to differentiate between schizophrenia and healthy control fMRI scans. But ViT model's performance metrics like accuracy, sensitivity and specificity ensures the accurate classification of positive cases.

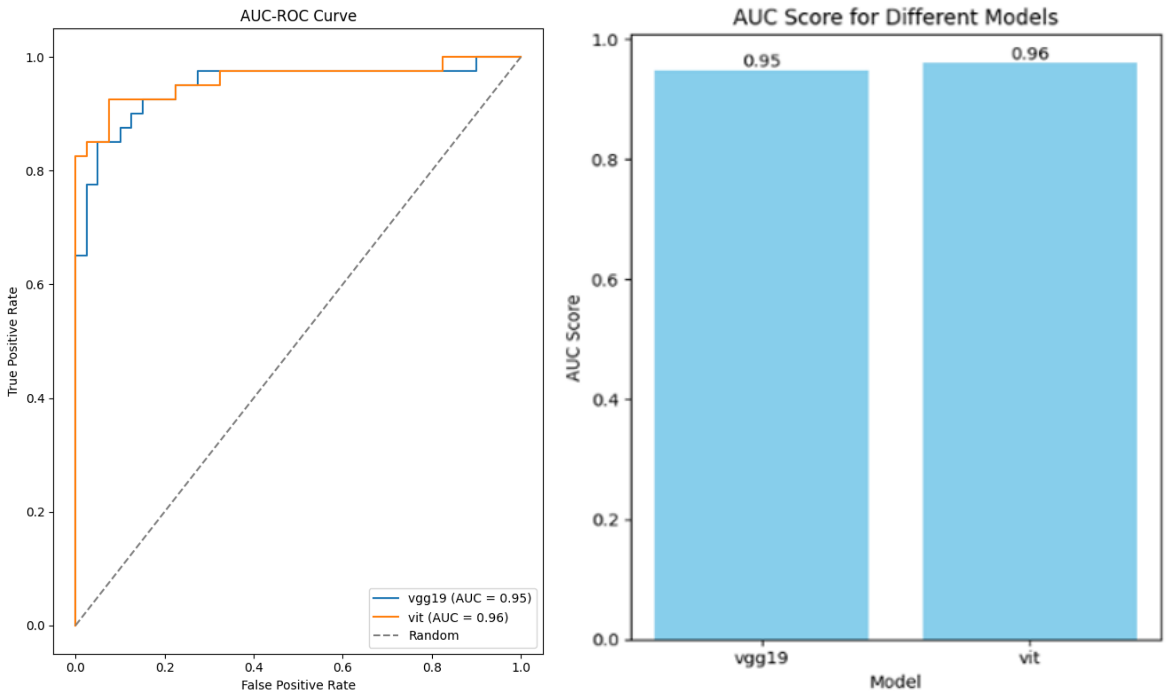

Fig. 4.8 AUC-ROC plots of VGG19 vs ViT model

4.6 Feature Fusion of ML and ViT Model

While both Vision Transformer (ViT) and VGG19 models have shown commendable performance in classifying MRI scans, ViT offers several distinct advantages. ViT leverages the self-attention mechanism, which allows it to capture global context and relationships within the entire image, providing a more comprehensive understanding of spatial arrangements and intricate details of brain tissues. This global perspective is a significant enhancement over the local receptive fields of VGG19's convolutional layers, which primarily capture local patterns.

Furthermore, ViT's architecture is highly flexible and can adapt more effectively to diverse image features, making it particularly well-suited for complex medical imaging tasks. The performance metrics underscore these benefits: ViT achieved a higher test accuracy (95.7%) compared to VGG19 (90.6%), along with superior sensitivity (97% vs. 88.5%) and specificity (94% vs. 92.5%). These metrics highlight ViT's proficiency in both identifying true positive cases and accurately discerning between healthy and schizophrenia-affected scans. Thus, ViT not only enhances the diagnostic accuracy but also ensures a more reliable and robust classification, crucial for real-time medical applications.

Thus, the ML model and ViT model was concatenated and fed to RFE to select best features. Among other classifier, Adaboost classifier have showed the highest accuracy, sensitivity, specificity of 95%. These results indicate that the AdaBoost model was exceptionally proficient in correctly identifying both healthy and schizophrenia-affected MRI scans, making it the most reliable classifier in this study.

Table 4-11 Performance analysis of ML and ViT features with RFE

ML+DL Feature	**ML (GLCM+LBP+HOG) + ViT ->[RFE]**					
Classifier	Logistic Regression	Random Forest	SVM	XGBoost	Adaboost	HistGrading Boosting
Accuracy	0.79	0.89	0.79	0.93	**0.95**	0.93
Sensitivity	0.93	0.90	0.93	0.93	**0.95**	0.93
Specificity	0.65	0.88	0.65	0.93	**0.95**	0.93

4.7 Website using ML and ViT Model

The deployment of machine learning (ML) and Vision Transformer (ViT) models within a Flask-based website framework for psychiatric diagnosis signifies a significant advancement in medical technology. This integration addresses the limitation of the previous system, which relied on VGG19 for feature extraction from fMRI scans. By incorporating ViT, the system gains the ability to overcome the restriction of selecting the appropriate axial slice for analysis, offering a more flexible and accurate diagnostic method.

ViT, a state-of-the-art deep learning architecture, revolutionizes image processing tasks by leveraging the power of self-attention mechanisms. Unlike traditional convolutional neural networks (CNNs) like VGG19, which operate on fixed-size grids of image patches, ViT processes images as sequences of tokens. This tokenization enables ViT to capture long-range dependencies within the image, allowing it to effectively extract features from fMRI scans regardless of the specific axial slice chosen.

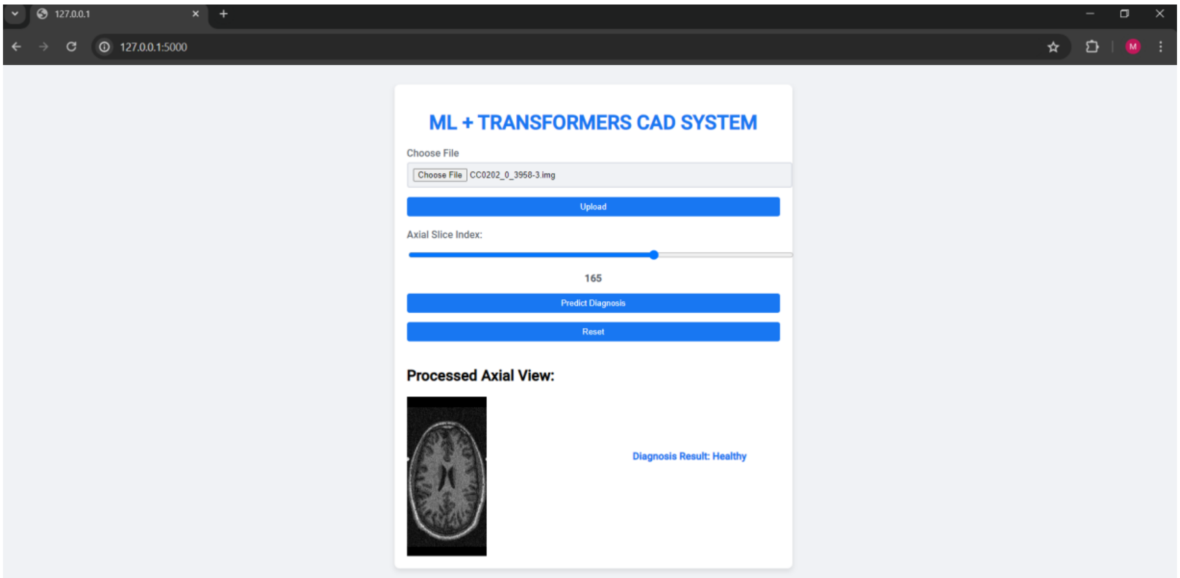

Fig. 4.9 UI of Website using ML and ViT with Healthy diagnostic result

In this system, psychiatrists have the flexibility to select an approximate range of around 5 correct axial slice indices where the ventricles are correctly visible. This flexibility alleviates the burden of precisely choosing a single slice for analysis, which was a limitation of the previous VGG19-based system. By providing a range of slices, the psychiatrist can ensure that the diagnostic process encompasses the relevant anatomical information necessary for accurate diagnosis.

Upon selecting the approximate range of axial slices and initiating the diagnosis process by clicking the "Predict diagnosis" button, the system sends the corresponding image file to both the ML and ViT models for feature extraction. The ML model, coupled with ViT, performs feature extraction using self-attention mechanisms to capture relevant patterns and structures within the fMRI scans. Subsequently, the extracted features are passed through a binary classifier for diagnosis.

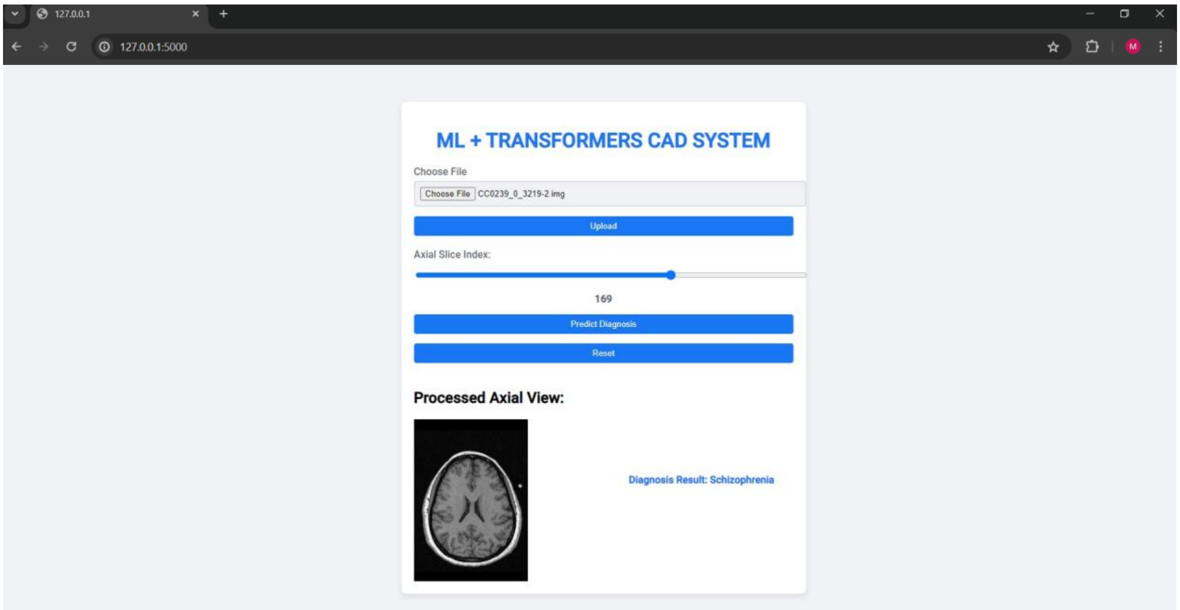

Fig. 4.10 UI of Website using ML and ViT with Schizophrenia diagnostic result

The processing time for generating diagnostic results remains efficient, typically ranging from 3 to 5 seconds. This rapid turnaround time ensures timely feedback for healthcare professionals, facilitating informed decision-making during clinical assessments. Additionally, the integration of ViT with ML enhances the system's diagnostic accuracy by enabling robust feature extraction from fMRI scans across a wide range of axial slices.

The deployment of ML and ViT models within a Flask-based website framework represents a significant advancement in psychiatric diagnosis technology. By overcoming the limitations of previous systems and offering enhanced flexibility and accuracy in feature extraction, this integrated method empowers healthcare professionals with valuable tools for improving patient care and outcomes.

4.8 Comparision with the state-of-the-art

Deep learning models have been extensively explored for diagnosing schizophrenia using MRI data, particularly fMRI datasets. This approach leverages the powerful feature extraction and classification capabilities of convolutional neural networks (CNNs), which have shown great promise in various medical imaging tasks. Specifically, studies by Manic et al. (2023) and Mudholkar, Siddant, and Agarwal (2024) have employed the VGG16 and VGG19 architectures. These studies reported notable accuracies of 91% for both models, indicating a high level of effectiveness in distinguishing between schizophrenia patients and healthy controls.

Similarly, Zheng et al. (2021) utilized the VGG16 model on a different fMRI dataset and reported a slightly lower accuracy of 84%. While this is somewhat lower than the results obtained by Manic et al. and Mudholkar, Siddant, and Agarwal, it still underscores the robustness of the VGG16 model in diagnosing schizophrenia.

These findings collectively underscore the effectiveness of VGG16 and VGG19 in discriminating between schizophrenia patients and healthy controls based on fMRI scans. Table 4.12 shows performance across studies, highlighting the robustness and reliability of these pre-trained CNN models for such diagnostic tasks. Among all these previous works, the feature fusion of ML and ViT method have showed the highest accuracy of 95%. As research in this area continues to advance, it is expected that the integration of deep learning models into clinical practice will enhance the accuracy and efficiency of schizophrenia diagnosis, ultimately improving patient outcomes. robustness of these pre-trained CNN models for such diagnostic tasks.

Table 4-12: Comparision of all state-of-the-art

Dataset	Modality	Methods	Performance	References
OpenNeuro / SCZ-20, HC – 20	fMRI	VGG16, VGG19, InceptionV3	Acc-91%	[22]
COBRE / SCZ-102, HC-98	fMRI	VGG16	Acc-84%	[9]
COBRE / SCZ-102, HC-98	fMRI	ML and VGG16	Acc-94.3%	[14]
OpenNeuro / 99	fMRI	VGG19	Acc-91%	[21]
NUSDAST/ SCZ-125, HC-125	fMRI	ML and ViT	Acc - 95%	**Proposed Work**

5. Conclusion

In conclusion, the presented study addresses the pressing challenges in schizophrenia diagnosis, aiming to enhance the efficiency and accuracy of detection through advanced machine learning and deep learning approaches. The motivation stems from the complexity of schizophrenia, a debilitating mental disorder with global ramifications.

The machine learning approach, employing feature extraction techniques such as GLCM, LBP, and HOG, coupled with Recursive Feature Elimination (RFE) for optimal feature selection, showcases promising results. The Histogram-based Gradient Boosting Classifier, integrated into the model architecture, achieves high accuracy, precision, recall, specificity of 89% and an impressive AUC-ROC Score of 0.96. This comprehensive approach provides valuable insights into the nuanced patterns within medical images, contributing to a well-performing model for schizophrenia classification.

The deep learning model, leveraging transfer learning with a pre-trained VGG19 architecture, augments its capabilities through data augmentation and meticulous hyperparameter tuning. The model demonstrates commendable accuracy of 91%, along with sensitivity of 88.5%, and specificity of 92.5% in detecting schizophrenia, highlighting its robustness in handling complex patterns within medical images. The fusion of machine learning and deep learning VGG19 approach have demonstrated an accuracy of 94% which showcases that the complementary features improve accuracy.

Furthermore, the vision transformer model combined with the machine learning model showed the highest accuracy, sensitivity and specificity of 95% ensuring the reliability on an automated system. The self-attention mechanism captured the contextual information among the brain regions while the CNN was relying on the spatial relationship between pixels.

Both methodologies contribute to the advancement of diagnostic tools for schizophrenia, addressing the limitations of traditional subjective evaluations. Future work may involve Integration of machine learning and deep learning techniques showcases the potential to revolutionize schizophrenia diagnosis, offering more efficient and accurate tools for clinicians. Further refining these models, exploring additional feature extraction techniques, and conducting rigorous clinical validations for seamless integration into real-world healthcare settings.

Bibliography

[1] Zoupa, Elli et al., “Cognitive Rehabilitation in Schizophrenia-Associated Cognitive Impairment: A Review,” *Neurology international*, vol. 15, no. 9, pp. 12-23, Dec. 2022.

[2] Ediri Arachchi et al., “A Systematic Characterization of Structural Brain Changes in Schizophrenia,” *Neurosci. Bull*, vol. 36, no. 10, pp. 1107–1122, Oct. 2020. Mohan N.M. and Kumar V.J., ‘A Novel Signal Conditioning Circuit for Push–pull-type Resistive Transducers’, *Measurement Science and Technology*, vol. 16, # 9, no. 9, pp. 1848–1852, 2005.

[3] A. Gautam and I. Chatterjee, “Medical Imaging and Schizophrenia: A Study on State-of-Art Applications,” *Cognizance of Schizophrenia: A Profound Insight into the Psyche*, pp. 271–281, Jan. 2023.

[4] Alves, C. L. et al., “Analysis of functional connectivity using machine learning and deep learning in different data modalities from individuals with schizophrenia,” *Journal of Neural Engineering*, vol. 20, no. 5, Oct. 2023.

[5] A. Venkataraman, T. J. Whitford, C. F. Westin, P. Golland, and M. Kubicki, “Whole brain resting state functional connectivity abnormalities in schizophrenia,” *Schizophrenia Research*, vol. 139, no. 1–3, pp. 7–12, Aug. 2012.

[6] Z. H. Chen et al., “Detecting Abnormal Brain Regions in Schizophrenia Using Structural MRI via Machine Learning,” *Computational Intelligence and Neuroscience*, vol. 19, pp. 10–16, 2020.

[7] J. M. Mateos-Pérez, M. Dadar, M. Lacalle-Aurioles, Y. Iturria-Medina, Y. Zeighami, and A. C. Evans, “Structural neuroimaging as clinical predictor: A review of machine learning applications,” *NeuroImage: Clinical*, vol. 20, pp. 506–522, 2018.

[8] R. De Filippis et al., “Machine learning techniques in a structural and functional MRI diagnostic approach in schizophrenia: A systematic review,” *Neuropsychiatric Disease and Treatment*, vol. 15, pp. 1605–1627, 2019.

[9] J. Zheng, X. Wei, J. Wang, H. Lin, H. Pan, and Y. Shi, “Diagnosis of Schizophrenia Based on Deep Learning Using fMRI,” *Computational and Mathematical Methods in Medicine*, vol. 12, pp. 21–27, 2021.

[10] J. Oh, B. L. Oh, K. U. Lee, J. H. Chae, and K. Yun, “Identifying Schizophrenia Using Structural MRI With a Deep Learning Algorithm,” *Frontiers in Psychiatry*, vol. 11, pp. 25–31, 2020.

[11] J. Zhang et al., “Detecting schizophrenia with 3D structural brain MRI using deep learning,” *Scientific Reports*, vol. 13, no. 1, 2023.

[12] M. Hu, K. Sim, J. H. Zhou, X. Jiang, and C. Guan, “Brain MRI-based 3D Convolutional Neural Networks for Classification of Schizophrenia and Controls*”,* in *42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC),* pp. 1742-1745, 2020.

[13] D. Liu, Y. Liu, S. Li, W. Li, and L. Wang, “Fusion of Handcrafted and Deep Features for Medical Image Classification,” *Journal of Physics: Conference Series*, vol. 1345, pp. 22-28, 2019.

[14] K. S. Manic, V. Rajinikanth, A. S. Al-Bimani, D. Taniar, and S. Kadry, “Framework to Detect Schizophrenia in Brain MRI Slices with Mayfly Algorithm-Selected Deep and Handcrafted Features,” *Sensors*, vol. 23, no. 1, pp 23-29, 2023.

[15] W. Chen et al., “A fusion of VGG-16 and ViT models for improving bone tumor classification in computed tomography,” *Journal of Bone Oncology*, vol. 43, pp. 22-28, Dec. 2023.

[16] A. Sagar, “ViTBIS: Vision Transformer for Biomedical Image Segmentation,” Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), *LNCS* vol. 12969, pp. 34 – 45, 2021.

[17] Q. Pu, Z. Xi, S. Yin, Z. Zhao, and L. Zhao, “Advantages of transformer and its application for medical image segmentation: a survey,” *BioMedical Engineering OnLine*, vol. 23, no. 1, Dec. 2024.

[18] A. Tyagi, V. P. Singh, and M. M. Gore, “Towards artificial intelligence in mental health: a comprehensive survey on the detection of schizophrenia,” *Multimedia Tools and Applications*, vol. 82, no. 13, pp. 20343–20405, 2023.

[19] L. Wang et al., “Northwestern University Schizophrenia Data and Software Tool (NUSDAST),” *Frontiers in Neuroinformatics*, vol. 7, pp. 55-61, Nov. 2013.

[20] Ashish Vaswani et al., “Attention is All you Need,” *Neural Information Processing Systems*, vol 10, pp 13-19, 2017.

[21] S. Kadry, D. Taniar, R. Damasevicius, and V. Rajinikanth, “Automated Detection of Schizophrenia from Brain MRI Slices using Optimized Deep-Features,” in *2021 IEEE 7th International Conference on Bio Signals, Images and Instrumentation*, vol 9., pp. 56-62, Mar. 2021.

[22] Y. Lyu, X. Yu, D. Zhu, and L. Zhang, “Classification of Alzheimer’s Disease via Vision Transformer: Classification of Alzheimer’s Disease via Vision Transformer,” in *ACM International Conference Proceeding Series*, pp. 463–468, Jun. 2022.

Publications

Authors: Mohan V, Nithin Rajaseharan, Taushiq Balamurugan, Varun S, Suguna G, Harikumar M E

Title: Validation of Vision Transformer Over Convolutional Neural Networks-based Pretrained Model for Schizophrenia Diagnosis

Conference: 4th Edition of IEEE Mysore Subsection Flagship conference

Organizer: IEEE Mysuru Subsection

Date of conference: 31-08-2024

Date of submission: 22-05-2024

Status: Submitted