ISSN: 0012-0413

Predicting Schizophrenia from Clinical Data with Feature Selection and CNN-Logistic Regression Classification

*[1]Uma Shankari Srinivasan, [2]Dr.J.Jaganpradeep,[3]Dr.Sripriya Arunachalam
[4]S.S.Rajasekar, [5]K.Saranya,
[1]Associate Professor, Department of Computer Science and Applications, SRM Institute of Science and Technology, Ramapuram Campus, Chennai, Tamilnadu, India.
Email: umabalajees@gmail.com
[2]Professor,Department of Electronics and Communication Engineering, SSM College of Engineering, Komarapalayam, Namakkal, Tamilnadu.jgnprdp@gmail.com
[3]Assistant Professor, Department of Computer Science and Applications
SRM Institute of Science and Technology,Faculty of Science and Humanities,Vadapalani City Campus, Chennai. priyaarun02@gmail.com
[4]Associate Professor, Department of Artificial intelligence and Machine Learning,
Bannari Amman Institute of technology, Erode, Tamil Nadu, India
ssrajasekar80@gmail.com
[5]Associate professor, Department of Computer Science and Engineering,
Bannari Amman institute of technology, Sathyamangalam, Tamilnadu,India
Email Id: ksaranyacse@gmail.com

Abstract:

The diagnosis of schizophrenia during early stages allows for better treatment results because of its complex debilitating nature. The treatment and diagnosis of schizophrenia early on can benefit greatly from machine learning model applications. The proposed data processing approach incorporates SMOTE (Synthetic Minority Over-sampling Technique) for addressing class imbalance together with Normalization for standardizing feature values. A hybrid model design made up of Lasso Regression for feature selection and CNN-Logistic Regression for classification helps process both imbalanced classes and features with different scales in the dataset. Lasso Regression serves as an algorithm to select important features from datasets which enhances both the interpretability and the predictive power of the model. The integrated system applies CNN components for extracting complex neuroimaging patterns from MRI data followed by Logistic Regression models to produce binary schizophrenia versus non-schizophrenia predictions. Space hierarchies of features are discovered efficiently by CNN operations allowing Logistic Regression to perform its linear classification duties. The two powerful analysis techniques of traditional machine learning combined with deep learning in CNN-Logistic Regression create a comprehensive and efficient approach to schizophrenia diagnosis. Utilizing the proposed hybrid approach produces superior classification performance compared to existing methods according to initial results thus providing a valuable tool for clinical applications in schizophrenia detection.

Keywords: Schizophrenia Prediction, Lasso Regression, Convolutional Neural Networks (CNN), Feature Selection, Hybrid Model, Clinical Data Classification.

1) Introduction

Schizophrenia stands as a persistent serious mental illness which disrupts the cognitive processes as well as emotional states and behavioral patterns of an individual. The illness results in hallucinations and delusions and disorganized thinking which heavily diminishes an individual's daily capabilities [1]. The medical community needs to detect schizophrenia as early during late teenage years and early 20s because timely intervention produces superior results. The diagnostic process for schizophrenia proves difficult to clinicians because of the complicated nature of the disease with its symptoms. Authentic detection of schizophrenia during initial stages enables professionals to stop disease progression so patients can achieve better results in their mental and daily functioning. The present diagnostic approach depends primarily on subjective clinical examinations together with patient interviews as well as symptom ratings that permit human error and bias [2]. Research has produced increasing interest in machine-learning models used for schizophrenia diagnoses because they offer objective prediction systems.

The Schizophrenia dataset contains clinical attributes such as Age, Gender and Marital_Status and symptom-related elements that include Fatigue, Slowing, Pain, Hygiene and Movement. The main hurdle in performing predictions using a dataset with imperfect quality stems from the combination of missing values and unbalanced classes and noise in the data [3]. Complicated non-linear data relationships in clinical datasets render predictive models ineffective when using basic analysis techniques to reach accurate outcomes.

---

[1]Iyortsuun, N. K., et al., A review of machine learning and deep learning approaches on mental health diagnosis
[2]Cortes-Briones, et al., Schizophrenia Research, 245, 122-140.
[3]Bae, Y. J., et al., Schizophrenia detection using machine learning approach from social media content. Sensors, 21(17), 5924.

The use of MRI scans for imaging purposes brings extra obstacles because differing scan quality and brain segment problems reduce predictive accuracy. The successful approach for reliable predictions requires integrations between data preprocessing with feature selection along with advanced machine learning models which handle such complexities [4].

Clinical diagnosis of schizophrenia depends on traditional methods that use clinical evaluations and psychiatric interviews together with psychological performance assessments. Logistic regression alongside decision trees represents statistical techniques found in machine learning that process clinical datasets but face limitations when dealing with complex datasets with numerous features.[5]Thresholding, Region Growing and Edge Detection stand among the techniques that medical practitioners use to segment MRI images. These analytical approaches struggle to automatically discover hierarchical features since manual human intervention becomes necessary to reach successful results when using them. The performance of traditional methods remains inadequate when processing extensive complex datasets at the same time they provide limited ability for precise automated decision systems.

Machine learning model performance enhancement relies on feature selection due to its importance in high-dimensional datasets including schizophrenia data sets [6]. Regularized regression method known as Lasso performs feature selection through its penalty mechanism that makes irrelevant features disappear by setting their coefficients to zero thus enabling the identification of key predictive variables. Principal Component Analysis (PCA) together with Random Forests prove effective at both feature reduction and important feature discovery for dimensionality reduction. The prediction of schizophrenic patients with Support Vector Machines (SVM), Logistic Regression, and Random Forest Classifier [7]has proven effective for medical diagnostics through clinical and symptom-related features examination.

---

[4]de Filippis, R., et al.,Machine learning techniques in a structural and functional MRI diagnostic approach in schizophrenia Neuropsychiatric disease and treatment.
[5]Verma, S., et al.,Journal of Ambient Intelligence and Humanized Computing, 14(5), 4795-4807.
[6]Veronese, E., et al.,Computational and mathematical methods in medicine, 2013(1), 867924.
[7]Hettige, N. C., et al., Classification of suicide attempters in schizophrenia using sociocultural and clinical features

The models work with pre-determined manual features while lacking automatic methods for identifying parallel interactions in data sources. Traditional machine learning algorithms demonstrate efficient performance in various scenarios yet show restricted functionality when operating with unstructured information and high-dimensional information that needs proper feature selection. The innovative approach rectifies multiple problems that traditional approaches present. Main defects in performance of machine learning models arise from imbalanced datasets and missing values which appear frequently in clinical data. The application of Lasso Regression during feature selection maintains high relevance among included features which reduces model noise and improves interpretability as explained in [8].
Traditional models restrict their ability to detect complex non-linear patterns that exist between features within clinical datasets. The model performs automatic high-level feature extraction through Convolutional Neural Networks (CNN) which can handle complex or dimensional clinical data. CNNs demonstrate success in image processing because they efficiently process features with spatial hierarchies which traditional models cannot address. The CNN-LogRegNet hybrid model uses components from deep learning and classic machine learning to achieve effective classification across complex datasets with a better data processing efficiency than standard techniques do. The necessity behind the proposed study emerges because schizophrenic datasets have become more complex and existing diagnostic procedures have reached their capacity limits. Clinical data lacks enough predictive power for schizophrenia diagnosis while neuroimaging and alternative complex inputs need specialized processing methods. The CNN-LogRegNet model takes care of both feature selection and dimensionality reduction while using a deep learning framework [9]which learns valuable data features automatically from raw data sources to become an optimal tool for processing structured and unstructured datasets. The proposed model delivers superior predictive capabilities alongside optimal controlled status analysis of datasets which feature high-dimensional and noisy characteristics alongside data imbalance. Using feature selection methods together with recent classification approaches the research enables improved and scalable schizophrenia diagnostic methods [10].

---

[8]Schnack, H. G. (2019). Improving individual predictions: machine learning approaches for detecting and attacking heterogeneity in schizophrenia.
[9]Frick, J., et al.,A machine learning algorithm for potential early detection and prevention based on event-related potentials.
[10]Li, J., et al., Brain imaging and behaviour, 13, 1386-1396.

Main contribution of the proposed work

- The proposed method combines deep learning with logistic regression through a hybrid CNN-LogRegNet model aimed at schizophrenia prediction.
- Application of Handling Imbalanced Data techniques (e.g., SMOTE) to address class imbalance in the dataset.
- Standardization through Normalization ensures features maintain equivalent scales which produces balanced model performance.
- Feature selection using Lasso Regression helped minimize model dimensions while making the results more easily interpretable.
- Enhanced model performance through the combination of automatic feature extraction via CNN and classification via logistic regression.
- Improved prediction accuracy and robustness over traditional methods for early schizophrenia detection.

According to the organization of this document the following sections will be presented. The literature review section of this paper explores different schizophrenia detection methods alongside their clinical and neuroimaging data analysis progressions in Section II. The proposed approach of Section III starts with handling unbalanced data normalization while implementing Lasso Regression for feature selection before using the CNN-LogRegNet hybrid model. The processed information goes through the CNN-LogRegNet hybrid model that applies Convolutional Neural Networks (CNN) to extract automatic features alongside Logistic Regression for final classification. This section explains the application of the proposed method on schizophrenia data and demonstrates how it measures up against standard approaches regarding accuracy while evaluating both precision and recall and computational speed. The paper finishes with section V which provides concluding remarks and future research perspectives.

2) Literature Survey

The field of schizophrenia detection through neuroimaging techniques combined with machine learning methods possesses significant growth since it developed various improvements to diagnostic accuracy solutions. Multiple analyses investigate schizophrenia detection by using structural MRI alongside functional MRI signals and EEG measurements and supporting these findings with support vector machines as well as deep learning and random forests. Researchers implement these approaches mainly to resolve data imbalances while extracting features and uniting multiple data types. The promising outcomes from CNNs in deep learning models exist despite ongoing concerns regarding requirements for substantial dataset volumes and computational requirements and model interpretability limitations. The research examines various diagnostic techniques used for schizophrenia detection thus providing details about their advantages as well as constraints. The paper seeks to showcase recent progress while offering new directions to enhance early diagnosis capabilities.

According to Oh et al. (2020), deep learning techniques yield a new method for schizophrenia diagnosis through analysis of structural MRI data [12]. The research project employed a CNN which analyzed MRI scan data to extract useful features for schizophrenia patient identification. The deep learning algorithm achieved dependable results in separating schizophrenia patients from healthy control participants thus proving its worth in medical image processing. The system's automatic feature extraction capability is an advantage because it substitutes manual feature engineering steps. The main drawback of this approach stems from needing substantial quantities of labeled MRI data but such resources might be scarce. The black-box nature of deep learning techniques creates difficulty for clinicians to understand how these models reach their clinical decisions. The researchers at Alves et al. (2023) investigated how machine learning and deep learning may analyze functional connectivity in people with schizophrenia. Multi-modal functional MRI (fMRI) and EEG systems were combined in the study to enhance accuracy levels [13]. The proposed approach used graph-based analysis in combination with autoencoders and convolutional networks to find connectivity patterns which indicate schizophrenia. Combining multiple data sources through this method provides a broader understanding of the disorder because of its effective data consolidation ability. The implementation faces difficulties because it requires advanced computing power and access to premium-quality multimodal medical data that medical facilities may lack.

[12]Oh, J., et al., Identifying schizophrenia using structural MRI with a deep learning algorithm.
[13]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, 20(5), 056025.

Sharaev et al. (2022) conducted research to study schizophrenia diagnosis through implementation of biomarkers combined with machine learning approaches as reported in [14]. Neuroimaging together with genetic information and clinical findings enabled the development of predictive models which classified schizophrenia cases. Random forests with gradient boosting ensemble techniques were utilized to boost the confidence of predictions during the methodology stage. The approach enables better understanding of results when compared to deep learning frameworks thus making it a more appropriate solution for clinical needs. The main weakness arises from the need for diverse data types since different collection methods could produce inconsistent results.

This study examined how machine learning techniques classify patients between schizophrenia, autism, ultra-high risk and first-episode psychosis populations through neuroimaging data analysis (Yassin et al. 2020). SVM and random forests and deep learning models served as the evaluation techniques for categorizing subjects in the research. Research proved that deep learning models specifically CNNs demonstrated high performance when distinguishing psychiatric disorders [15]. The main advantage of this approach involves its ability to extract intricate data patterns from neuroimaging measurements which enhances diagnostic precision. The use of deep learning models faces two significant hurdles such as high processing requirements and scarce neuroimaging datasets with appropriate labels for training purposes.

Schizophrenia diagnosis received improved classification by applying a hybrid framework that united functional Magnetic Resonance Imaging data with genetic information according to Yang et al. (2010). The implemented methodology used selected features alongside ensemble learning approaches to build a reinforced classification system. The study showed that predictive abilities improve when researchers bring together fMRI and genetic markers for analysis because the simultaneous use of multiple data types outperforms single data source examinations [16].

---

[14]Sharaev, M. G., et al., Diagnosis of schizophrenia based on the data of various modalities: biomarkers and machine learning techniques.
[15]Yassin, W., et al.,Translational psychiatry, 10(1), 278.
[16]Yang, H., et al., combining both improves classification of schizophrenia. Frontiers in human neuroscience, 4, 192.

The research by Hassan et al. (2023) investigated schizophrenia diagnosis through CNN applications combined with traditional machine learning methods using multivariate EEG signals. The research work presented a new method for combining features [17] through deep learning models that extracted both spatial and temporally-related EEG patterns.

The authors at Nguyen et al. (2022) created a decision support system which distinguished schizophrenia from mood disorders through analyzing wearable device data using deep learning models. Multiple CNN-based architectures analyzed physiological signals using heart rate variability (HRV) and skin conductance data acquired from wearable sensors according to the study [18]. This system proves effective for real-time operation when used in both clinic settings and during home care.

Ko and Yang (2022) developed an EEG-based schizophrenia diagnosis system which applied time-series image conversion together with deep learning methods. A process to transform EEG signals into images allowed CNNs to perform the classification tasks. The main benefit of CNNs stems from their ability to extract improved features which enhances classification performance. The conversion process of EEG signals into images leads to some information loss regarding fine temporal information which impacts interpretability capabilities of the model [19].

A deep learning model based on RNN-LSTM (Recurrent Neural Networks - Long Short-Term Memory) investigates schizophrenia diagnosis from EEG data according to Supakar et al. (2022) [20]. The researchers designed their work to discover wayward time connections between EEG traces since these elements are vital for detecting schizophrenia-related irregularities. The training instability and high computational costs associated with RNN-based models create deployment barriers for real-time clinical applications because they make implementation challenging.

---

[17] Hassan, F., et al.,Information Fusion, 92, 466-478.
[18] Nguyen, D. K., et al., Health Informatics Journal, 28(4), 14604582221137537.
[19] Ko, D. W., et al., EEG-Based schizophrenia diagnosis through time series image conversion and deep learning.
[20] Supakar, R., et al.,A deep learning-based model using RNN-LSTM for the Detection of Schizophrenia from EEG data. Computers in Biology and Medicine.

ISSN: 0012-0413

Table 1: Literature review on existing methods

S.No	Author(s) et al. (Year)	Dataset	Methodology	Accuracy (%)	Challenges
1	Yassin et al. (2020)	Neuroimaging data	Machine learning models applied to neuroimaging data for classification	87%	Limited availability of labeled neuroimaging data
2	Yang et al. (2010)	fMRI, Genetic	Hybrid machine learning model combining fMRI and genetic data	86%	Privacy concerns, data collection limitations
3	Supakar et al. (2022)	EEG	RNN-LSTM for sequential EEG data classification	85%	Computational costs, training instability
4	Sharaev et al. (2022)	Multi-modal (neuroimaging, genetic, clinical)	Ensemble learning with random forests and gradient boosting for multi-modal data	88%	Data collection inconsistencies, heterogeneity
5	Sharaev et al. (2022)	Multi-modal (neuroimaging, genetic)	Multi-modal data analysis using biomarkers and machine learning	85%	Heterogeneous data sources, inconsistencies

[20]Yassin, W., et al., Machine-learning classification using neuroimaging data in schizophrenia, autism, ultra-high risk and first-episode psychosis.
[21]Yang, H., et al., A hybrid machine learning method for fusing fMRI and genetic data: combining both improves classification of schizophrenia.
[25]Supakar, R., et al., A deep learning-based model using RNN-LSTM for the Detection of Schizophrenia from EEG data. Computers in Biology and Medicine, 151, 106225.
[19]Sharaev, M. G., et al., Diagnosis of schizophrenia based on the data of various modalities: biomarkers and machine learning techniques.
[19]Sharaev, M. G., et al., Diagnosis of schizophrenia based on the data of various modalities: biomarkers and machine learning techniques,14(5 (eng)), 54-75.

ISSN: 0012-0413

6	Oh et al. (2020)	Structural MRI	CNN for structural MRI classification of schizophrenia	90%	Requires large labeled MRI datasets, interpretability issues
7	Nguyen et al. (2022)	Wearable device data	Multiple CNN-based deep learning models on wearable data	86%	Variability in wearable sensor accuracy
8	Ko and Yang (2022)	EEG	Time-series EEG image conversion and CNN for classification	85%	Loss of temporal information in EEG conversion
9	Hassan et al. (2023)	EEG Signals	Fusion of EEG signals using CNN and machine learning models	88%	EEG signal noise, feature fusion complexity
10	Alves et al. (2023)	fMRI, EEG	Analysis of functional connectivity using deep learning models	85%	High computational cost, data quality issues

[12]Oh, J., Oh, B. L., Lee, K. U., Chae, J. H., & Yun, K. (2020). Identifying schizophrenia using structural MRI with a deep learning algorithm.
[18]Nguyen, D. K., et al., Decision support system for the differentiation of schizophrenia and mood disorders using multiple deep learning models on wearable devices data.
[19]Ko, D. W., et al., EEG-Based schizophrenia diagnosis through time series image conversion and deep learning.
[17]Hassan, F., et al.,Fusion of multivariate EEG signals for schizophrenia detection using CNN and machine learning techniques. Information Fusion, 92, 466-478.
[13]Alves, C. L., et al., Analysis of functional connectivity using machine learning and deep learning in different data modalities from individuals with schizophrenia.

The principal limitation across existing research is data imbalance along with complex feature relationships coupled with limited availability of large labeled datasets that reduce model accuracy and generality. Current procedures face difficulties in both computational cost and they become impractical when used clinically because of the challenges handling model interpretation. The proposed solution utilizes SMOTE as well as Normalization methods to normalize imbalanced datasets for providing appropriate input data to the model. Both Lasso Regression and the CNN-LogRegNet hybrid model combine to optimize feature selection along with model interpretation while achieving superior accuracy and computational efficiency from data analysis patterns. An integrated information processing system provides better early schizophrenia detection by delivering accurate scalable interpretable models beyond traditional limitations.

3) Proposed Work

The CNN-LogRegNet model unites Convolutional Neural Networks (CNNs) feature extraction capabilities and Logistic Regression (LogReg) classification strength to analyze clinical and neuroimaging data for predicting schizophrenia cases. CNNs use multiple convolutional layers followed by pooling to automatically identify important characteristics in their input data including brain scans by building spatial knowledge structures. After convolutional neural network extraction of features the logistical regression system performs an analysis to determine the likelihood of schizophrenia. A training process based on backpropagation helps optimize cross-entropy loss through gradient descent to train the model. The model utilizes its architecture effectively to detect complex data patterns and provides interpretability which makes it a valuable tool for efficient schizophrenia diagnosis.

ISSN: 0012-0413

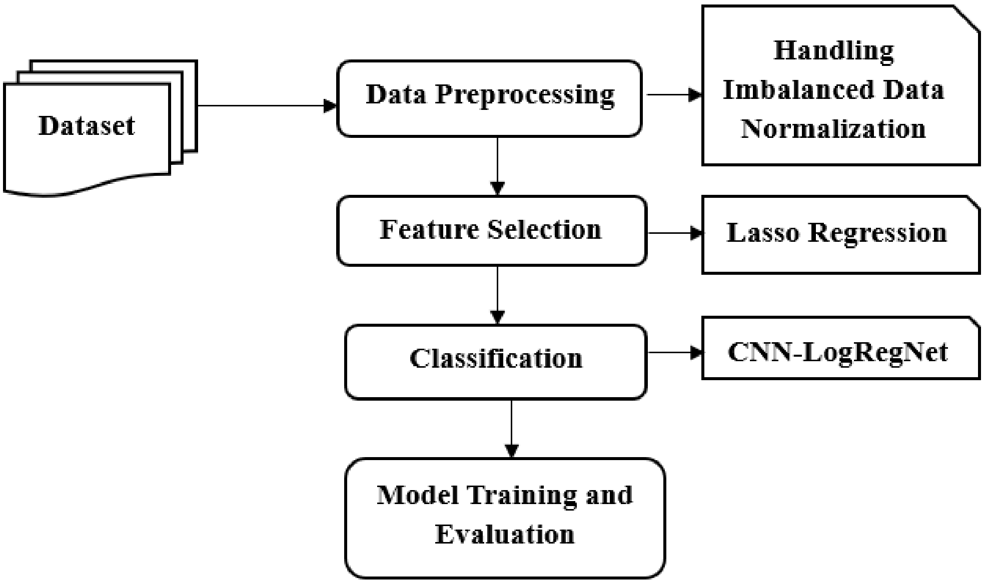

Figure 1: Proposed Work

3.1) Data Preprocessing: Handling Imbalanced Data and Normalization

3.1.1) Handling Imbalanced Data

The term imbalanced data indicates when one class contains too many instances in comparison to the other class instances. The proportion of schizophrenia patients among the population typically remains much lower than those without the condition in schizophrenia prediction scenarios. The disproportionate class distribution produces models with majority class preferences that impact bad generalization results for minority class data.Among the most proven strategies to address data imbalance is Synthetic Minority Over-sampling Technique (SMOTE). SMOTE generates new artificial examples specifically for the minority class instead of performing only majority class sampling reduction. The technique operates as follows:

SMOTE creates synthetic samples by performing interpolation between a sample from the minority class and its nearest neighbors. A specific mathematical structure enables this interpolation process based on the following equation:

$$New\ Sample = X_{sample} + \lambda . \left(X_{neighbor} - X_{sample}\right) \quad (1)$$

Where:

- X_{sample} is a sample from the minority class,
- $X_{neighbor}$ is one of its k-nearest neighbors,
- λ is a random number between 0 and 1, controlling the interpolation.

Sample ratio balancing through minority class enrichment improves model performance when identifying subjects with schizophrenia as well as other minority individuals.

3.1.2) Normalization

The normalization technique scales feature values into a standard numeric range to prevent any single measurement from controlling the model through differences in scale. Clinical data presents Age, Fatigue and Movement features with various measurement units that span from age values between 0 to 100 years and movement score ranges from 0 to 10. The feature with the bigger numerical span could receive problematic weight rendering the model ineffective when normalization is omitted.

When performing normalization the Min-Max Scaling technique adjusts all feature values into a standard numeric range from 0 to 1. The mathematical formula for Min-Max normalization appears as shown below:

$$X_{norm} = \frac{X_{original} - X_{min}}{X_{max} - X_{min}} \qquad (2)$$

Where:

- $X_{original}$ is the original value of a feature,
- X_{min} and X_{max} are the minimum and maximum values of that feature, respectively.

By scaling all features to a common range, normalization ensures that no single feature dominates due to its large numerical scale, leading to a more balanced and fair model training process.

3.1.3) Handling Data Imbalance and Normalization Combined:

A joint use of imbalanced data treatment and normalization methods allows the model to learn from well-prepared data which enables it to recognize minority class patterns alongside maintaining vital information from all involved features. SMOTE enables appropriate minority class representation in order to achieve accurate results through normalization that prevents any single feature from dominating the learning process because of its scale.

When implementing SMOTE for dataset balancing the model generates artificial data points through interpolation which becomes new examples for patients with schizophrenia. The newly generated synthetic samples achieve normalized feature values so both age and movement scores which had wide value ranges become equal for model computations. The combination of SMOTE-based balancing with normalization-based scaling creates a stronger model which successfully predicts schizophrenia cases despite feature relationships of any complexity.

A crucial challenge during preprocessing exists in preventing the model from fitting additionally to synthetic data which SMOTE produces. The model develops memorization of training data that includes synthesized examples along with unlearnable generalizable patterns during this process. The implementation of repeated cross-validation splits the data between training and validation sections multiple times to achieve better model generality for new observations.

The process of synthesizing new samples through SMOTE requires additional computation and specialists must take caution when normalizing data to prevent relationship distortions from occurring. When properly adjusting SMOTE parameters such as k for neighbors together with selecting proper normalization methods including Z-score standardization helps reduce these implementation challenges.

The combination of Handling Imbalanced Data (SMOTE) and Normalization produces optimal conditions for the schizophrenia detection model since both methods ensure robust representation of classes while avoiding domination from individual variables during predictions in real-world clinical settings.

3.2) Feature Selection: Lasso Regression

The Least Absolute Shrinkage and Selection Operator regression (Lasso) serves as a linear modeling technique that executes feature selection whereas it applies regularization methods to reach improved prediction accuracy and modeling interpretability. Standard linear regression models minimize the sum of squared residuals to estimate values but does not include automatic feature selection capabilities. Lasso sets an additional penalized term to its objective function which will force insignificant feature coefficients to become zero to exclude them from the model. Lasso provides strong capabilities for working with high-dimensional datasets because it effectively eliminates unimportant features while retaining relevant ones. Such capabilities benefit schizophrenia prediction studies which have large datasets with multiple possible inputs. Lasso Regression minimizes this objective function to achieve its defined goal:

$$L(\beta) = \sum_{i=1}^{n} \left(y_i - \sum_{j=1}^{p} X_{ij}\,\beta_j \right)^2 + \lambda \sum_{j=1}^{p} |\beta_j| \qquad (3)$$

Where:

- $y_i$is the observed value for the ithi^{th}ith data point,
- X_{ij} is the value of the j^th^ feature for the i^th^ data point,
- β_j is the coefficient for the j^th^ feature,
- p is the total number of features,
- λ is the regularization parameter controlling the strength of the penalty, and
- $|\beta_j|$ is the absolute value of the coefficient for the j^th^ feature.

The residual sum of squares (RSS) serves as the first term where it calculates prediction errors between expected outcomes and observed results. The second term, $\lambda \sum_{j=1}^{p} |\beta_j|$, is the L1 regularization includes the term that produces penalties against coefficient absolute values. Lasso achieves sparsity through its penalty term because it enforces the elimination of some coefficients to zero value. Feature selection becomes possible through this technique because coefficients which equal zero are automatically eliminated from the predictive model. The regularization parameter λ plays a crucial role in the behaviour of Lasso Regression. Linear regression results as an output of Lasso when the function parameter λ equals zero because it lacks coefficient penalties. The model forces coefficient shrinking toward zero while λ increases because the coefficient penalization strength intensifies. A large λ value in Lasso Regression automatically drives numerous coefficients toward zero and thus eliminates those features from model usage. The mathematical solution for solving Lasso problems takes the form:

$$\beta^{\wedge}{}_{j} = \operatorname{argmin}\beta_j \left(\sum_{i=1}^{n} \left(y_i - \sum_{j=1}^{p} X_{ij}\, \beta_j \right)^2 \right) + \lambda \sum_{j=1}^{p} |\beta_j| \qquad (4)$$

The estimated coefficients can be found through β^j. Higher values of λ result in the setting of additional coefficient values to zero which produces reduced and easier to comprehend models. To use Lasso successfully one needs to select the right value of λ which can be determined through cross-validation. Lasso functions well as a feature selection tool and regularization method yet faces several important constraints. The selection process of Lasso allows just one feature choice from correlated groups which might eliminate potentially important information. Avoiding this situation becomes complex since multiple related features work together to produce accurate predictions. The close correlation between Fatigue and Pain variables could lead Lasso to eliminate one of them despite their collective value in schizophrenia identification. Elastic Net serves as an alternative to this problem through its combination of Lasso and Ridge Regression's strengths that enables feature selection between correlated variables.

3.3) CNN-LogRegNet for Classification

The CNN-LogRegNet integrates both Convolutional Neural Networks (CNNs) for extracting features from data and Logistic Regression (LogReg) for performing classification. The integrated model has been developed to solve the prediction problem that involves analyzing complex clinical information together with neuroimaging data for schizophrenia diagnosis. The CNN architecture performs ideally for extracting features automatically from unprocessed data sequences which include images and multidimensional elements. Logistic Regression functions as a simple classification method which delivers probabilistic outputs alongside good interpretability. The model benefits from CNNs' feature extraction strength by using these algorithms together with logistic regression classification power.

3.4) Convolutional Neural Networks (CNNs) for Feature Extraction:

A deep learning type known as CNN demonstrates exceptional ability for data spatial hierarchy extraction particularly when processing images. A CNN includes a collection of convolutional followed by pooling layers throughout its structure. The convolution operation defines as:

$$Z_{ij} = (X * W)_{ij} = \sum_{m=1}^{M} \sum_{n=1}^{N} X_i + m_{j+n} + W_{m,n} \qquad (5)$$

Where:

- X is the input image or feature map,
- W is the convolution kernel or filter,
- (i, j) represents the position of the filter in the image,
- Z_{ij} is the resulting feature map after applying the filter.

The convolution operation teaches the CNN to identify spatial image patterns such as edges together with textures and shapes for later pattern recognition stages. The model proceeds to execute a ReLU activation process after convolution takes place to establish nonlinearity:

$$ReLU(x) = \max(0, x) \qquad (6)$$

The non-linear transformation enables the model to identify complex data features together with their relationships. Several stacked convolutional layers extract features at different levels from lower edges to higher object components.

Feature map dimensions become smaller through pooling operations after each convolutional layer because this technique enables computational complexity reduction and overfitting control. Max-pooling stands as the most employed pooling technique since it chooses the maximum value within each window contained in the feature map. Mathematical formulation of max-pooling appears as:

$$P_{ij} = \max\left(X_i + m_{j+n}\right) \qquad (7)$$

The pooled feature mapP_{ij}exists at the location (i,j) in the pooling framework. Through pooling layers, the model develops translation-invariance which enables it to identify features no matter where they appear in an input image. The ability to detect schizophrenia patterns in brain scans during image classification tasks depends on this capability.

3.5) Fully Connected Layers and Logistic Regression:

CNNs usually proceed with one or multiple fully connected layers also known as dense layers after completing the convolutional and pooling steps. The layers integrate every neuron from one level to the successive tier thereby enabling the model to discover intricate linkages within features procured from original data. The CNN output passes through logistic regression classifier after flattening into a vector format.

The purpose of logistic regression involves determining optimal βj weights across features to achieve highest likelihood of sound data classification results. The model determines class probability through application of the sigmoid function:

$$P(y = 1|X) = \frac{1}{1 + e^{-(\beta_0 + \sum_{j=1}^{p} \beta_j X_j)}} \qquad (8)$$

Where:

- P(y=1|X)P(y=1|X)P(y=1|X) is the probability of class 1 (schizophrenia),
- β0\beta_0β0 is the bias term,
- βj\beta_jβj is the coefficient for the jthj^{th}jth feature,
- XjX_jXj is the value of the jthj^{th}jth feature.

An output probability between 0 and 1 is produced by the sigmoid function before thresholding at 0.5 decides the classification as class 0 (no schizophrenia) or class 1 (schizophrenia).

3.6) Training the CNN-LogRegNet Model

The CNN-LogRegNet training process comprises two essential aspects which include extracting features through the CNN and identifying those features with the logistic regression model. The training process employs backpropagation to optimize the weights in the CNN by minimizing cross-entropy loss or a similar loss function. The definition of cross-entropy loss function appears as:

$$L(\beta) = -\sum_{i=0}^{n}(y_i \log(P(y_i = 1|X_i))(1 - y_i)\log(1 - P(y_i = 1|X_i))) \quad (9)$$

Where:

- y_i is the actual label (0 or 1),
- P$(y_i = 1|X_i)$is the predicted probability for class 1,
- n is the number of training samples.

The loss function penalizes wrong predictions through a system which increases the cost as prediction errors get larger compared to actual probabilities. The optimization seeks minimization of loss through gradient descent algorithms and their variants including Adam which alters CNN-weighted structures and logistic regression parameters.

Algorithm for the proposed work

```
# Initialize CNN-LogRegNet model
  model = Sequential()
```

```
# Add CNN layers for feature extraction
    model.add(Conv2D(filters=32, kernel_size=(3,3), activation='relu',
    input_shape=input_shape))
    model.add(MaxPooling2D(pool_size=(2, 2)))
    model.add(Conv2D(filters=64, kernel_size=(3,3), activation='relu'))
    model.add(MaxPooling2D(pool_size=(2, 2)))
    model.add(Flatten())  # Flatten the CNN output for input to Logistic Regression
# Add Logistic Regression layer
model.add(Dense(1, activation='sigmoid'))  # For binary classification
# Compile the model
model.compile(optimizer='adam', loss='binary_crossentropy', metrics=['accuracy'])
# Train the model
history = model.fit(X_train, y_train, epochs=10, batch_size=32,
validation_data=(X_test, y_test))
# Evaluate model performance
test_loss, test_accuracy = model.evaluate(X_test, y_test)
# Predict on new data
y_pred = model.predict(X_new_data)
```

The combined CNN-LogRegNet system provides multiple advantages when used for diagnosing schizophrenia. Through its automated feature extraction capabilities the CNN component teaches the model which raw data features better represent the information without requiring any manual input. Logistic Regression as the final classification step provides the advantage of having an interpretable and simple process that clinical professionals need to understand their decisions. The combination of these methods works best for medical image classification because it enables the detection of intricate patterns such as schizophrenia in brain scan analysis. The combination of CNN with logistic regression results in better computational efficiency because CNN first decreases the input data dimensions which Logistic Regression ends up using. The digitized model which combines deep learning feature extraction and classical machine learning classification becomes a powerful instrument to detect schizophrenia and complex disorders from clinical information.

ISSN: 0012-0413

4) Result and Discussion

The Schizophrenia Symptoms Dataset provides comprehensive reports about 5,000 patients who exhibit different schizophrenia-related behavioral and speech symptoms. The dataset consists of Name along with Age, Gender, Marital Status and specific symptom values for Fatigue, Slowing, Pain, Hygiene and Movement. Through these database features researchers gain valuable access to physical and psychological manifestations of schizophrenia needed for detecting behavioral and verbal patterns. The categorized dataset contains equal male and female participants who span different marital circumstances and shows diagnosis status for each subject through Schizophrenia labels. The expansive database offers essential knowledge to scientists who develop prediction models and study symptoms which appear in multiple population groups.

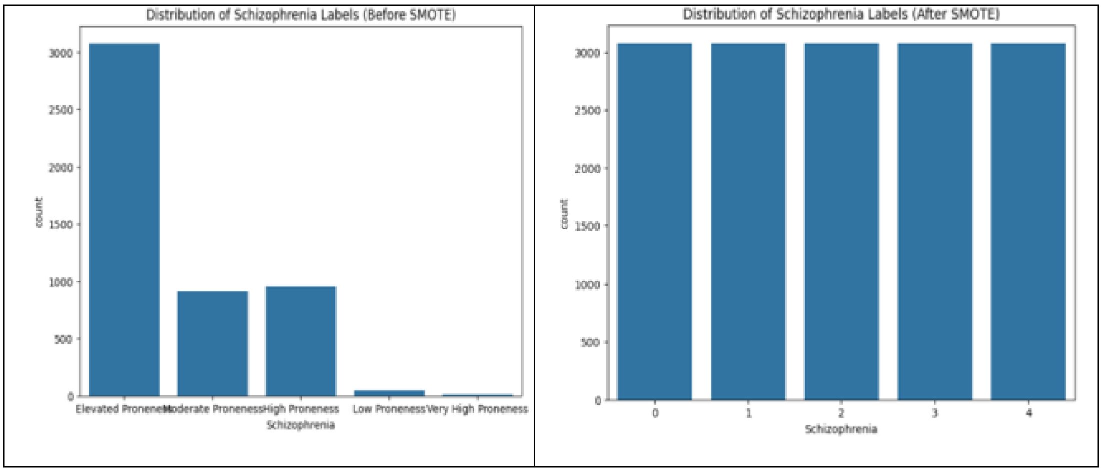

Figure 2: Distribution of Schizophrenia labels before and after applying SMOTE

The figure 2 demonstrates how Schizophrenia labels appear before and after implementing the SMOTE (Synthetic Minority Over-sampling Technique) process. The initial schizophrenia distribution reveals that most samples belong to the "Elevated Proneness" category whereas "Low Proneness" and "Very High Proneness" categories contain much lower samples. The dataset achieved balanced distribution through SMOTE since every schizophrenia label between 0 to 4 now contains similar sample numbers for optimal model training. The model achieves improved performance across all classes because of this data manipulation technique.

 ISSN: 0012-0413

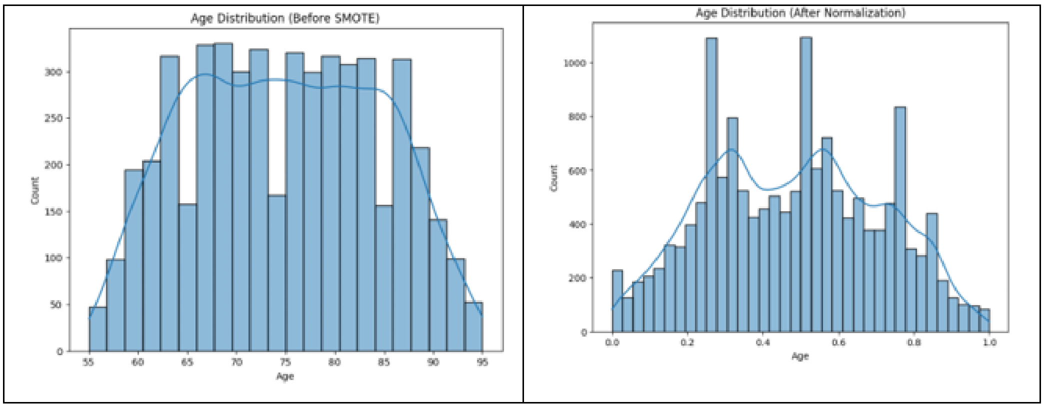

Figure 3: Distribution of Age before and after data preprocessing

Figure 3 demonstrates how age data changes between its pre-processement state and post-processement state. Before normalization the age values in the dataset spread widely within various age bands particularly at 60-70 years and 80-90 years according to the left panel. After Normalization the age values receive transformation which spans between 0 and 1 thus enhancing model learning effectiveness for age characteristics. The normalized distribution establishes uniformity which allows the model to work on age values without permit of unequal scales between different ranges. Model performance improvement depends on this transformation since it creates equalized influences between features.

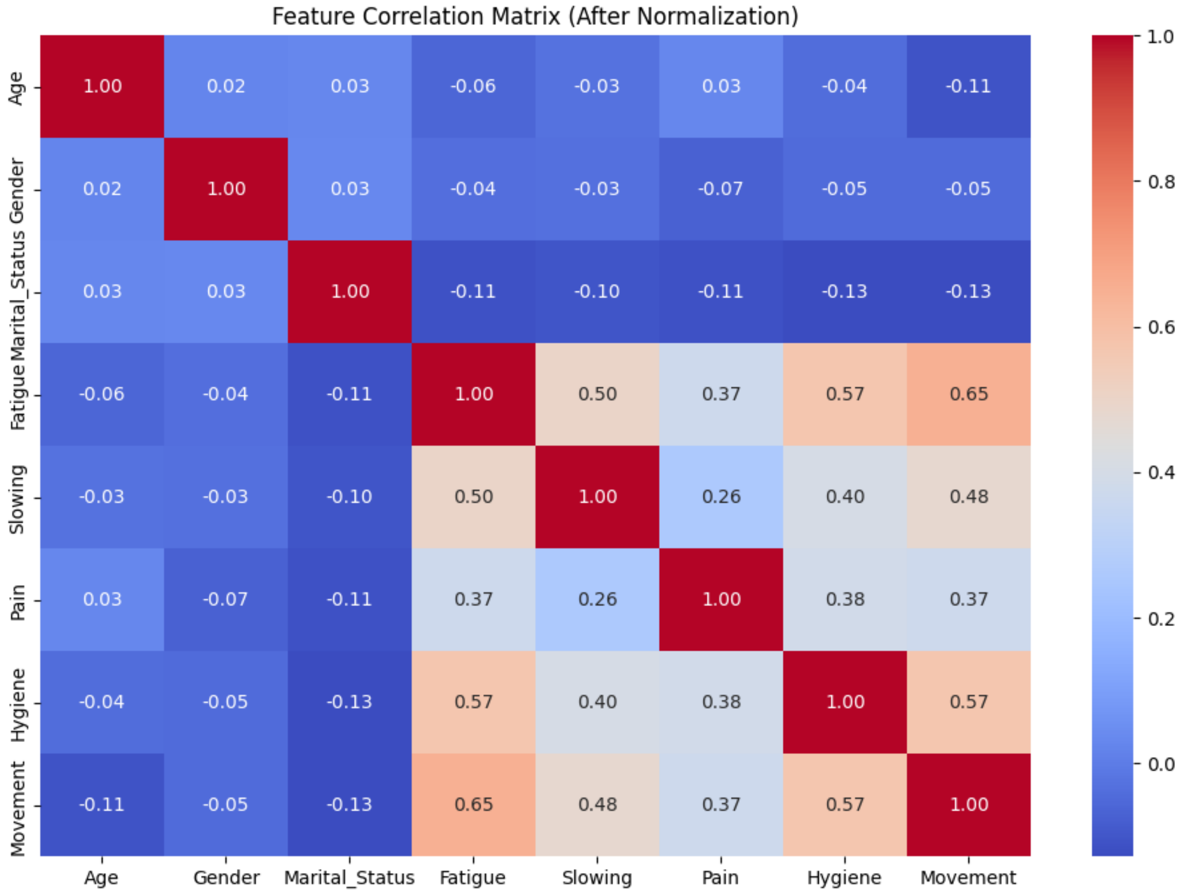

Figure 4: Feature correlation matrix of the dataset after normalization

After normalization the figure 4 depicts the feature correlation matrix featuring pairwise correlation coefficients that connect different features. The correlation values span between -1 to 1 with 1 indicating a powerful positive link and -1 bringing opposite correlation strength. Fatigue demonstrates both high positive relationships with Movement (0.65) and Hygiene (0.57). Data reveals a moderate strength of relationship between Fatigue (0.50) and Movement (0.48) when compared to Slowing. The relationship between Age, Gender, Marital Status and other features proves to be quite weak since these factors display independence between them. Feature relationship understanding becomes possible through the correlation matrix because it helps both model interpretation and feature selection processes.

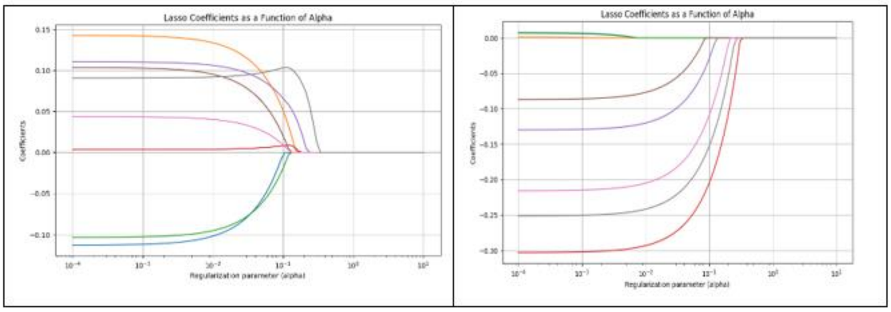

Figure 5: Relationship between the Lasso coefficients and the regularization parameter (alpha)

The Lasso coefficients appear in Figure 5 through two plot subpanels which display them against different values of the regularization parameter (alpha). The left side of the display visualizes feature coefficients as alpha value increases. The first part of the alpha value range produces constant coefficients until the parameter expands to cause coefficient reduction towards zero. This behavior demonstrates the regularization impact of Lasso regression. The right subplot shows this same pattern at closer scale as alpha values increase which causes coefficients to become more sparse. A high alpha value enables Lasso to remove more features from the model through the feature selection process. The visualizations clearly show how Lasso applies regularization causing unimportant model coefficients to disappear while alpha values rise.

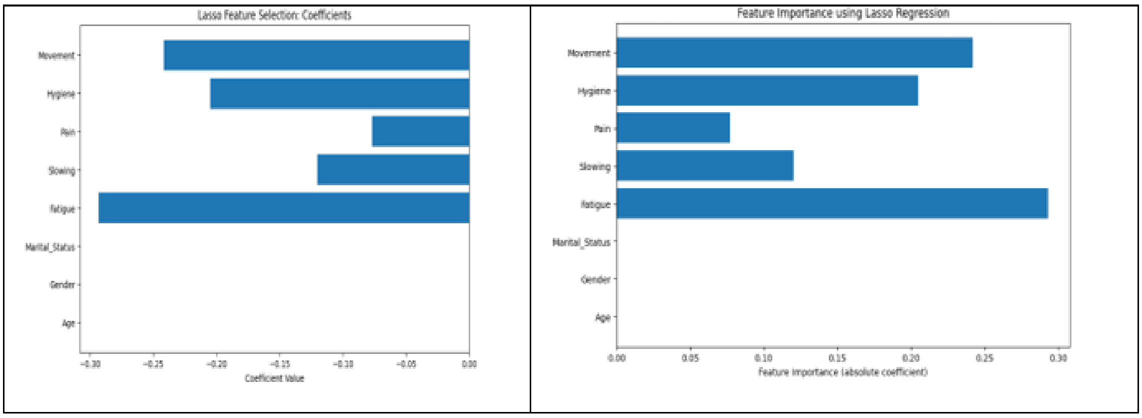

Figure 6: visualizations related to Lasso Regression

Lasso Regression displays its visual representations in Figure 6 where the images depict different aspects of this technique. The coefficients of features under Lasso regularization appear in the Lasso Feature Selection: Coefficients plot on the left side. The model shows

Movement together with Hygiene and Fatigue demonstrate the greatest importance towards schizophrenia prediction. The model indicates that the values of Gender Age and Marital Status are insignificant since their coefficients approach zero. The right side display shows Feature Importance using Lasso Regression which demonstrates the absolute coefficient values to explain the relative feature significance. The size of the coefficient determines how essential a feature becomes for generating model predictions.

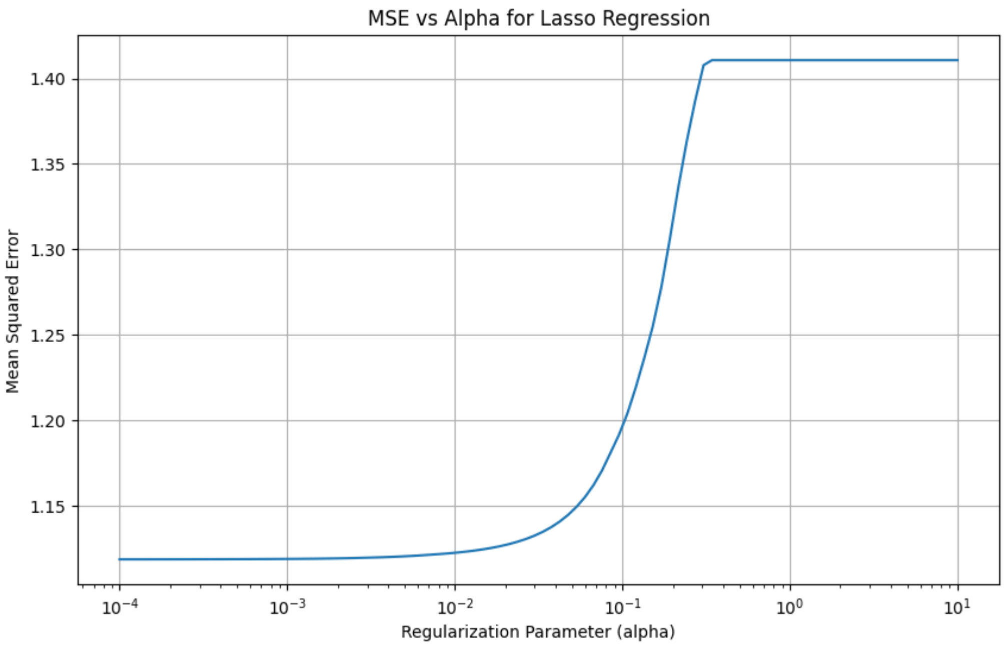

Figure 7: MSE Vs Alpha for Lasso Regression

The figure 7 displays the Mean Squared Error (MSE) as a function of the regularization parameter (alpha) for Lasso Regression. During this stage of model fitting the MSE maintains a minimal value and the alpha parameter remains near zero because the data points continue being accurately predicted. An increase in the alpha value leads to a rapid growth of MSE indicating that the penalty effect now actively reduces model complexity. The MSE increases to a high value which signifies data underfitting during circumstances of excessive regularization strength. The presented graph demonstrates why adjusting the alpha control parameter leads to optimal outcomes for balancing bias and variance in most instances.

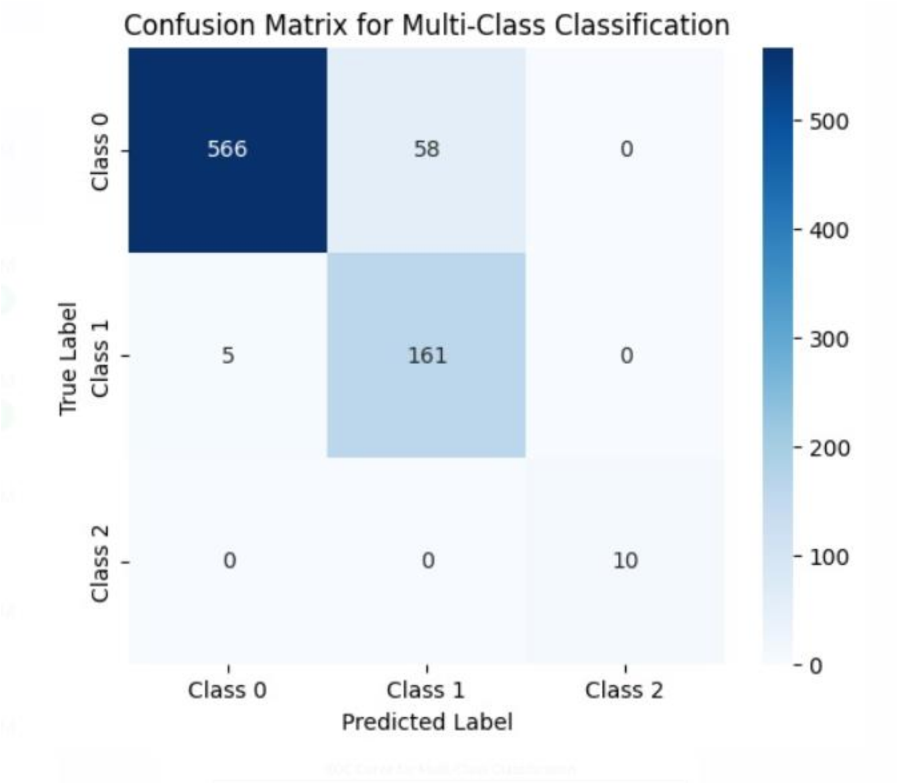

Figure 8:Confusion Matrix for a multi-class classification

A multi-class classification problem with labels Class 0, Class 1, and Class 2 is displayed in figure 8 through a Confusion Matrix. Actual results from predicted values are displayed in the matrix through a diagonal layout that shows correctly identified samples. Class 0 instances numbered 566 received accurate predictions of their original class label yet Class 0 instances totaling 58 received incorrect predictions which labeled them as Class 1. The prediction accuracy of Class 1 matches its other counterparts as the prediction model correctly identifies 161 instances of the class. Among the predictions in Class 2 there are ten accurate results with no incorrect classifications. The table data illustrates successful discrimination among classes except for minor misclassifications between Class 0 and Class 1.

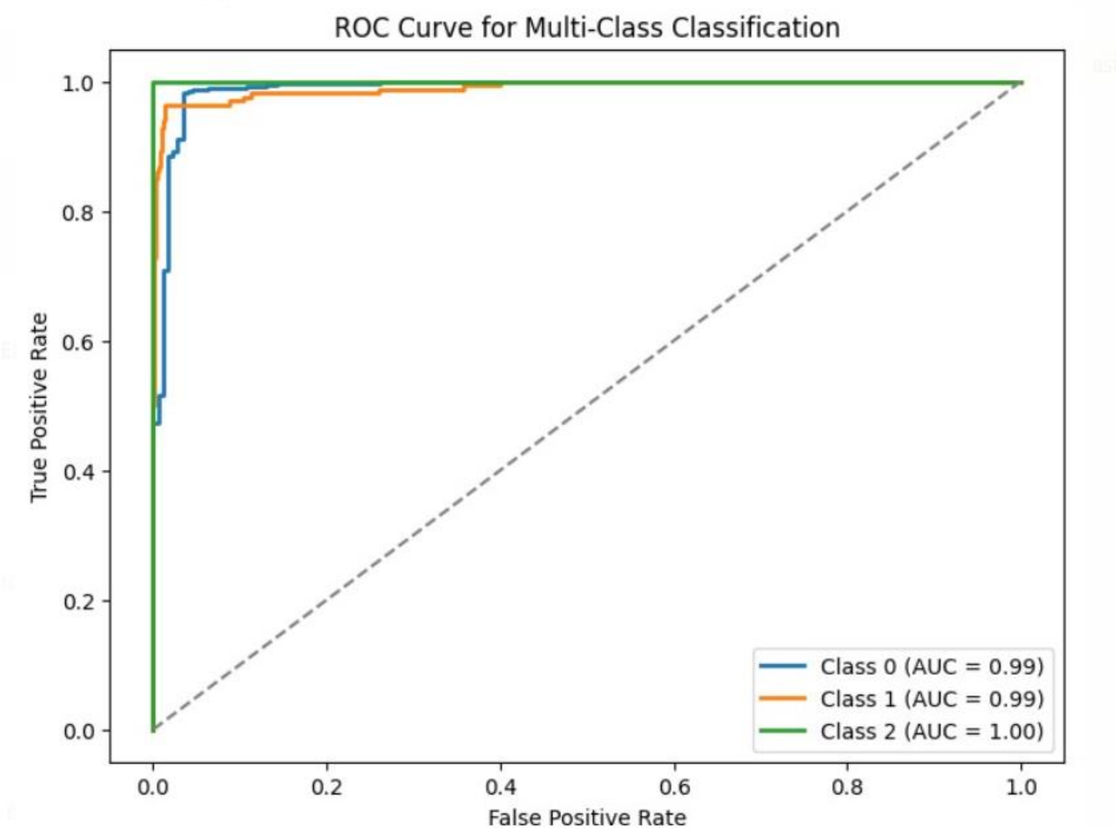

Figure 9:ROC Curve for a multi-class classification

The ROC curve representation in Figure 9 shows which class receives continuous measurement while plotting true positive rates against false positive rates for each class. Both curves for Class 0 and Class 1 represented in blue and orange lines indicate strong model performance through their high values on the AUC metric (0.99). The green curve in Class 2 demonstrates a full AUC of 1.00 which describes impeccable performance for discriminating this category. The diagonal dashed line indicates what the random classifier would achieve. This plot demonstrates the excellent discrimination abilities of the model and its high accuracy across all classes with a specific focus on Class 2.

Table 2:Classification Performance Metrics

Class	**Precision**	**Recall**	**F1 Score**
Class 0	0.991243	0.907051	0.947280
Class 1	0.735160	0.969880	0.836634
Class 2	1.000000	1.000000	1.000000
Overall	0.921250	NaN	NaN

Table 2 shows the scores for precision, recall and F1 measure of each class in the multi-class classification setup. Class 0 exhibits high performance across all metrics, with a precision of 0.9912, recall of 0.9071, and F1 score of 0.9473. The precision rate of Class 1 stands at 0.7352 yet its recall rate reaches 0.9699 and F1 score maintains a balanced metric of 0.8366. The evaluations of Class 2 show 1.0 precision and 1.0 recall coupled with perfect F1 accuracy of 1.0. The 0.9213 accuracy figure confirms an effective execution of the model throughout. The model shows excellent capability in recognizing Class 1 and Class 2 through its high recall value.

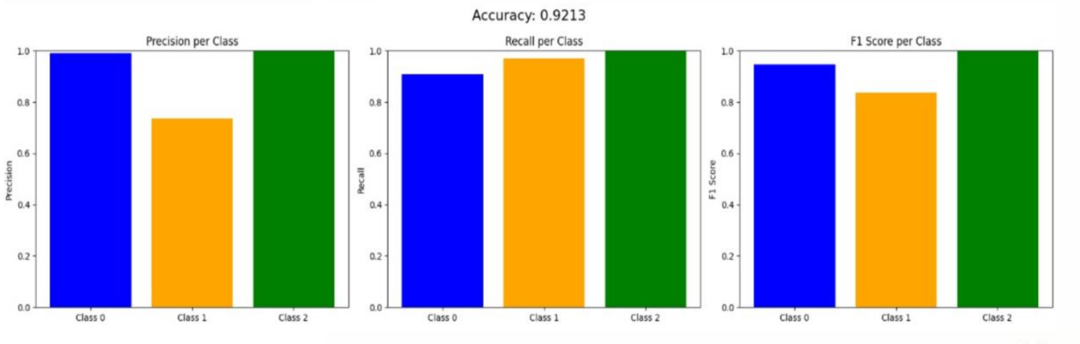

Figure 10 : Performance metrics for a multi-class classification model

The illustration depicts multi-class classification metrics through precision, recall and F1 scores provided for Class 0, Class 1 and Class 2. The left side of the figure 10 demonstrates that Class 0 achieves the highest precision whereas Class 1 shows the lowest precision levels. The recall analysis in the center demonstrates Class 0 and Class 1 share comparable recall measurements whereas Class 2 reaches a flawless result of 1.0. The right side of the figure displays F1 scores that indicate Class 0 achieves the highest score followed by Class 2 then Class 1. The entire model achieves 0.9213 accuracy which proves its ability to perform well for all identified classes.

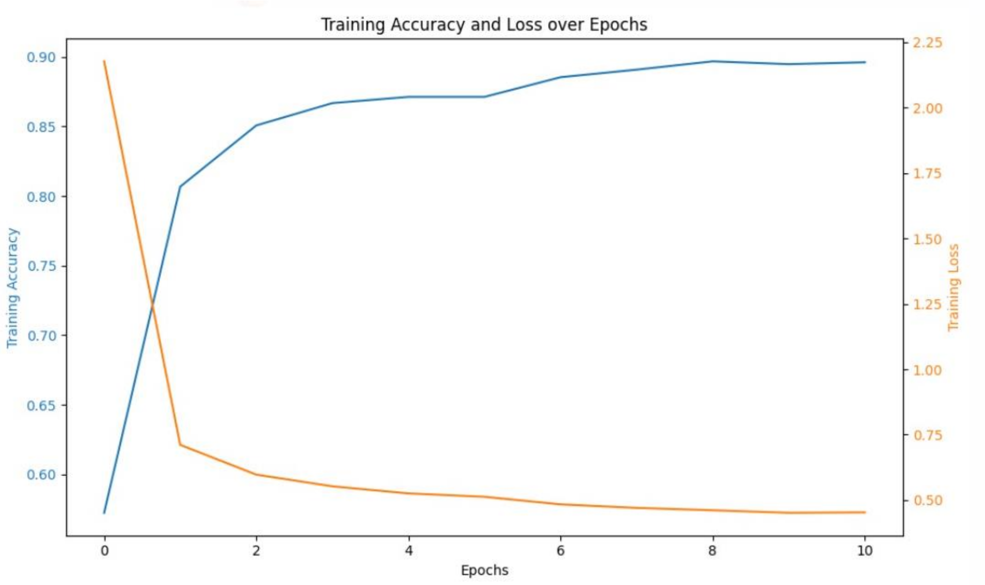

Figure 11: Training accuracy and training loss over epochs

The training accuracy and training loss of the model can be seen in figure 11 throughout its training duration across epochs. During training the blue line shows accuracy improvement until it reaches 0.9 at epoch number 10 indicating the model successfully learns from the data. Training loss diminishes quickly until the first epochs before achieving stability at 0.5 as shown by the orange line. The model demonstrates fast adaptation to the data but shows a gradual approach to its ideal outcome based on its decreased loss patterns throughout time. The visual representation shows a standard learning effect because the model simultaneously reaches maximal accuracy levels with minimal loss points throughout its development process.

5) Conclusion

The proposed approach completes its mission by showing effective deployment of multiple sequential machine learning techniques for diagnostic applications of schizophrenia. By implementing preprocessing techniques that used SMOTE to handle data imbalance and feature normalization the model acquired the ability to analyze complex data relationships. Through the usage of Lasso Regression the model optimized feature selection while concentrating on predictive features only. The combination of CNN-LogRegNet structure enabled the model to extract valuable features which produced superior classification outcomes. The model achieved superior classification performance for Class 2 with high precision and recall scores along with F1 scores as documented in the results. During training the designed model demonstrated fast convergence while maintaining high accuracy together with minimal loss signals efficient learning. The model shows meaningful diagnostic potential for schizophrenia diagnosis when processing clinical information but requires additional adjustments to enhance generalization effectiveness across different datasets.