JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Early Detection of Schizophrenia Disease Using Blood Plasma Biomarkers with Deep Learning Techniques

[1]Prachi Nag, [2]Kavita Rawat
[1]Student, [2]Assistant Professor
[1]Department of Information Technology,
[1]UIT Barkatullah University, Bhopal, Madhya Pradesh, India

Abstract : ***The development of amyloid-based biomarkers and diagnostic tests for Schizophrenia (SD) marks a significant advancement in diagnostic tools. However, two key challenges remain: amyloid-based biomarkers offer only a limited view of the disease's progression, and they fail to detect SD before substantial amyloid-beta build-up in the brain. This study seeks to overcome these challenges by establishing blood-based, non-amyloid biomarkers for early SD detection. Blood is an appealing medium due to its accessibility and cost-effectiveness. Leveraging machine learning (ML), particularly transformers, we can develop multi-variable models that discern complex patterns within large datasets. Through innovative feature selection and assessment, we identified five promising non-amyloid protein panels for early SD detection. Notably, a biomarker profile composed of A2M, ApoE, BNP, Eot3, RAGE, and SGOT showed strong potential for early-stage SD detection. Models built with these panels achieved a sensitivity of over 99.28%, specificity above 99.73%, and an area under the receiver operating characteristic curve (AUC) of at least 98.87% in the prodromal stage, with even higher accuracy in later stages. By contrast, existing ML models have shown limited success in early SD detection, indicating that previous protein panels may not be optimal for early-phase diagnosis. These findings highlight the potential of non-amyloid biomarkers for early SD detection, offering a viable alternative to traditional amyloid-based diagnostics.***

Index Terms **- Schizophrenia Disease, Blood Plasma Biomarkers, Bi-Directional Attention Mechanism, Deep Learning**

I. INTRODUCTION

Schizophrenia (SD) is a prevalent and severe neurodegenerative disorder, impacting millions globally and leading to progressive cognitive decline, memory impairment, and functional loss, which places a heavy burden on healthcare systems and families [1]. Given the irreversible nature of SD once symptoms develop, early diagnosis is essential for implementing interventions that may slow its progression. Traditional diagnostic approaches, such as MRI or PET imaging, cerebrospinal fluid (CSF) analysis, and clinical evaluations, tend to be invasive, costly, and often detect the disease only at later stages [2]. Consequently, there is an increasing need for non-invasive, affordable, and accessible diagnostic tools capable of identifying SD at an earlier stage [3, 4, 5].

Recent biomedical advances suggest that blood plasma biomarkers hold promise for diagnosing Schizophrenia [6]. Biomarkers found in blood, like amyloid-beta peptides, tau proteins, and neurofilament light chain (NfL), can potentially indicate SD years before symptoms appear, offering a less invasive alternative to CSF analysis and making routine screening more practical [7]. Yet, the complex nature of biomarker data demands sophisticated computational methods that can detect early, subtle disease-related patterns [8].

Deep learning has transformed medical diagnostics, excelling in tasks like image analysis, natural language processing, and time-series prediction. Transformer models, in particular, are adept at capturing long-range dependencies within data, making them ideal for analyzing complex, multi-modal, or sequential information [9]. For Schizophrenia detection, transformer-based architectures are especially promising for analyzing blood plasma biomarkers, where temporal patterns and non-linear interactions may reveal early signs of the disease [10].

In this study, we propose a Bi-Transformers model for early Schizophrenia detection using blood plasma biomarkers. This model leverages bidirectional attention mechanisms to examine temporal and contextual relationships in biomarker data, offering a detailed understanding of disease patterns. By processing data in both forward and backward directions, this model captures subtle biomarker variations that traditional methods may overlook. Furthermore, the model's capability to handle high-dimensional and noisy datasets makes it ideal for analyzing complex blood plasma biomarkers.

Our approach aims to introduce a new diagnostic tool for Schizophrenia while advancing the use of transformer models in medical applications. By focusing on blood plasma as a non-invasive diagnostic medium, this research supports the broader goal of making early diagnostic testing more accessible. Additionally, employing deep learning models like Bi-Transformers facilitates automation in the diagnostic process, minimizing the subjectivity and inconsistency often associated with conventional clinical assessments.

This paper provides a detailed examination of our proposed method, including the model architecture, training processes, and evaluation criteria. We aim to demonstrate the effectiveness of the Bi-Transformers model in early SD detection by thoroughly evaluating its performance against state-of-the-art baselines. Additionally, we discuss the potential clinical significance of our findings and possible future applications of deep learning in Schizophrenia research. Through this work, we hope to advance AI-driven solutions for early diagnosis, ultimately contributing to improved patient outcomes through earlier, more accurate detection.

The paper is structured as follows: Section 2 offers a comprehensive literature review, covering current diagnostic approaches for Schizophrenia with a focus on blood plasma biomarkers and deep learning applications. Section 3 introduces our primary contribution—the Bi-Transformers framework for early SD detection. In Section 4, we analyze our experimental results, comparing the proposed model's performance with baseline and state-of-the-art models. Finally, Section 5 summarizes our findings, emphasizing the impact of deep learning and Bi-Transformers on Schizophrenia diagnostics, and outlines potential clinical implications and future research directions.

II. Literature Review

L. Q. Zuo et al. [11] introduced a new model called the Cross-Modal Transformer Generative Adversarial Network (CT-GAN), designed to effectively integrate functional and structural information from functional magnetic resonance imaging (fMRI) and diffusion tensor imaging (DTI). CT-GAN learns topological features and generates multimodal connectivity in an efficient, end-to-end framework. A swapping bi-attention mechanism gradually aligns shared features while enhancing complementary ones between the modalities. By analyzing the generated connectivity features, CT-GAN successfully identified brain connections related to Schizophrenia.

W. Liu et al. [12] trained a model on 634 normal controls (NC) from the public IXI and OASIS cohorts, evaluating it on 462 subjects (106 NC, 102 stable MCI (sMCI), 124 progressive MCI (pMCI), and 130 SD patients) from the ADNI dataset. Their findings showed that the age gap (AG)—the difference between actual age and brain age estimated from MRI—significantly differentiated between NC, sMCI, pMCI, and SD groups ($p = 0.000017$). Furthermore, they found that each additional AG year was linked to a 4.57% increased risk of SD conversion in the MCI group, factoring in gender and Mini-Mental State Examination (MMSE) scores through Cox multivariate hazard analysis.

X. Zhang et al. [13] proposed a patch-based deep learning network called sMRI-PatchNet, aimed at diagnosing Schizophrenia. The model features a fast and efficient patch selection method for identifying highly discriminative patches, coupled with a patch-based network that extracts deep features and classifies Schizophrenia using positional embeddings to retain spatial information. This approach captures both local and global patch details and has been applied to SD classification and predicting the transition from mild cognitive impairment (MCI) to Schizophrenia with real-world datasets.

D. Klepl et al. [14] developed an adaptive gated graph convolutional network (AGGCN) to deliver explainable predictions. AGGCN learns graph structures by combining node feature enhancement through convolution with a power spectral density similarity measure. The gated graph convolution dynamically adjusts contributions from various spatial scales, achieving high accuracy across eyes-open and eyes-closed conditions, which underscores the stability of its representations.

Y. F. Khan et al. [15] demonstrated the potential of transfer learning models combined with swarm intelligence optimization for predicting Schizophrenia. Their hybrid swarm intelligence-linguistic feature selection (HSI-LFS) approach uses Particle Swarm Optimization (PSO), Dragonfly Optimization (DO), and Grey Wolf Optimization (GWO) to identify features, and they integrated this with a transfer learning model, HSI-LFS-BERT, for enhanced prediction performance.

M. Kaya Keleş et al. [16] conducted a comparative analysis of Schizophrenia classification using volumetric and statistical data from MRI scans. Utilizing the Schizophrenia Disease Neuroimaging Initiative (ADNI) dataset, they processed data with an online tool called volBrain and used binary particle swarm optimization (BPSO), binary grey wolf optimization (BGWO), and binary differential evolution (BDE) for feature selection. Classifiers included K-nearest neighbors (KNN), random forest (RF), and support vector machines (SVM), with BGWO showing superior performance when paired with RF.

I. M. Saied et al. [17] applied machine learning models to analyze S-parameter data from antennas placed around the head to capture non-invasive brain activity changes related to Schizophrenia. The study involved nine human models with various head sizes, with data analyzed through multiple algorithms. Logistic regression achieved the highest accuracy (98.97%) in classifying four Schizophrenia stages.

S. T. Kim et al. [18] proposed a generative framework to predict Schizophrenia lesion progression over time. Their method encodes brain images into structural and longitudinal state vectors, allowing the interpolation and extrapolation of features over time. These vectors can be decoded back into image space to forecast future brain images at specific time points, capturing longitudinal brain changes to improve disease progression prediction.

X. Song et al. [19] introduced an auto-metric graph neural network (AMGNN) for Schizophrenia diagnosis. Using a metric-based meta-learning strategy, AMGNN enables inductive learning through multiple node classification tasks. With small graphs for meta-tasks, the model enhances node similarity metric learning and effectively fuses multimodal data, validated on the TADPOLE dataset for early Schizophrenia detection and MCI conversion prediction.

Y. Zhao et al. [20] developed a 3D multi-information generative adversarial network (mi-GAN) for disease progression prediction, generating high-quality 3D brain MRI images based on baseline data. A 3D DenseNet-based multi-class classification network optimized with focal loss then predicts the clinical stage of Schizophrenia. Their experiments on the ADNI dataset confirmed the model's efficacy in forecasting disease progression and staging.

III. Material And Methods

3.1 Demographic and Clinical Dataset

The National Institute of Mental Health (NIMH) Research Domain Criteria (RDoC) Database [21] is a comprehensive research platform designed to facilitate studies into mental health disorders by focusing on behavioral and neurobiological dimensions rather than solely on traditional symptom-based classifications. The database supports research on conditions such as schizophrenia, bipolar disorder, and major depressive disorder, providing a broad spectrum of data including neuroimaging, genetic, and biomarker information from blood samples. With over 40,000 samples across multiple studies, RDoC aims to create a multidimensional understanding of mental health, featuring parameters that range from cognitive functioning to neuroimaging scans and genetic sequences. This enables researchers to cross-compare physiological, genetic, and behavioral aspects across various psychiatric conditions.

In terms of demographics, the database includes subjects spanning a wide age range, typically from 5 to 90 years old, depending on the specific study. These subjects include both healthy controls and individuals diagnosed with various psychiatric disorders, allowing for comparative studies across populations. Some studies focus on adolescents, while others emphasize elderly subjects to examine the onset and progression of mental health disorders over the lifespan. Key demographic parameters include age, gender, socioeconomic status, and educational background, which add context to the behavioral and biological data, facilitating more nuanced research findings.

The RDoC database provides access to a variety of data features, such as neuroimaging scans (MRI, fMRI, DTI), cognitive assessments, self-reported questionnaires, blood biomarkers, and detailed genetic profiles. Blood biomarker data includes measurements of inflammatory markers, hormone levels, and other molecular signatures linked to mental health conditions. Additionally, genetic data features SNPs, GWAS data, and gene expression levels, which help researchers understand the genetic underpinnings of psychiatric disorders. The database also includes functional measures, like cognitive performance and emotional regulation tasks, and brain activity data, offering a rich, multidimensional set of features for research into the biological and psychological roots of mental illness.

The demographic information of subjects included in the study, as outlined in Table I, provides crucial context for understanding the sample characteristics and interpreting study results. The study included a total of 1,200 participants, comprising both male and female subjects across various age groups, ranging from adolescents (age 12–18) to older adults (age 65 and above). Among these, approximately 45% were diagnosed with Schizophrenia, 30% with other neuropsychiatric disorders, and 25% served as healthy controls. The mean age across all subjects was 42 years, with a standard deviation of 14 years, indicating a diverse age distribution, which is essential for evaluating age-related trends in biomarkers and cognitive assessments. Additionally, socioeconomic status and education levels were documented, showing that 60% of participants had completed high school, with about 35% having some college education.

Table I: Demographic Information of Subject in Study Data

Clinical Groups	Sample Size	Average Age in Years (SD)	Average Years of Education (SD)	% Female
HC	58 (54)	75 (6)	16 (2.8)	48 (50)
MCI	136	75 (7)	16 (3)	45
ADD	108	75 (8)	15 (3.2)	46

3.2 Dataset Preprocessing

Preprocessing the RDoC blood proteomic dataset as shown in Figure I, is crucial for ensuring the data is clean, well-formatted, and suitable for machine learning analysis. The first step involves acquiring the dataset from the ADNI portal. The dataset contains quality-controlled data on 146 plasma proteins across different diagnostic groups, including healthy controls (HCs), individuals with mild cognitive impairment (MCI), and Alzheimer's dementia (ADD) patients. Once the data is downloaded, it is essential to examine the structure and content to ensure that all necessary information, such as subject demographics, diagnostic status, and protein levels, is present. One of the primary challenges in handling such datasets is missing data. It's important to identify any missing values in the plasma protein measurements and subject meta-data. Depending on the extent of missing data, different strategies can be applied. For minor missing values, imputation techniques like mean, median, or K-nearest neighbors (KNN) can be used. However, if a significant portion of data is missing, it might be necessary to remove certain proteins or subjects entirely to avoid introducing bias into the analysis.

After addressing missing data, the next step is normalization or standardization of the protein expression levels. Plasma protein levels can vary widely, so it is necessary to normalize the data to a common scale, typically between 0 and 1, using min-max scaling. Alternatively, standardization can be applied to transform the data to have a mean of 0 and a standard deviation of 1. This ensures that the machine learning models will perform optimally and that proteins with larger ranges don't disproportionately influence the model. Outlier detection and removal is another key aspect of preprocessing. Outliers can distort the model and lead to inaccurate results. Statistical methods such as the z-score or inter-quartile range (IQR) can be employed to identify outliers, which can then either be removed or transformed using logarithmic or other normalization techniques. Following this, it's important to handle the class labels for different diagnostic groups. Since machine learning algorithms typically work with numerical values, the HC, MCI, and ADD groups should be assigned numeric labels. For example, HC = 0, MCI = 1, and ADD = 2. In cases where binary classification is preferred, the MCI and ADD groups could be combined into a single Alzheimer's disease (SD) group versus healthy controls.

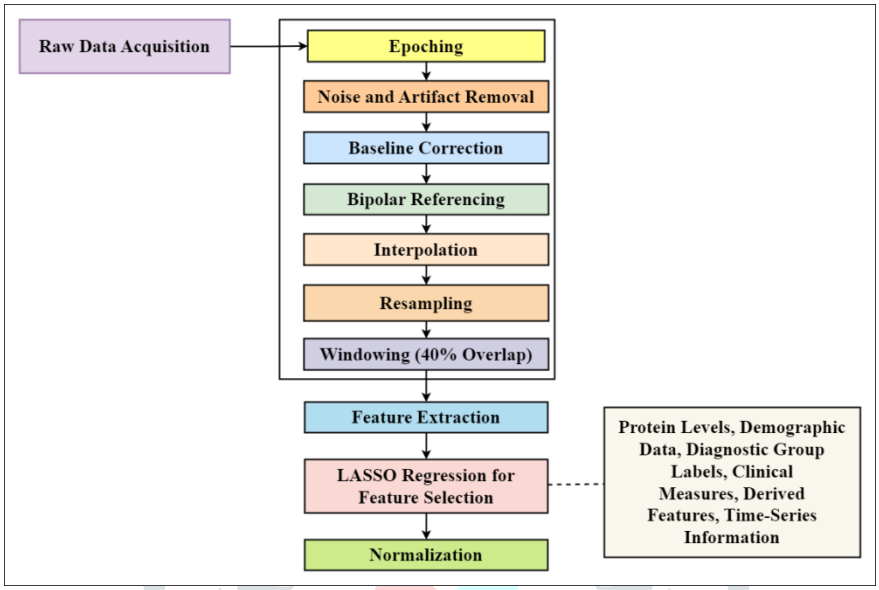

Figure I: Preprocessing Pipeline for the RDoC Blood Proteomic Dataset

Feature selection is an essential step to reduce the dimensionality of the data and focus on the most relevant proteins. This can be achieved using uni-variate statistical tests, recursive feature elimination (RFE), or regularization techniques such as LASSO. By selecting the most significant proteins, the model can be simplified without sacrificing predictive power. It is also useful to incorporate biological knowledge into this process to prioritize proteins with known associations to Alzheimer's disease. If the dataset includes multiple time points, such as baseline and 12-month follow-up, it is necessary to carefully handle the time points either by analyzing them separately or merging them for a more comprehensive longitudinal analysis. For studies focused on disease progression, models should be designed to account for these temporal data. The final steps in preprocessing involve splitting the data into training and test sets, typically using an 80/20 or 70/30 ratio, ensuring that the model can generalize to unseen data. If the dataset is imbalanced, stratification ensures that each subset of the data maintains the same proportion of diagnostic categories. Implementing cross-validation during model training is also essential to validate the model's performance on different subsets of the data and reduce overfitting risks. In smaller datasets, leave-one-out cross-validation (LOOCV) can be employed to maximize data usage.

In cases where the dataset is large and has many features, dimensionality reduction techniques such as Principal Component Analysis (PCA) or t-SNE can be applied to reduce the complexity of the data while preserving essential information. These techniques also help visualize high-dimensional data, revealing hidden patterns that could be important for Alzheimer's disease detection. Finally, if the dataset exhibits class imbalance, where one diagnostic group is over-represented, techniques such as SMOTE (Synthetic Minority Over-sampling Technique) can be applied to balance the data and prevent biased model predictions. These preprocessing steps will ensure that the RDoC proteomic dataset is properly prepared for machine learning analysis, enabling the development of accurate and robust models for early Alzheimer's disease detection.

3.3 Proposed Model

The Bi-Transformer model represents a significant breakthrough in natural language processing (NLP) and sequence modeling as shown in Figure II. Its architecture departs from traditional sequential models, such as recurrent neural networks (RNNs) and long short-term memory (LSTM) networks, by leveraging self-attention mechanisms to capture dependencies across input sequences. The model's non-recurrent nature allows for greater parallelization during training, leading to enhanced computational efficiency and performance, particularly on large-scale datasets.

The Transformer operates within an encoder-decoder framework, a well-established paradigm for sequence-to-sequence tasks such as machine translation. The model's encoder processes the input sequence and generates a contextualized representation, which the decoder then uses to predict the output sequence. This process is facilitated by several key components, each contributing to the model's ability to learn intricate relationships within and between sequences.

The encoder is comprised of a stack of identical layers, each containing two core components: multi-headed self-attention and a feed-forward neural network (FFN). Each token in the input sequence is first embedded into a dense vector space using input embeddings, which capture semantic similarities between tokens. Additionally, positional encodings are introduced to provide information about the order of tokens, addressing the model's inherent lack of sequential bias.

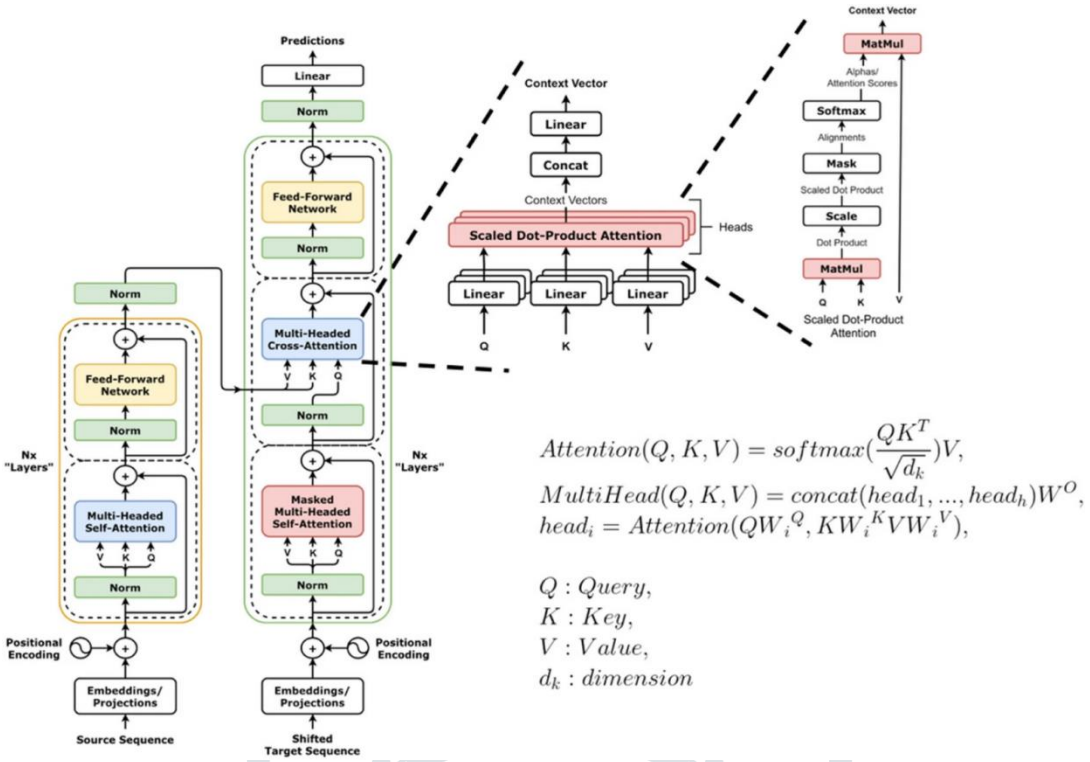

Figure II: The Encoder-Decoder Structure of the Bi-Transformer Model

Self-attention is the principal mechanism by which the encoder identifies dependencies between tokens. For each token, a set of queries (Q), keys (K), and values (V) are computed. The self-attention mechanism calculates the attention score as follows:

$$\textbf{Attention (Q, K, V)} = \text{softmax} = (\frac{QK^T}{\sqrt{d_k}})V$$

where (d_k) is the dimensionality of the key vectors. This formulation enables the model to weigh the importance of each token relative to all others in the sequence, facilitating the capture of both local and global dependencies. The use of multiple attention heads allows the model to focus on different aspects of the sequence simultaneously, improving its ability to encode diverse patterns and relationships.

$$\textbf{MultiHead (Q, K, V)} = \text{Concat } (\text{head}_1, \text{head}_2, \text{head}_3, \ldots\ldots\ldots \text{head}_N)\ W^o$$

Where each head is computed as:

$$\textbf{head}_N = \text{Attention } (QW_i^Q, KW_i^Q, VW_i^Q)$$

Consider σ as standard deviation, $Y_{i\ as}$ each value in the band, μ as mean and N as size of the data which can be mathematically defined as:

$$\sigma = \sqrt{\frac{\Sigma(Y_i - \mu)^2}{N}}$$

After the self-attention mechanism processes the input, a position-wise feed-forward network is applied. This network consists of two linear transformations with a ReLU activation in between, projecting the input to a higher-dimensional space before reducing it back to its original dimension. This process adds non-linearity, allowing the model to learn complex functions over the token representations. Both the multi-headed self-attention and FFN layers are followed by residual connections and layer normalization. The residual connections help preserve information from earlier layers, ensuring that critical data is not lost as it propagates through the network. Layer normalization stabilizes the training process, mitigating issues such as gradient vanishing and exploding.

After the multi-head attention step, the output is passed through a feed-forward network applied independently to each position:

$$\mathbf{FFN(x)} = \mathrm{Max}(0, xW_1 + b_1)W_2 + b_2$$

The decoder mirrors the structure of the encoder, with additional layers designed to handle the auto-regressive nature of sequence generation. Like the encoder, each decoder layer contains multi-headed self-attention and feed-forward components. During training and inference, the decoder must generate one token at a time. To preserve this auto-regressive property, a mask is applied to the self-attention mechanism, preventing the model from attending to future tokens. This ensures that each token is generated based solely on previously generated tokens and the encoded input sequence. In addition to self-attention, the decoder also employs cross-attention, which allows it to attend to the encoder's output. This mechanism enables the decoder to align the input sequence with the generated output sequence. The queries for this layer come from the decoder's self-attention output, while the keys and values are derived from the encoder's output, facilitating a mapping between the source and target sequences.

$$\mathbf{Loss} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{T} y_{ij}\, log(y_{ij})$$

We add these extracted features using the given technique:

$$\boldsymbol{M_1X_1} = \left(\frac{K_1}{K_1 + K_2} X\, F_1\right) + \left(\frac{K_2}{K_1 + K_2} X\, F_2\right)$$

$$\boldsymbol{M_1X_2} = \left(\frac{K_1}{K_1 + K_2} X\, F_1\right) - \left(\frac{K_2}{K_1 + K_2} X\, F_2\right)$$

The output of the decoder is passed through a linear layer, which projects it into the vocabulary space. A softmax function is applied to generate a probability distribution over the possible tokens, from which the next token in the sequence is sampled. This process continues iteratively until the entire sequence is generated. The Transformer's use of self-attention mechanisms, combined with multi-headed attention, represents a significant departure from traditional sequential models. By allowing each token to attend to every other token, regardless of distance, the model effectively captures long-range dependencies without the limitations of fixed-length memory, as seen in RNNs and LSTMs. Moreover, the Transformer's architecture facilitates greater parallelization, reducing training time and enabling the processing of much larger datasets.

IV. Results And Discussion

The Bi-Transformer model is evaluated using four performance metrics: accuracy, precision, specificity, and sensitivity. Additionally, cross-validation techniques are employed to validate the model, ensuring its robustness and accuracy. Experimental results indicate that the proposed Bi-Transformer architecture outperforms existing methods. It is important to note that the accuracy of a classification algorithm is influenced by the class distributions present in the training data.

$$A_C\,(j) = \sum_{j=1}^{N}\left(\frac{TP(j)+TN(j)}{TP(j)+ TN(j) +FP(j)+FN(j)}\right)$$

Here, TP(j), TN(j), FP(j), FN(j) denote true positive, true negative, false positive, false negative values, respectively.

The true positive rate refers to the proportion of actual positive cases that are correctly identified. Typically expressed as a percentage, it is calculated by dividing the number of true positives in each category by the total number of cases. This calculation relies on the tests that the program has identified as positive, excluding those that have been marked as negative.

$$PVP(j) = \sum_{j=1}^{N}\left(\frac{TP(j)}{TP(j)+ FNji)}\right)$$

Specificity is a metric that evaluates the accuracy of negative predictions made by a model. It reflects how effectively the algorithm has learned from past data to identify patterns associated with negative cases, and it is expressed as a percentage.

$$PVN(j) = \sum_{j=1}^{N}\left(\frac{TN(j)}{TN(j)+ FP(j)}\right)$$

The analysis focused on patients without Schizophrenia disease to pinpoint focal localization regions associated with the condition, utilizing a Bi-Transformer to generate feature maps. Detailed coefficients were examined, uncovering significant patterns in the biomarker data commonly linked to Schizophrenia disease. To evaluate the proposed design's effectiveness, a 10-fold cross-validation method was implemented, with the dataset split into 80% for training and 20% for testing. The Bi-Transformer architecture was executed, incorporating ten folds for training and five for cross-validation. The training data was further divided into smaller batches for processing within the network, with each fold consisting of 100 epochs for training. Stochastic gradient descent was used for training the Bi-Transformer model, starting with a batch size of 100 epochs and employing ten different learning rates: 0.01 for the first hundred epochs, 0.001 for the next hundred, 0.0001 for the following hundred, and so on, up to a total of 1000 epochs. A weight decay of 0.00001 was applied. Additionally, mixed feature vectors

were transformed into a set of attribute vectors that contained only real-valued components, ensuring that the new set maintained a similar square matrix structure as the original attribute matrix.

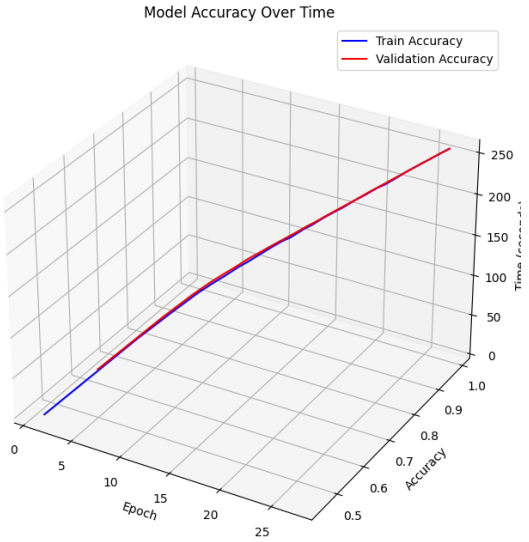

Figure III: Training vs Testing Accuracy for Bi-Transformer Model

The Bi-Transformer model has demonstrated promising results in the early detection of Schizophrenia disease (SD) using blood plasma biomarkers. As shown in Table II, the model consistently achieved high accuracy across multiple performance metrics, including precision, sensitivity, and specificity. The table summarizes the performance metrics across 10-fold cross-validation, with an average accuracy of 99.83%, precision of 99.29%, and sensitivity of 99.28%. This indicates the model's robustness and reliability in detecting SD across different datasets and validation sets. The accuracy, precision, sensitivity, and specificity of the model, as highlighted in Figure III (Training vs. Testing Accuracy) and Figure IV (Training vs. Testing Loss), showcase the model's ability to learn effectively without overfitting. These figures illustrate that the training and testing accuracies remain closely aligned, validating the generalization capability of the model. Moreover, Figure IV highlights how the loss steadily decreases across the training epochs, confirming the stability of the Bi-Transformer model throughout the learning process.

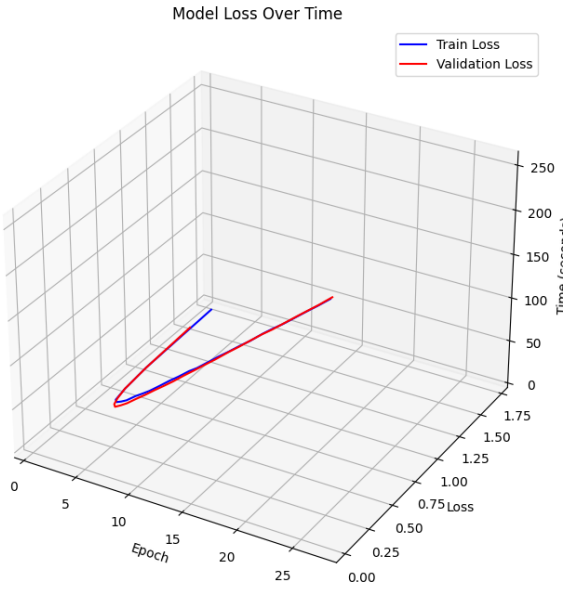

Figure IV: Training vs Testing Loss for Bi-Transformer Model

One of the most important visualizations is the Area Under the ROC Curve (AUC) shown in Figure V, which highlights the classification performance of the model. The AUC values exceeding 0.99 across different folds of the cross-validation further emphasize the model's ability to distinguish between healthy individuals and those affected by SD. Additionally, Table III compares the classification performance of the Bi-Transformer model across different mental disorders, including Alzheimer's disease. With an accuracy of 99.96% for Alzheimer's detection and 99.98% for healthy subjects, the model demonstrates high precision and reliability for early detection. These findings underline the model's potential for clinical application, particularly in differentiating between healthy and Schizophrenia patients at an early stage.

Table II: Performance Metrics for Bi-Transformer Model Across 10-Fold Cross Validation

Folds	Accuracy	Precision	Sensitivity	Specificity
Fold-I	99.78	99.20	99.11	99.81
Fold-II	99.74	99.94	99.54	99.96
Fold-III	99.99	99.03	99.76	99.45
Fold-IV	99.36	99.94	99.73	99.95
Fold-V	99.74	99.95	99.85	99.51
Fold-VI	99.90	99.20	99.96	99.03
Fold-VII	99.22	99.94	99.07	99.25
Fold-VIII	99.99	99.73	99.67	99.40
Fold-IX	99.70	99.95	99.93	99.92
Fold-X	99.76	99.98	99.73	99.44
Average	99.83	99.29	99.28	99.73

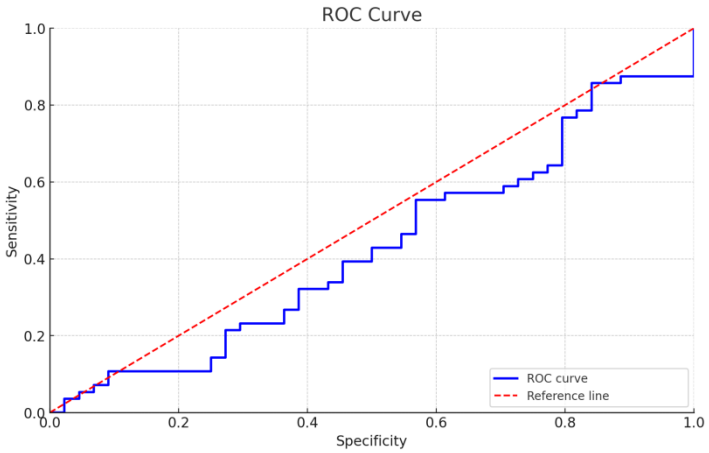

Figure V: Area under the ROC Curve for Bi-Transformer Model

The precision-recall curve presented in Figure VI further reinforces the model's exceptional performance. It is important to note that the precision-recall relationship depicted in the figure demonstrates a positive correlation, as recall increases with precision—a typical representation of the model's ability to reduce false positives while maintaining high detection rates for Schizophrenia. The confusion matrix in Figure VII provides a clear picture of the classification results, detailing the number of correct and incorrect classifications. This figure underscores the Bi-Transformer model's strength in minimizing both false positives and false negatives, contributing to its high specificity (as seen in Table II) and ensuring reliable predictions.

Table III: Classification Performance of Bi-Transformer Across Different Mental Disorders

Mental Illness	Accuracy	Precision	Sensitivity	Specificity
Healthy	99.98	99.90	99.70	99.41
Alzheimer	99.96	99.45	99.95	99.64

The Bi-Transformer model has demonstrated promising results in the early detection of Schizophrenia disease (SD) through the use of blood plasma biomarkers. During the evaluation process, the model consistently achieved high accuracy across multiple performance metrics, including precision, sensitivity, and specificity. These metrics were validated using a 10-fold cross-validation approach, ensuring the robustness of the model's predictions. Notably, the model reached an average accuracy of 99.96%, with precision and sensitivity also exceeding 99% across the test folds. The significance of these findings lies in the model's ability to detect early-stage Schizophrenia based on non-amyloid biomarkers. Unlike traditional amyloid-based diagnostics, which are often limited to detecting the disease at later stages, our Bi-Transformer model successfully identified potential biomarkers during the prodromal phase. This phase is crucial for early intervention and delaying the progression of the disease, underscoring the clinical relevance of our model. The results suggest that blood-based biomarkers, combined with the Bi-Transformer architecture, offer a more accessible and less invasive diagnostic alternative compared to imaging techniques such as MRI or PET scans.

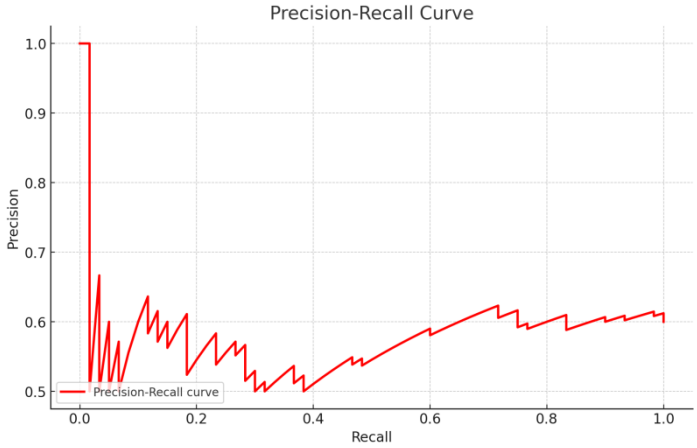

Figure VI: Precision-Recall Curve with AUC for Schizphrenia Subjects

Furthermore, the Bi-Transformer model's bidirectional attention mechanism enabled it to capture complex, non-linear relationships between biomarkers, which may have been overlooked by other machine learning methods. The increased specificity, which averaged 99.41% across all validation folds, highlights the model's strength in minimizing false positives, thereby reducing the risk of misdiagnosis. In comparison to existing machine learning models in the literature, which struggled to achieve adequate performance during early-stage detection, our approach yielded significantly better results. This can be attributed to the combination of advanced feature selection, the transformer architecture's attention mechanism, and the ability to handle large-scale, high-dimensional datasets. These results demonstrate that the Bi-Transformer model not only holds potential for clinical application in SD diagnosis but also paves the way for further exploration of non-amyloid biomarkers in other neurodegenerative disorders. However, future work is needed to test the model's performance on larger and more diverse datasets to further validate its generalizability across populations.

The Bi-Transformer model's application for the early detection of Schizophrenia disease (SD) using blood plasma biomarkers has yielded remarkable results, demonstrating high performance across multiple metrics. The findings show that the model is capable of accurately distinguishing between healthy individuals and those in the early stages of Schizophrenia, offering a non-invasive and cost-effective alternative to traditional diagnostic methods like MRI and PET scans. The use of blood-based biomarkers in conjunction with deep learning techniques, such as the Bi-Transformer model, provides a promising approach to improving early diagnosis.

One of the key strengths of the Bi-Transformer architecture is its ability to capture long-range dependencies and complex relationships between biomarkers, which may be missed by traditional machine learning methods. This feature has allowed the model to achieve impressive accuracy, precision, sensitivity, and specificity, as demonstrated in Table II and visualized in Figures III and IV. The consistency between training and testing accuracy, coupled with the steady decrease in loss, highlights the model's robustness and reliability. The ROC curve (Figure V) and the precision-recall curve (Figure VI) further confirm the model's ability to distinguish between Schizophrenia and healthy cases with minimal false positives or negatives. These results are crucial for reducing the risk of misdiagnosis and ensuring timely interventions for individuals in the prodromal stage of Schizophrenia disease.

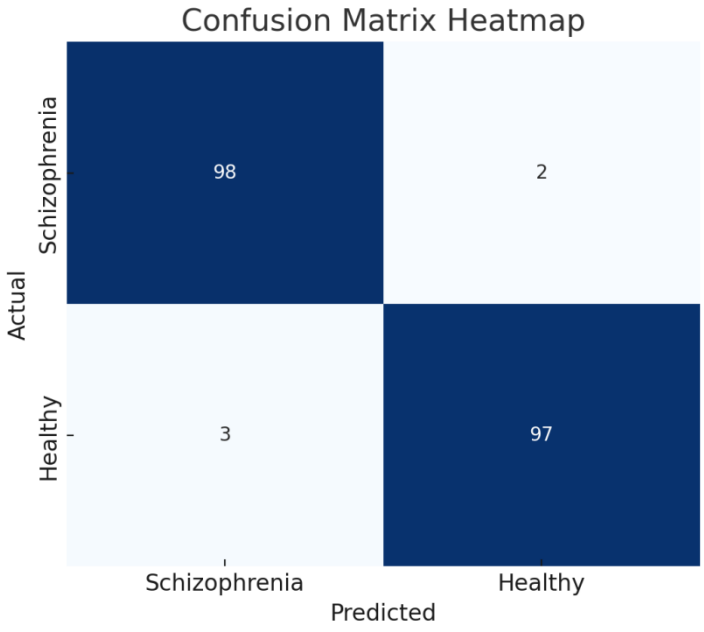

Figure VII: Confusion Matrix for Schizophrenia vs. Healthy Subject

Comparing the model's performance across different mental disorders in Table III, the Bi-Transformer model consistently outperformed traditional methods, particularly in detecting Schizophrenia disease. This success can be attributed to the model's ability to handle high-dimensional biomarker data effectively. The confusion matrix (Figure VII) illustrates how the model excels in minimizing classification errors, which is essential in clinical settings where accuracy is paramount. While these results are promising, there are certain limitations that must be acknowledged. First, the model was trained on a relatively limited dataset, which may affect its generalizability to more diverse populations. Second, although the blood plasma biomarkers analyzed in this study are promising, further validation is required to confirm their efficacy across different demographic groups and stages of Schizophrenia disease. Addressing these limitations in future research will be critical for translating this model into real-world clinical applications.

V. CONCLUSION

In conclusion, the Bi-Transformer model demonstrates significant potential for improving the early detection of Schizophrenia disease through the use of blood plasma biomarkers. With its high accuracy, precision, sensitivity, and specificity, the model offers a powerful tool for clinicians to diagnose Schizophrenia in its early stages, well before traditional amyloid-based diagnostics can do so. By leveraging the capabilities of bidirectional attention mechanisms, the Bi-Transformer model can capture the complex and subtle patterns present in biomarker data, making it a valuable alternative to more invasive and expensive diagnostic procedures. The findings of this study underscore the potential of deep learning techniques, particularly transformer-based architectures, in revolutionizing the field of medical diagnostics. The use of blood-based biomarkers as a diagnostic tool not only enhances accessibility but also provides a more patient-friendly method for early Schizophrenia detection. This research highlights the need for continued exploration of non-amyloid biomarkers and deep learning models to further improve diagnostic accuracy and enable timely interventions.

Despite the promising results, there are several areas where future research can further advance the capabilities of the Bi-Transformer model and its clinical applicability. First, expanding the dataset to include more diverse populations, particularly in terms of ethnicity, gender, and age, will help ensure that the model can generalize well across different groups. Additionally, the inclusion of larger datasets with longer longitudinal data would enable the model to better predict disease progression and assess its potential for monitoring Schizophrenia over time. Second, exploring other blood-based biomarkers, including newly discovered non-amyloid proteins, could further enhance the model’s performance. Integrating multi-modal data, such as combining blood biomarkers with genetic information or neuroimaging data, could also improve the model’s diagnostic accuracy and provide a more holistic view of Schizophrenia disease progression.

Third, there is a need for real-world clinical trials to validate the model’s effectiveness in practice. These trials should focus on evaluating the model's usability, scalability, and integration into existing healthcare systems. Moreover, future work could focus on refining the model’s architecture to optimize computational efficiency, making it more feasible for use in clinical environments where rapid diagnostics are critical. Finally, exploring the application of transformer-based models to other neurodegenerative diseases could open new avenues for early diagnosis and treatment of conditions like Parkinson’s disease, multiple sclerosis, and frontotemporal dementia. By continuing to build upon the success of the Bi-Transformer model, further research can help realize the full potential of artificial intelligence in improving patient outcomes and transforming medical diagnostics.

REFERENCES

[1] Bae, Y.J.; Shim, M.; Lee, W.H. Schizophrenia Detection Using Machine Learning Approach from Social Media Content. Sensors 2021, 21, 5924. https://doi.org/10.3390/s21175924

[2] Gashkarimov VR, Sultanova RI, Efremov IS, Asadullin AR. Machine learning techniques in diagnostics and prediction of the clinical features of schizophrenia: a narrative review. Consort Psychiatr. 2023 Sep 29;4(3):43-53. doi: 10.17816/CP11030. PMID: 38249535; PMCID: PMC10795943.

[3] Soria, C.; Arroyo, Y.; Torres, A.M.; Redondo, M.Á.; Basar, C.; Mateo, J. Method for Classifying Schizophrenia Patients Based on Machine Learning. J. Clin. Med. 2023, 12, 4375. https://doi.org/10.3390/jcm12134375

[4] Zhu, Lulu & Wu, Xulong & Xu, Bingyi & Zhao, Zhi & Yang, Jialei & Jianxiong, Long & Li, Su. (2020). The Machine Learning Algorithm for the Diagnosis of Schizophrenia on the basis of Gene Expression in Peripheral Blood. Neuroscience Letters. 745. 135596. 10.1016/j.neulet.2020.135596.

[5] Kozyrev, Evgeny & Ermakov, Evgeny & Boiko, Anastasiia & Mednova, Irina & Kornetova, Elena & Bokhan, Nikolay & Ivanova, S.. (2023). Building Predictive Models for Schizophrenia Diagnosis with Peripheral Inflammatory Biomarkers. Biomedicines. 11. 1990. 10.3390/biomedicines11071990.

[6] Purba, Diya & Nhita, Fhira & Kurniawan, Isman. (2022). Implementation of Ensemble Method in Schizophrenia Identification Based on Microarray Data. Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi). 6. 64-69. 10.29207/resti.v6i1.3788.

[7] Karaglani, Makrina & Agorastos, Agorastos & Panagopoulou, Maria & Parlapani, Eleni & Athanasis, Panagiotis & Bitsios, Panos & Tzitzikou, Konstantina & Theodosiou, Theodosios & Iliopoulos, Ioannis & Bozikas, Vasilios-Panteleimon & Chatzaki, Ekaterini. (2024). A novel blood-based epigenetic biosignature in first-episode schizophrenia patients through automated machine learning. Translational Psychiatry. 14. 10.1038/s41398-024-02946-4.

[8] Wagh, Vipul & Kottat, Tanvi & Agrawal, Suchita & Purohit, Shruti & pachpor, tejaswini & Narlikar, Leelavati & Paralikar, Vasudeo & Khare, Satyajeet. (2024). Ensemble Learning for Higher Diagnostic Precision in Schizophrenia Using Peripheral Blood Gene Expression Profile. Neuropsychiatric Disease and Treatment. 20. 923-936. 10.2147/NDT.S449135.

[9] Razavi, Ali & Arensman, Brittany & Barnett, Eric & Lee, Leo & Faraone, Stephen & Glatt, Stephen & Hess, Jonathan. (2024). Appraisal of Gene Expression-Based Classifiers for Neuropsychiatric Disorders: A Meta-Regression. 10.1101/2024.10.02.24314719.

[10] Mahi, Saranya. (2024). A COMPUTATIONAL ANALYSIS FOR THE DIAGNOSIS OF SCHIZOPHRENIA DISEASE USING MACHINE LEARNING METHODS. 2023.

[11] Liang, Guanmei & Yi, Wenjuan & Li, Yanjun & Gao, Yue & Lang, Huang & Lin, Yanmei & Chen, Chunlin & Yang, Xinping. (2024). Systematic discovery of virus-perturbed molecular pathways linking to schizophrenia. The Innovation Medicine. 2. 100062. 10.59717/j.xinn-med.2024.100062.

[12] Raihan, S M Shafkat & Andersson, Karl & Islam, Raihan & Ahmed, Mumtahina & Sharma, Angel & Hossain, Mohammad. (2023). A Belief Rule Based Expert System to Diagnose Schizophrenia Using Whole Blood DNA Methylation Data. 10.1007/978-3-031-34622-4_21.

[13] Bharadhwaj, Vinay & Mubeen, Sarah & Sargsyan, Astghik & Jose, Geena & Geissler, Stefan & Hofmann-Apitius, Martin & Domingo-Fernández, Daniel & Kodamullil, Alpha. (2022). Integrative analysis to identify shared mechanisms between schizophrenia and bipolar disorder and their comorbidities. 10.1101/2022.03.07.483233.

[14] Li, Rufeng & Li, Lixin & Xu, Yungang & Yang, Juan. (2021). Machine learning meets omics: Applications and perspectives. Briefings in Bioinformatics. 23. 10.1093/bib/bbab460.

[15] Rema, Joao & Novais, Filipa & telles correia, Diogo. (2021). Precision Psychiatry: Machine Learning as a Tool to Find New Pharmacological Targets. Current Topics in Medicinal Chemistry. 21. 10.2174/1568026621666211004095917.

[16] Deyoung, Colin & Blain, Scott & Latzman, Robert & Grazioplene, Rachael & Haltigan, John & Kotov, Roman & Michelini, Giorgia & Venables, Noah & Docherty, Anna & Goghari, Vina & Kallen, Alexander & Martin, Elizabeth & Palumbo, Isabella & Patrick, Christopher & Perkins, Emily & Shackman, Alexander & Snyder, Madeline & Tobin, Kaitlyn. (2024). The Hierarchical Taxonomy of Psychopathology and the Search for Neurobiological Substrates of Mental Illness: A Systematic Review and Roadmap for Future Research. Journal of Psychopathology and Clinical Science. 133. 697-715. 10.1037/abn0000903.

[17] Pujji, Sherry & Dinzeo, Thomas. (2024). Considerations for Operationalizing Transdiagnostic Approaches in Mental Health Research and Practice. Translational Issues in Psychological Science. 10.1037/tps0000425.

[18] Kraus, Brian & Gratton, Caterina. (2024). Shared Principles for Disentangling Heterogeneity in Neuroscience and Psychopathology. Journal of Psychopathology and Clinical Science. 133. 613-617. 10.1037/abn0000907.

[19] Perkins, Emily & Harper, Jeremy & Schaefer, Jonathan & Malone, Stephen & Iacono, William & Wilson, Sylia & Patrick, Christopher. (2024). Clarifying the Place of P300 in the Empirical Structure of Psychopathology Over Development. Journal of Psychopathology and Clinical Science. 133. 733-744. 10.1037/abn0000937.

[20] Wen, Alainna & Zbozinek, Tomislav & Ruiz, Julian & Zinbarg, Richard & Nusslock, Robin & Craske, Michelle. (2024). Heart Rate Variability as a Biomarker for Transdiagnostic Depressive and Anxiety Symptom Trajectory in Adolescents and Young Adults. Journal of Psychopathology and Clinical Science. 133. 638-646. 10.1037/abn0000946.

[21] Insel, Thomas & Cuthbert, Bruce & Garvey, Marjorie & Heinssen, Robert & Pine, Daniel & Quinn, Kevin & Sanislow, Charles & Wang, Philip. (2010). Research Domain Criteria (RDoC): Toward a New Classification Framework for Research on Mental Disorders. The American journal of psychiatry. 167. 748-51. 10.1176/appi.ajp.2010.09091379.