**Leveraging Machine Learning to Decode Genetic and Environmental Factors in Neurodegenerative and Mental Health Disorders**

*Dr Swapnita Srivastava[1], Prof Dr Divya Midhun[2], Dr. Pawan Whig[3]*
[1,2] Lincoln University, Malaysia ;[3] VIPS-TC, India
swapnitasrivastava@gmail.com, divya@lincoln.edu.my, pawan.whig@vips.edu

**Abstract**: Attention Deficit Hyperactivity Disorder (ADHD), bipolar disorder, schizophrenia, autism spectrum disorder (ASD), and Parkinson's disease (PD) are all complicated illnesses that are impacted by both environmental and hereditary variables. Environmental factors frequently influence the intensity and course of symptoms, even if genetic predisposition plays a major role in many illnesses. To create more potent, disease-modifying treatments, it is essential to comprehend how genetics and environment interact. In order to better understand the genetic, clinical, and environmental foundations of these illnesses and provide more individualized treatment regimens and precise diagnoses, machine learning (ML) and deep learning (DL) approaches are being used more and more. In the context of Parkinson's disease (PD), where the integration of clinical, genetic, imaging, and biochemical data has produced more thorough insights, this study looks at current developments in ML and DL models. It also emphasizes the growing awareness of adult ADHD and the heredity of bipolar disease, schizophrenia, and ASD. Every ailment has different diagnostic difficulties, which are sometimes made worse by symptoms that coincide with those of other conditions. Even though bipolar illness and ASD have high heredity rates, contextual variables such drug abuse, social isolation, and sleep difficulties are important in causing and intensifying symptoms. The possibilities of wearable sensors, multi-modal data integration, and explainable AI techniques for enhancing the diagnosis and treatment of various disorders are also covered in this research.

**Keywords**: *Parkinson's Disease; Heritability; Machine Learning; Mental Health Disorders; Genetic-Environmental Interaction.*

**Introduction**

The progressive neurodegenerative condition known as Parkinson's disease (PD) is typified by both motor and non-motor symptoms. Even with major advances in research, the majority of current therapies concentrate on managing symptoms. To enable a shift toward disease-modifying interventions, large datasets that capture PD's complexity and progression are essential. These datasets include clinical, genetic, imaging, and biochemical data from various patient cohorts and provide a complete understanding of the disease's impact [1].

In the context of mental health, heritability refers to the likelihood of developing a disorder due to genes inherited from one's parents. However, having a genetic predisposition does not guarantee the development of a disorder. A common analogy states that "genes load the gun, but the environment pulls the trigger." While the implications of this analogy may be debated, it effectively illustrates the significant role environmental factors play in determining whether a disorder develops and how severe its manifestations may be [2]. This understanding empowers individuals to take proactive steps to influence their mental health outcomes, even in the presence of genetic vulnerability. High heritability indicates increased susceptibility but does not equate to a 100% chance of developing the disorder.

**1. Bipolar Disorder:** Bipolar disorder ranks among the most heritable mental health conditions, with an estimated heritability of 80 to 90%. This means that having a first-degree relative—such as a parent, sibling, or child—with bipolar disorder significantly increases the risk of developing the condition [3]. A first-degree relative shares approximately half of one's DNA. Interestingly, a child's diagnosis of bipolar disorder can provide insights into a parent's condition, especially when there is a history of undiagnosed or misdiagnosed symptoms.

Diagnosing bipolar disorder often involves significant delays, with research indicating an average of six years from symptom onset to accurate diagnosis. This delay occurs because bipolar disorder frequently overlaps with other conditions, such as unipolar depression or anxiety [4]. For instance, if depression is the initial symptom, clinicians may diagnose major depressive disorder unless there is a family history of bipolar disorder. Family history serves as a critical indicator of genetic predisposition. It is estimated that 20 to 30% of individuals diagnosed with major depressive disorder eventually receive a bipolar disorder diagnosis.

The age of onset varies between bipolar disorder subtypes. Bipolar 1 typically emerges between 15 and 24 years, while bipolar 2 often appears later, between 45 and 54 years. Bipolar 2 is characterized by more depressive episodes compared to bipolar 1. For example, a 49-year-old individual with a long history of depression may reconsider their diagnosis if their 19-year-old child experiences a manic episode. In such cases, antidepressants, which are commonly prescribed for depression, may worsen symptoms by triggering mania or agitated depression. Mood stabilizers, either alone or in combination with antidepressants, are often more effective for managing bipolar disorder [5].

**2. Schizophrenia:** Schizophrenia has an estimated heritability of 70 to 80%, slightly lower than bipolar disorder but still significant. Males usually have symptoms in their late teens to early 20s, while females often experience them in their mid-20s to early 30s [6]. The average age of onset varies by gender. Environmental factors, such as substance use, can trigger or exacerbate symptoms. For instance, smoking marijuana or using stimulants may induce psychotic episodes, particularly in individuals with a genetic predisposition.

**3. Autism Spectrum Disorder (ASD):** ASD also demonstrates high heritability, with estimates exceeding 80%. Research into the genetic underpinnings of ASD is extensive, and early interventions, such as behavioral therapy, speech therapy, and developmental support, can significantly improve outcomes. Environmental triggers for ASD include changes in routine, social demands, and sensory overload, such as exposure to loud noises. Minimizing exposure to these triggers can help manage symptoms.

**4. Attention Deficit Hyperactivity Disorder (ADHD):** ADHD has a heritability estimate of 70% or higher. While historically viewed as a childhood condition, there is now greater recognition of ADHD in adults. Many individuals in their 40s and beyond may have struggled with undiagnosed ADHD, often realizing their condition only after their children are diagnosed. ADHD symptoms in adults may include difficulties with task switching, reading, or maintaining focus during conversations. However, some individuals may excel in areas that align with their strengths.

Triggers for ADHD symptom flare-ups include sleep disturbances, boredom, and under stimulation. For example, repetitive tasks at work may lead to restlessness, while certain food dyes or dietary patterns may exacerbate symptoms in some individuals. Prioritizing sleep hygiene and minimizing exposure to triggers can help manage ADHD symptoms.

**Environmental Triggers and Preventive Measures:** Environmental factors show a crucial role in the beginning and severity of mental health conditions. For bipolar disorder, even minor changes in sleep patterns can destabilize mood. ADHD is closely linked to disruptions in the body's circadian rhythm, making consistent sleep routines essential. Substance use, such as cannabis or alcohol, can worsen symptoms of bipolar disorder and schizophrenia. Sensory overload and social isolation are common triggers for schizophrenia, while changes in routine and overwhelming sensory experiences can exacerbate ASD symptoms.

**Related work**

The diagnosis and monitoring of Parkinson's Disease (PD) have seen significant advancements through the integration of ML, DL, and multi-modal data analysis. Below table 1 provides summary of recent studies, highlighting their objectives, methodologies, advantages, and limitations.

*Table 1.* Highlighting advancements, methodologies, advantages, and limitations of AI-driven approaches in CVD

| Reference | Objective | Methodology | Advantage | Limitations |
|---|---|---|---|---|
| **[7] De Fazio et al. (2025)** | Develop a smart glove for PD evaluation using flexible sensors. | Integrated flexible MEMS piezoelectric and inertial sensors in a TPU-based glove to track hand and arm movements during MDS-UPDRS tests. Embedded ML algorithms classify movements with high accuracy. | High accuracy in movement classification (>95%), portable system, real-time assessment, and cost-effective | Small sample size (7 individuals), limiting generalizability and robustness. |

| | | | | |
|---|---|---|---|---|
| | | | solution for tracking PD severity. | |
| **[8] Khanom et al. (2025)** | Enhance PD diagnosis using an ensemble boosting machine with explainable ML. | Combined AdaBoost and LDA to build an interpretable boosting framework using 195 clinical records from the UCI dataset. Feature selection via SHAP and LIME improves model understanding. | Achieved 99.44% accuracy; provides global and local explanations, improving trust in diagnostic predictions. | Dataset size is limited, and reliance on clinical data may miss multi-modal insights (e.g., imaging, biomechanical data). |
| **[9] Arnab et al. (2025)** | Review multi-modal approaches for PD diagnosis using ML and DL. | Explored various ML/DL methods across multiple datasets, emphasizing the integration of imaging, genetic, and clinical data to enhance early detection of PD. | Comprehensive review of cutting-edge ML/DL techniques, highlights potential for multi-modal diagnostic accuracy improvement. | Primarily a review; lacks direct experimental validation or specific benchmarks. |
| **[10] Rojas-Velazquez et al. (2025)** | Use microbiome analysis and ML to identify PD-related biomarkers. | Applied DADA2 for sequence processing. Extra Trees classifier validated selected features across four datasets. | Identified microbial signatures associated with PD; demonstrated potential of microbiome-based diagnosis with >80% accuracy. | Moderate accuracy (AUC: 0.62–0.74); variability across datasets; requires further validation and larger samples. |
| **[11] Dattola et al. (2025)** | Develop unsupervised learning techniques for tremor classification in PD using wearable sensors. | K-means clustering on accelerometer data classified tremor severity. Compared Vision Transformer and CNN models for handwriting analysis. | Objective tremor analysis using wearables; 99.9% accuracy for handwriting analysis with Vision Transformers. | Limited dataset for wearable-based tremor analysis; multiclass classification challenges (accuracy: 57.1%). |
| **[12] McFleder et al. (2025)** | Examine DBS's role in halting immune dysregulation in PD. | Analyzed immune profiles in PD patients using RNA-sequencing and immunohistochemistry. Linked DBS to reduced neuroinflammation and altered immune profiles. | Provides insights into DBS as a disease-modifying intervention; links immune modulation to therapeutic outcomes. | Focused on DBS-treated patients; findings may not apply to early-stage PD or non-DBS therapies. |
| **[13] Tahedl et al. (2025)** | Predict clinical progression in PD using domain-specific models. | Used longitudinal data and the Mosaic approach to identify immune-related markers predicting motor symptom progression. | Highlights potential of biomarkers for disease progression prediction; integrates multi-dimensional data. | Limited to immune markers; lacks external validation on independent cohorts. |
| **[14] Yin et al. (2025)** | Introduce Cognitive Few-Shot Learning (CFSL) for PD diagnosis. | Combined landmark learners with few-shot learning models for interpretable paralinguistic feature analysis. Compared | Superior performance (4–12% improvement); interpretable | Focus on limited features (paralinguistic); challenges in |

| | | | |
|---|---|---|---|
| | against seven FSL models on three datasets. | framework for PD and cleft lip diagnosis. | generalizing beyond studied datasets. |

**Key Contribution**

This review paper provides a comprehensive comparative analysis of the latest machine learning (ML) models applied to cardiovascular disease (CVD) diagnosis, risk prediction, and management. By systematically evaluating state-of-the-art ML techniques.

**Method, Experiments and Results**

**Dataset:** The dataset appears to be related to Parkinson's disease detection, as it contains voice features and a status column indicating whether a person has Parkinson's or not.

Table 2: Dataset description

**Pre-processing:**

**Encoding:** Ensure that categorical variables (e.g., sex, chest_pain_type, etc.) are properly encoded.

**Scaling/Normalization:** Consider scaling or normalizing numerical features like age, resting_blood_pressure, cholesterol, etc., especially if using algorithms sensitive to feature magnitudes.

**Handling Missing Values:** Although the dataset currently has no missing values, it's always good practice to check for and handle any missing data in real-world scenarios.

**Different ML Models:**

comparison of the Random Forest (RF), K-Nearest Neighbors (KNN), SVM with RBF Kernel (SVM_RBF), Decision Tree (DT), and Multilayer Perceptron (MLP) algorithms in terms of their key characteristics, advantages, and limitations describe in table 3 and table 4 provides ML model performance metric.

**Table 3:** ML model Key characteristics, Advantage, and Limitation

| Algorithm | Key Characteristics | Advantages | Limitations |
|---|---|---|---|
| | - Instance-based learning | - Simple to implement and understand | - Computationally expensive for large datasets |
| **KNN** | - Non-parametric | - No training phase | - Sensitive to irrelevant features and noise |
| | - Lazy learner | - Effective for small datasets | |
| | - Kernel-based method | - High accuracy for complex datasets | - Computationally intensive |
| **SVM_RBF** | - Effective for non-linear data | - Robust to overfitting in high-dimensional spaces | - Requires careful tuning of hyperparameters (e.g., C, gamma) |
| | - Margin maximization | | |
| | - Tree-based model | - Easy to visualize and interpret | - Prone to overfitting |
| **DT** | - Splits data based on feature values | - Handles both categorical and numerical data | - Delicate to small changes in data |
| | - Interpretable | | |
| | - Ensemble of decision trees | - High accuracy and robustness | - Computationally expensive |
| **RF** | - Bagging technique | - Handles missing data and outliers well | - Less interpretable than single decision trees |
| | - Reduces overfitting | | |
| **MLP** | - Feedforward neural network | - Can model complex, non-linear relationships | - Requires large amounts of data |

| | | |
|---|---|---|
| - Multiple layers of neurons<br>- Non-linear mapping | - Scalable to large datasets | - Computationally expensive and hard to interpret |

**Table 4:** ML model performance metric

| Metric | KNN | SVM_RBF | DT | RF | MLP |
|---|---|---|---|---|---|
| **Accuracy** | Moderate | High | Moderate | High | High |
| **Interpretability** | Low | Low | High | Moderate | Low |
| **Training Speed** | Fast (no training) | Slow | Fast | Moderate | Slow |
| **Scalability** | Poor (large data) | Moderate | Moderate | High | High |
| **Overfitting Risk** | Low | Low | High | Low | Moderate |

**Result:**

**1. Accuracy:** Measures overall correctness of the model.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

Measures the proportion of correct predictions.

**2. Matthews Correlation Coefficient (MCC):**

$$MCC = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (2)$$

A balanced metric for binary classification, even when classes are imbalanced.

**3. F1 Score:** Balances precision & recall for better performance evaluation.

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (3)$$

The harmonic mean of precision and recall, balancing FP and FN.

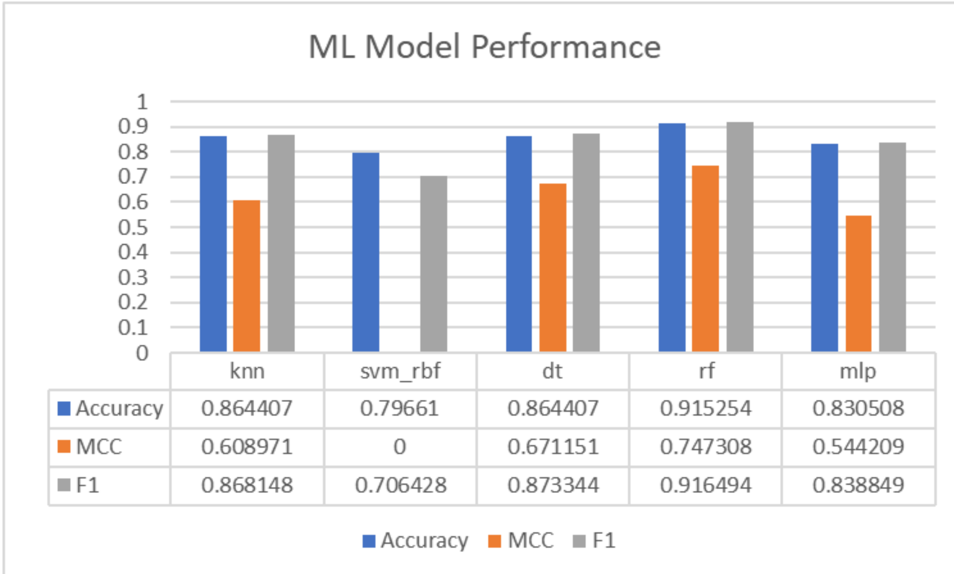


*Figure 2. ML Model Performance.*

**Discussions**

Our knowledge of neurodegenerative and mental health conditions has been completely transformed by the combination of genetic, environmental, and machine learning data. In this regard, Parkinson's disease (PD) continues to be one of the most researched conditions, and advances in machine learning and multi-modal data collecting are opening the door to more precise diagnosis and improved prediction models. New insights into the clinical development of Parkinson's disease can be gained by the use of wearable sensors, such as smart gloves for movement analysis and machine learning-based models for early identification. However, the issues persist, notably in terms of data generalization, as studies generally have low sample sizes, which might impair the robustness of prediction models.

While there are known genetic predispositions for illnesses including bipolar disorder, schizophrenia, and autism spectrum disorder (ASD), environmental variables like substance abuse, sleep disorders, and social stresses can drastically change how the disease manifests. Improving patient outcomes requires treatment regimens to take these environmental factors into account. Bipolar disorder in particular frequently coexists with other mental diseases, making diagnosis more difficult. Bipolar illness can take up to six years to diagnose, which emphasizes the need for more sophisticated diagnostic methods that include clinical symptoms, family history, and new biomarkers.

Though more research is required to increase model interpretability, dataset variety, and the capacity to accurately forecast long-term illness development, machine learning has enormous promise to close the gap in these diagnostic problems.

**Conclusions**

The increasing importance of genetic and environmental variables in the diagnosis and treatment of mental health and neurodegenerative diseases is highlighted in this study. Personalized medicine's future lies in machine learning models, particularly those that integrate genetic, clinical, and environmental data. Utilizing multi-modal data, whether from wearable technology or microbiome research, has the potential to improve illness monitoring, identify new indicators for disease development, and enable early diagnosis, as PD has shown. Given the heritability of disorders such as ASD, schizophrenia, and bipolar disorder, early detection based on family history may result in better results when paired with focused therapy.

However, even though machine learning has advanced significantly, it still has to be improved for use in clinical contexts. Addressing the constraints related to sample size, model transparency, and the integration of many data sources is essential. To make AI models more accessible and useful for physicians, future research should concentrate on enhancing their interpretability and resilience. Furthermore, knowing how genetic predispositions and environmental triggers interact might assist customize treatment and prevention strategies for high-risk people. Therefore, the next generation of diagnostic and therapeutic tools will require multidisciplinary collaboration between geneticists, doctors, and data scientists.

**References**

[1] Pezel, T., Toupin, S., Bousson, V., Hamzi, K., Hovasse, T., Lefevre, T., ... & Garot, J. (2025). A Machine Learning Model Using Cardiac CT and MRI Data Predicts Cardiovascular Events in Obstructive Coronary Artery Disease. Radiology, 314(1), e233030.
[2] Oke, O. A., & Cavus, N. (2025). Electrocardiogram image classification for six classes of heart diseases. Iran Journal of Computer Science, 1-21.
[3] Singh, J., Kumar, V., Sinduja, K., Ekvitayavetchanukul, P., Agnihotri, A. K., & Imran, H. (2025). Enhancing Heart Disease Diagnosis Through Particle Swarm Optimization and Ensemble Deep Learning Models. In Nature-Inspired Optimization Algorithms for Cyber-Physical Systems (pp. 313-330). IGI Global Scientific Publishing.
[4] Hempel, P., Ribeiro, A. H., Vollmer, M., Bender, T., Dörr, M., Krefting, D., & Spicher, N. (2025). Explainable AI associates ECG aging effects with increased cardiovascular risk in a longitudinal population study. npj Digital Medicine, 8(1), 25.
[5] Matusik, P. S., Mikrut, K., Bryll, A., Popiela, T. J., & Matusik, P. T. (2025). Cardiac Magnetic Resonance Imaging in Diagnostics and Cardiovascular Risk Assessment. Diagnostics, 15(2), 178.
[6] Mishra, A. P., & Panigrahi, S. (2025). Computer-Aided Ensemble Method for Early Diagnosis of Coronary Artery Disease. In Computational Intelligence for Oncology and Neurological Disorders (pp. 253-266). CRC Press.
[7] De Fazio, R., Del-Valle-Soto, C., Mastronardi, V. M., De Vittorio, M., & Visconti, P. (2025). A smart glove to evaluate Parkinson's disease by flexible piezoelectric and inertial sensors. Sensors International, 6, 100324.
[8] Khanom, F., Uddin, M. S., & Mostafiz, R. (2025). PD_EBM: An Integrated Boosting Approach Based on Selective Features
[9] Arnab, S. B., Palash, M. I. A., Islam, R., Ovi, H. H., Yousuf, M. A., & Uddin, M. Z. (2025). Analysis of Different Modality of Data to Diagnose Parkinson's Disease Using Machine Learning and Deep Learning Approaches: A Review. Expert Systems, 42(2), e13790.
[10] Rojas-Velazquez, D., Kidwai, S., Liu, T. C., El-Yacoubi, M. A., Garssen, J., Tonda, A., & Lopez-Rincon, A. (2025). Understanding Parkinson's: The microbiome and machine learning approach. Maturitas, 193, 108185.
[11] Dattola, S., Ielo, A., Quartarone, A., & De Cola, M. C. (2025). Integrating Wearable Sensor Signal Processing with
[12] McFleder, R. L., Musacchio, T., Keller, J., Knorr, S., Petschner, T., Chen, J. Z., ... & Ip, C. W. (2025). Deep brain stimulation halts Parkinson’s disease-related immune dysregulation in the brain and peripheral blood. Brain, Behavior, and Immunity, 123, 851-862.
[13] Tahedl, M., Bogdahn, U., Wimmer, B., Hedderich, D. M., Kirschke, J. S., Zimmer, C., & Wiestler, B. (2025). Domain-Specific Prediction of Clinical Progression in Parkinson's Disease Using the Mosaic Approach. Brain and Behavior, 15(1), e70289.
[14] Yin, P., Song, J., Bouteraa, Y., Qian, L., Martín, D., & Khishe, M. (2025). A cognitive few-shot learning for medical diagnosis: A case study on cleft lip and palate and Parkinson’s disease. Expert Systems with Applications, 262, 125713.