A NOVEL APPROACH FOR BIPOLAR DISORDER PREDICTION USING PSO-GKFCM HYBRID CLUSTERING AND WEIGHTED FEATURE SELECTION

[1]**R.Saranya, PhD**, Research Scholar, Department of Computer Science, NGM College, Pollachi

[2]**Dr.S. Niraimathi,**, Associate Professor, Department of Computer Science, NGM College, Pollachi

Abstract

Bipolar disorder is a common and serious mental illness affecting millions worldwide. Severe mood swings, such as bouts of sorrow or mania, may significantly influence a person's daily functioning and quality of life. This study proposes an effective technique for predicting bipolar disorder by combining complex data preprocessing, feature selection, clustering. The approach includes an upgraded decision tree and a weighted k-means approach for missing data preparation and analysis. Relevant feature selections are made using a weighted binary BAT approach, reducing the data's dimensionality while preserving essential qualities. The data is then clustered using a hybrid approach combining the particle swarm optimization (PSO) algorithm with the Gaussian Kernel fuzzy clustering algorithm (GKFCM), resulting in a clustering accuracy of 98.6%. Overall, the proposed approach offers an effective and efficient way to predict bipolar disorder with promising results, and it has the potential to be extended to other related illnesses.

Keywords: Bipolar Prediction, Particle Swarm Optimization, GKFCM, Clustering

I. INTRODUCTION

Bipolar disorder is a common and serious mental illness affecting millions worldwide. Severe mood swings, such as bouts of sorrow or mania, may significantly influence a person's daily functioning and quality of life [1]. Early detection and treatment of a bipolar disease may improve a patient's prognosis and reduce the chance of long-term complications [2]. As a result, accurate illness prediction systems are required for the timely diagnosis and treatment of bipolar disorder [3].

Machine learning algorithms have shown significant promise in correctly predicting and diagnosing various diseases, including bipolar disorder, in recent years [4]. Yet, the success of machine learning algorithms is heavily reliant on the quality of input data and the effectiveness of preprocessing and feature selection techniques [5]. This study offers an effective disease prediction strategy for bipolar disorder that combines complex strategies to solve these difficulties [6].

The proposed disease prediction process is divided into four major modules: preprocessing and missing value analysis using an improved decision tree and weighted k-means algorithm, feature selection using a weighted binary BAT algorithm, clustering using a hybrid

method of particle swarm optimization (PSO) with Gaussian Kernel fuzzy clustering algorithm (GKFCM) [7][8].

Using an upgraded decision tree and weighted k-means method, the preprocessing and missing values analysis module cleans and preprocesses the data [9]. The feature selection module selects the most critical characteristics using a weighted binary BAT technique, minimizing the complexity of the input while keeping the essential features [10]. To improve the accuracy of the clustering process, the clustering module adopts a hybrid technique combining PSO and GKFCM [11] [12].

Overall, the proposed approach for illness prediction is a reliable and efficient strategy for predicting bipolar disorder [13]. The results of this work may provide significant insights into the development of machine learning algorithms for sickness prediction, and they may apply to other illnesses with similar features [14-16].

The primary contributions and objectives of this manuscript may be summarized as follows.

- Preprocessing and Missing values analysis using an Improved Decision tree and weighted k-means algorithm
- The critical feature selections are selected using a weighted binary BAT algorithm.
- The data has clustered with the hybrid method as PSO with GKFCM

The remainder of this paper is structured as follows. Numerous authors address a variety of bipolar diagnosis strategies in Section 2. The proposed model is shown in Section 3. Section 4 summarizes the results of the investigation. Section 5 concludes with a discussion of the result and future work.

II. BACKGROUND STUDY

Borges-Junior et al. (2018) explored using clinical data to forecast depressive relapse in bipolar disorder patients. Chen et al. (2020) used whole-brain functional connectivity to classify bipolar disorder. He et al. (2016) found differences in resting-state functional network connectivity between unmedicated patients with bipolar and major depressive disorders. Huang et al. (2018) used speech responses to detect unipolar and bipolar depressive disorders. Jie et al. (2015) discriminated bipolar disorder from major depression based on multimodal brain imaging data using SVM-FoBa feature selection. Laksshman et al. (2017) identified genomic mutations for bipolar disorder via deep learning. Librenza-Garcia et al. (2017) conducted a systematic review of the impact of machine learning techniques on the study of bipolar disorder. Lobentanzer et al. (2019) used integrative transcriptomics to reveal sexually dimorphic control of the cholinergic/neurokine interface in schizophrenia and bipolar disorder. Mansur et al. (2020) proposed a disease model of dysregulated energy expenditure for bipolar disorder. Nunes et al. (2018) used structural MRI to identify bipolar disorders in a 13-site machine learning study. O'Shea and McInnis (2016) used iPSC models to explore the neurodevelopmental origins of bipolar disorder. Rotenberg et al. (2021) explored machine learning to predict depressive relapses in bipolar disorder patients. Schwarz et al. (2019) found reproducible grey matter

patterns that index a multivariate, global alteration of brain structure in schizophrenia and bipolar disorder.

2.1 Problem specification

Bipolar disorder is a prevalent and debilitating mental illness that affects millions of people worldwide. Accurately predicting the disorder is essential for timely diagnosis and effective treatment. However, bipolar disorder diagnosis is a complex process that relies on the analysis of various patient attributes, including behavioral, psychological, and genetic factors. Additionally, the availability of large amounts of data and the presence of missing values further complicate the analysis process. Therefore, there is a need for an effective and efficient technique that can accurately predict bipolar disorder by integrating various data preprocessing, feature selection, clustering, and classification algorithms. The goal is to improve the accuracy of bipolar disorder diagnosis and provide a useful tool for healthcare professionals to support clinical decision-making.

III. MATERIALS AND METHODS

In this chapter, we present the methodology used to predict bipolar disorder using a combination of complex data preprocessing, feature selection, clustering, and classification algorithms. The proposed approach includes an upgraded decision tree and a weighted k-means approach for missing data preparation and analysis. Relevant feature selections are made using a weighted binary BAT approach, reducing the data's dimensionality while preserving essential qualities. The data is then clustered using a hybrid approach combining the particle swarm optimization (PSO) algorithm with the Gaussian Kernel fuzzy clustering algorithm (GKFCM), resulting in a clustering accuracy of 98.6%. The proposed flow chart is represented by figure 1.

3.1 Dataset:

The benchmark datasets are downloaded from Kaggle.com website https://www.kaggle.com/datasets/arashnic/the-depression-dataset. The dataset contains control and condition EEG datasets. The dataset contains 50.6 MB size.

3.2 Preprocessing and Missing values analysis using an Improved Decision tree and weighted k-means

3.2.1 Improved Decision Tree

ID3 algorithm is the gold standard for decision trees. The ID3 method evaluates the splitting attribute based on the principle of information gain, which borrows from information entropy. The entropy of a thermodynamic system measures the degree of chaos in that system. It is used to quantify risk in the study of information. Information entropy was first proposed by Claude Shannon in 1948 and is defined as the probability of a sequence of independent random events. As information is better organized, entropy decreases. On the other hand, the larger the disorder in a system, the higher the information entropy. Hence, information entropy may be interpreted as a quantitative index of system organization.

Let there be m groups, each containing n data points, that make up the sample set s. Next, assign a sample size of ni to each of m different categories Ci, where I = 1, 2,... Assigning n samples to m categories efficiently requires an information entropy of

$H(s) = \sum_{i=1}^{m} p_i log_2 p_i$, ----- (1)

And the probability that all of the data points in sample s belong to the class Ci is given by pi = ni / n. The sample will be broken down into subsets according to property A's values if specified as the splitting attribute. The entropy of the information needed to classify each sample subset properly is determined.

$H(A_j) = \sum_{i=1}^{m} p_i \, log_2 p_i$ ------- (2)

where Aj represents a subset constructed from different values of characteristic A and pi is the probability that the sample belongs to class Ci. Hence, the needed information entropy to differentiate all samples based on feature A is

$H(A) = \sum_{v=1}^{v} \frac{sv}{s} H(A_j)$ ------ (3)

Where v is a unique value for attribute A, Sv is the number of samples included within the subset of unique attribute values, and Sv / S is the weight assigned to the attribute value. Gains in understanding are, in a nutshell,

$Gain = H(s) - H(A)$ ------ (4)

Where v is a unique value for attribute A, Sv is the number of samples included within the subset of unique attribute values, and Sv / S is the weight assigned to the attribute value. Gains in understanding are, in a nutshell.

Information theory provides the foundation for both mutual information and information entropy. With it, the degree to which two variables are related may be measured. Except in rare cases, information theory can choose a sample of characteristics from raw data and directly extract the necessary information without changing the data at all. However, mutual information is primarily used for feature selection in data mining, and its potential in other areas is little explored. To choose the collection of characteristics with the least amount of duplication, mutual information is utilized to measure the mutual inclusion connection between different attributes.

Mutual information may also symbolize the degree of association between random variables. Suppose the mutual information between an attribute of the sample set and a category is 0. In that case, the attribute weakly connects with the category since it is distributed evenly across all categories. Suppose there is a large disparity in the distribution of one dimension's unique categories. In that case, they have a wealth of mutual information, indicating a strong relationship between the characteristic and the category. The "best" splitting attribute of the decision tree is determined by calculating the mutual information between the category and many characteristics. Mutual knowledge describes how well one sample characteristic fits into another category:

$I(x, y) = log \frac{p(x_i, y_j)}{p_{(xi)} p(y_j)}$ ------ (5)

The mutual information between an attribute x and a category is denoted as MI (x), where x and w are the numbers of attribute values, respectively. Regarding attribute evaluation in the decision tree classification method, mutual information is used instead of information gain. At the same time, it's easier to understand how to put up a decision tree.

3.2.2 Weighted k-means

A variation on the standard k-means clustering method, weighted k-means assigns different values to each data point at different stages of the clustering process. The traditional k-means technique gives each data item the same weight regardless of how important or unimportant it is. But, in practice, certain information may be more crucial than others when delineating clusters. The weighted k-means approach solves this issue since it gives different values to each data point depending on its importance. The weights may be taught using an algorithm for learning, or they can be predetermined. The algorithm prioritizes the most critical data points and downplays the less important ones during the clustering stage.

The weighted k-means approach re-assigns data points based on their weighted distance to the nearest cluster centre and iteratively updates the cluster centres. After a certain number of iterations, or when the cluster centres have not changed significantly, the algorithm stops. Increased robustness against outliers, better cluster quality, and faster convergence are advantages of the weighted k-means approach over the standard k-means algorithm. Applications such as picture segmentation, data mining, and bioinformatics have heavily used this technique.

When given a set of n items D = x1, x2_, xn and a positive integer K, a k-partitioning algorithm will partition D into exactly K disjoint subsets D1,_, DK. Mark this split with. Items inside the same cluster are more related to one another than to all other subjects, and each subset defines a cluster. Creating a cost function that measures the quality of clustering for each dataset subset may make the problem of determining more manageable. Each item's (gene's) attribute is expressed here as an integer. Therefore each object may be described by a real-number row vector of dimension d, where d is the total number of attributes that object has. Let's pretend there are no missing values in the dataset and that all items have the same number of characteristics. Let the set represent n items (xi, I = 1, _, n).

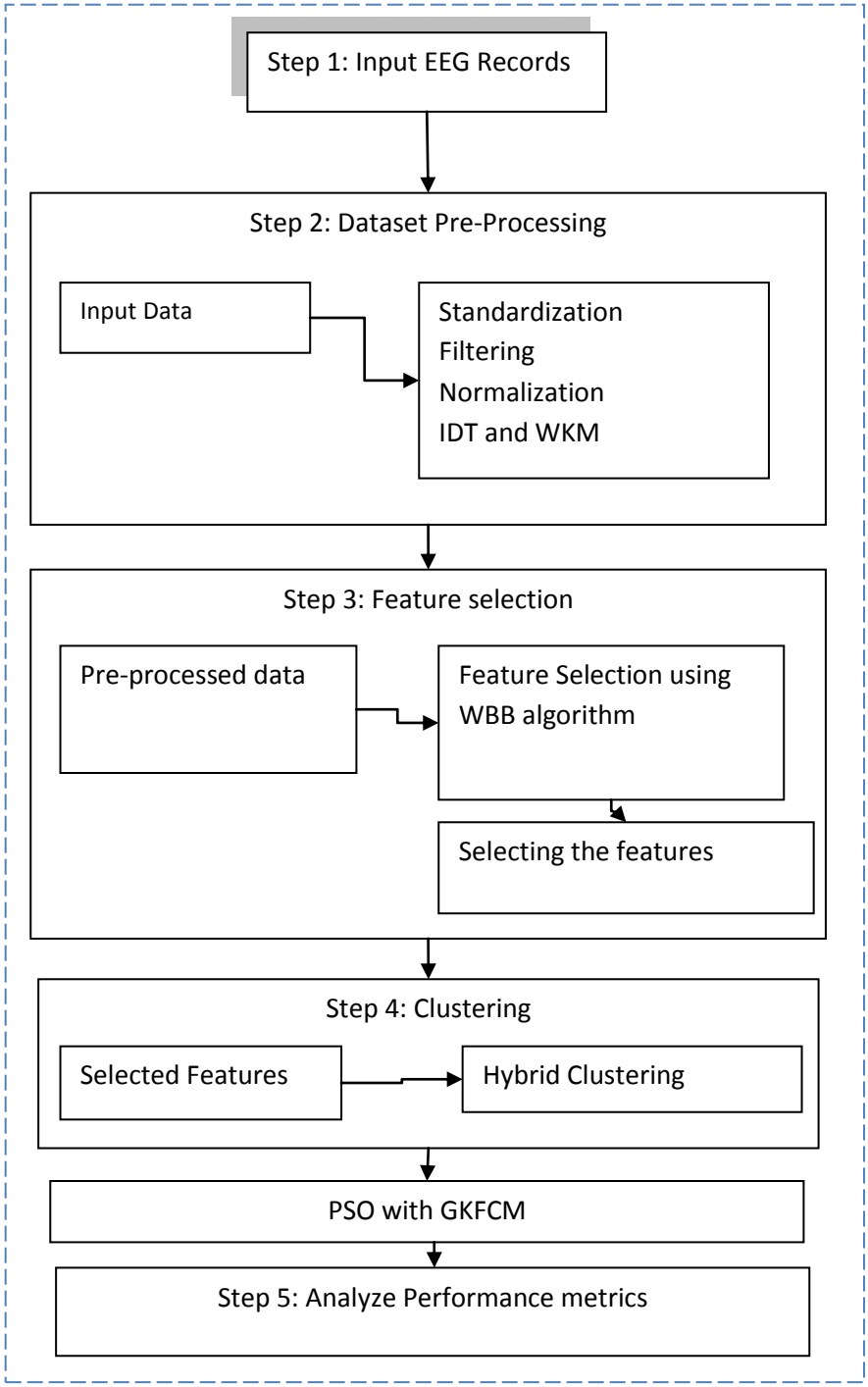

Figure 1: Proposed flow diagram

For simplicity, we will refer to xi's jth attribute as xij. The notation X = (xij) denotes an object set D attribute matrix. For a weighted k-means clustering technique, the cost function may be written as:

$j_G(\Delta) = \sum_{k=1}^{k} \sum_{x_i \in D_k} (x_i - m_k) G (x_i - m_k)$ -------- (6)

$m_k = \frac{1}{n_k} \sum_{x_i \in D_k} x_i$ ------- (7)

G is a symmetrically positive weighted matrix, where nk and mk are the means and the number of items in Dk, respectively. The purpose of a weighted k-means technique is to locate a division represented by * and a symmetrical positive matrix G* fulfilling Equation (8) such that

$j_g(\Delta^*) = \min_\Delta\{j_g(\Delta)\}$ ------- (8)

Multiplying a partition by a weighted matrix G results in a different value for JG(). The weighted matrix must so be normalized. In this inquiry, the G determinant is set at 1, i.e.

$\big(det(G)\big) = 1$ ------- (9)

With fixed G = I in (6), condition (7) is automatically satisfied, and equations (8) and (9) comprise the cost function and optimal aim, respectively, of a standard k-means algorithm.

3.3 Feature selection using weighted binary BAT algorithm

Finding the best answer to a problem is the goal of the optimization technique known as Weighted Binary BAT (WBB). It's a twist on the classic Binary Bat Algorithm (BBA) that uses a weighting factor to adjust the searching strategy of the bats. In the first step of the WBB procedure, a pool of binary strings representing potential answers to the problem is populated. After assessing each solution using a fitness function, the software assigns a weight to each bat based on its fitness score. Better solutions are more likely to be selected for the next generation due to the weighting factor.

The bats then use echolocation to find their way around the search area and zero in on promising leads. The WBB algorithm thoroughly probes the search space by combining random search with local search. Bats can strike a delicate balance between exploration and exploitation by adjusting their frequency and pulse rate during the search. The bats adjust their positions and fitness scores to reflect the best solution throughout each cycle. The weighting factor is also adjusted such that the best solutions still have a higher chance of getting picked.

Unless a stopping criterion is met, such as reaching a maximum number of iterations or finding a solution that meets a predetermined threshold, the WBB algorithm will continue to run.

Echolocation is a unique ability of bats (tiny bats). To determine an item's distance, bats emit a short sound and listen for the echo. The BAT Algorithm, created by Yang (2010) [68], is a revolutionary meta-heuristic optimization technique inspired by the skills of bats (BA). A group of bats will utilize their echolocation abilities to look for nectar or insects. Here, we discuss some idealized ideas based on bats' echolocation properties and behaviour.

$Freq_i = Freq_{min} + (Freq_{max} - Freq_{min}) * \beta$ ------ (10)

$vl_i(t+1) = vl_i(t) + (pos_i(t) - G_{best}) * Freq_i$ ------- (11)

$pos_i(t+1) = pos_i(t) + vl_i(t+1)$ ------- (12)

At the end of each iteration, the frequency of the ith bat, denoted by F reqi, is updated following (12); [0, 1] is a random number between 0 and 1; and Gbest is the optimal solution. The ith bat's speed and position at the tth iteration are denoted by V li(t) and P osi(t), respectively. The BA's exploitability might be guaranteed by Equations 2-4. To increase the potential for exploitation, a random walk method was subsequently implemented (13).

$pos_{new} = pos_{old} + \in * L^t$ ------ (13)

$L_i(t+1) = \alpha * L_i(t)$ ------ (14)

$R_i(t+1) = R_i(0)[1 - e^{-rt}]$ ------ (15)

Where is an arbitrary positive integer between 1 and 1. If the new solution is improving, then the loudness and pulse rate of the ith bat at the tth iteration, L and R, are adjusted accordingly, as shown in Eqs. (14) and (15). Both are predetermined constants.

In the case of the bat method, digital bats may fly across the search space by using position and velocity vectors (or updated position vectors) in the continuous or accurate domain (BA). So, they may include the velocity with the optimal prior position to update the new positions of bats. Yet, in binary or discrete spaces, the location can only be represented by a one or a 0. Hence, updating the location of binary spaces is not the same as doing so for continuous spaces. The BAT binary version (BBA) is similar to the BA algorithm except for the transfer function and position update. Modifying the transfer function is to convert continuous search space to discrete search space (15).

$S\left(Vl_i^k(t)\right) = \frac{1}{1+e^{-vl_i^k(t)}}$ ------- (15)

V lk i (t) is the velocity of bat i at kth dimension on tth iteration and is updated by (15) in each iteration.

3.4 Clustering with the hybrid method as PSO with GKFCM

3.4.1 Particle Swarm Optimization Algorithm

The cooperative behaviours are seen in nature-inspired Particle Swarm Optimization (PSO). It's a method of Meta-heuristic search inspired by the coordinated behaviours of animals as diverse as fish, birds, and insects. It was first suggested in 1995 by social psychologist James Kennedy and electrical engineer Dr. Russell Eberhart. Swarm intelligence (SI) is a term that is widely used to describe it. It is a stochastic global optimization method that uses a population model. It was designed to study how fish, birds, and insects communicate with one another and signal when they have found food. In search of sustenance, it has been observed that animals often travel in packs to seemingly unrelated sites. A PSO strategy strikes a compromise between exploitative and exploratory strategies using global search (using the knowledge of neighbours) and local search (using one's knowledge). Some industries have successfully used PSO to boost efficiency. Some examples of these disciplines include classification, feature selection, prediction analysis, text clustering, multi-model optimization, clustering, association rule mining, data mining, pattern recognition, rule extractor, and image processing.

The PSO draws its fundamental biological inspiration from the idea that people's actions are shaped not just by their own prior experiences but also by those of the people they come into touch with. Since its first release, PSO has been modified to meet the needs of a wide range of applications.

$Plg = (cck_1, cc\ \ k_2, cck_{k\square}, \ldots cckm)$ ------- (16)

In this case, cckm represents the centre of the kth cluster. Position vector xkh represents the cckh. Kth particle hth element location is indicated by xkh. Each particle in the swarm has a distinct movement and position. The article goes in whichever direction its velocity takes it. The Vkh is the V-h vector's component for the kth particle. Particle location (X-vector) and current velocity (V-vector). The particle's updated location is calculated by adding the V-vector to the X-

vector at the end of each generation. In Equation 4.3, rand1 and rand2 are arbitrary numbers. These are used in exhaustive lookups. Inertia's unpredictable value has a role. Particle masses are determined by the parameters c1 and c2 during trajectory design.

$V_{k\square} = \emptyset x v_{k\square} + c_1\, x\, rand_1(Pgbest - x_{g\square})$ ------- (17)

$X_{g\square} = X_{g\square} + V_{g\square}$ -------- (18)

The particle's location is updated with each generation. We use equations 17 and 18 to add the particle's current velocity (vgh) to its location (xgh). As a result, the particle's speed altered. The ideal position of each particle is recorded in the search space. Ppbestg(t) is gthparticle's all-time favourite spot, and it also happens to be its best position. The optimal position of the particle at time step t is computed using equation 18, where f is the empirical function.

$if\left(Ppbest_g(t+1) = Ppbest_g(t)\right)$ -------- (19)

$f(x_g(t+1) \geq f\left(Ppbest_g(t)\right)$ ------- (20)

$else$

$f(x_g(t+1) < f\left(Ppbest_g(t)\right)$ ----- (21)

The most critical choice is Pgbest. It is selected to single out the top particle from the group. The best-known position of the swarm is calculated using Equation 4.8, and "the best" is modified accordingly.

3.4.2 Gaussian Kernel Fuzzy C-Means Clustering Method

There are two parts to the gaussian kernel fuzzy C-means clustering approach: Clustering is used to categorize data sets after the Gaussian Kernel Function transforms the original space X into the higher dimensional space F. The GKFCM employs a non-linear mapping operation in the kernel space to highlight these distinctions. Classification problems involving samples with similar characteristics are well-suited to this method. A non-linear mapping function is defined as follows:

$\theta: x_k \rightarrow \theta(x_k)\epsilon F$ ------- (22)

where xk is the sample of the original space X. The objective function of the clustering algorithm is given by

$j_m(U,v) = \sum_{i=1}^{c}\sum_{k=1}^{n}\mu_{ik}^{m}||\theta\,(x_k) - \theta(v_i)||^2$ ------- (23)

where *n*i is the core point of the original sample space; c is the clustering number; n is the sample

As represented by the kth sample, membership in the ith fault class is denoted by mik. m is a weighted parameter, and mik satisfies mik 2 [0, 1], 0 n k=1 mik n, and c i=1 mik = 1, k = 1, 2, _, n.

The kernel function is defined as follows:

$K(x,y) = \emptyset(x)^T\emptyset(y)$ -------- (24)

Therefore, the Euclidean distance in Equation (25) is given by

$\left|\left|\emptyset(x_k) - \theta(vi)\right|\right| = K(x_k,x_k) + K(v_i,v_i) - 2k(x_k,v_i)$ ------ (25)

Examples of popular kernel functions include the Gaussian kernel function, the polynomial kernel function, and the Sigmoid kernel function. The Gaussian kernel function and its Equation are reported in this analysis (25).

IV. RESULTS AND DISCUSSION

In this section, we present and discuss the results obtained from our proposed approach for predicting bipolar disorder using a combination of complex data preprocessing, feature selection, clustering, and classification algorithms. The approach was implemented using the Python programming language, utilizing a variety of libraries and frameworks such as Pandas, Scikit-learn, and Keras. Furthermore, we present the results of the clustering process, demonstrating the effectiveness of the hybrid approach combining particle swarm optimization (PSO) with Gaussian Kernel fuzzy clustering algorithm (GKFCM) in identifying patterns in the data. We also discuss the implications of these results.

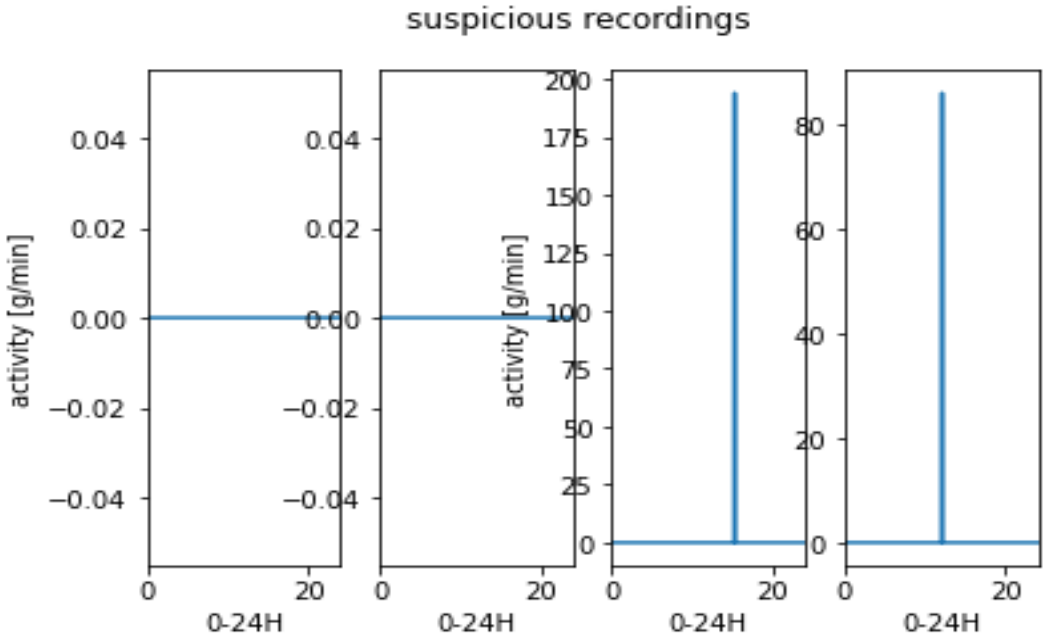

Figure 2: Suspicious recording

Figure 2 shows suspicious recording typically refers to an audio or video recording that raises concerns or doubts about its authenticity, legality, or intentions. Such recordings may be intentionally altered or manipulated to misrepresent or deceive or be recorded secretly or without consent.

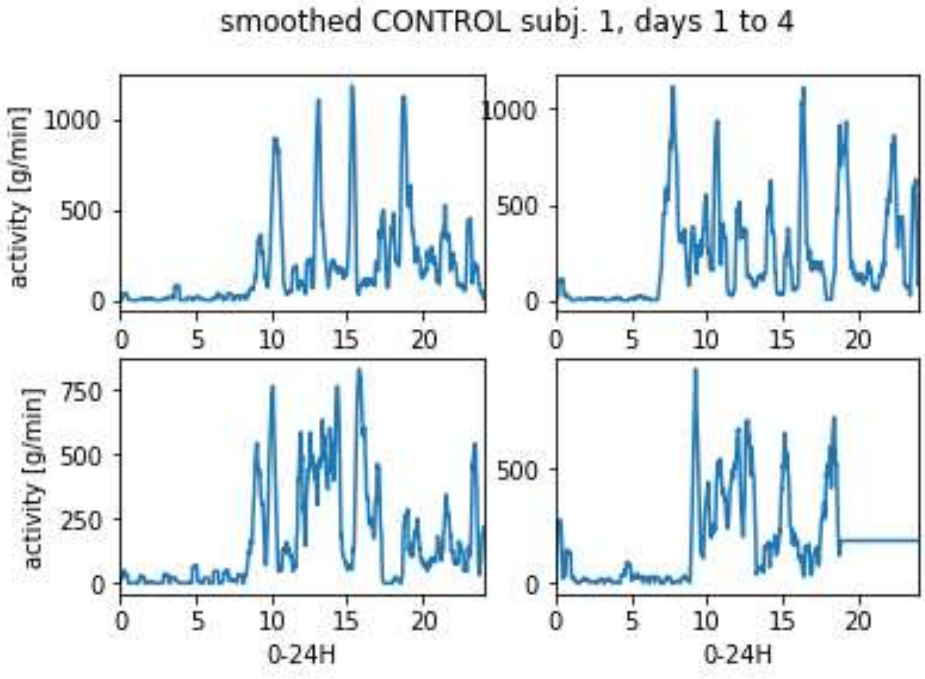

Figure 3: Smoothed control

Figure 4 depicts the Smoothed control approach utilized in bipolar illness to help manage symptoms and regulate mood. It entails taking a steady and controlled approach to medication modification rather than making sudden adjustments that might cause mood swings.

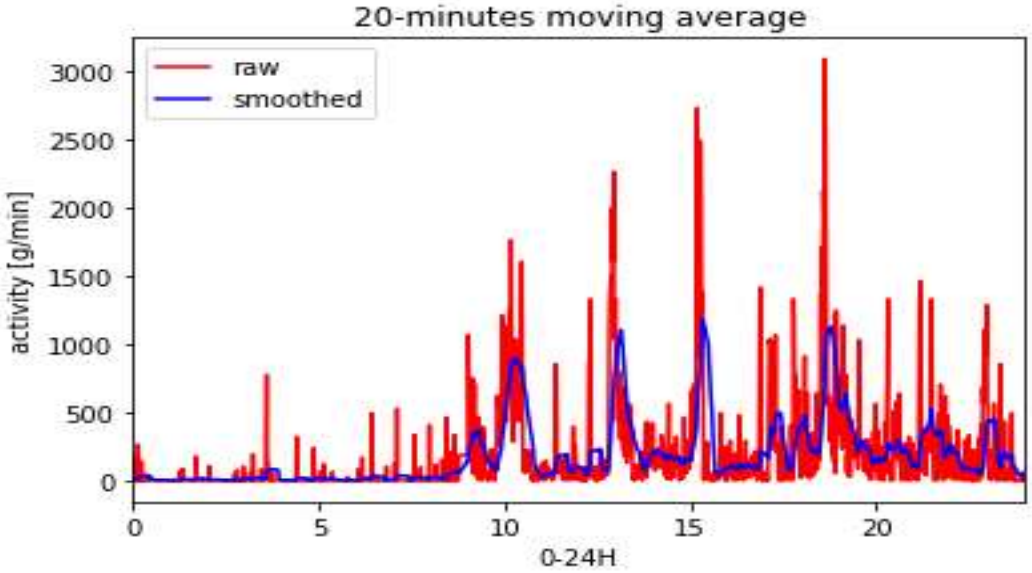

Figure 4: 20 minutes moving average

Moving Average (MA) is a technical analysis method that creates a continually updated average price to smooth out price data. A moving average might be used to measure mood changes over time in the context of bipolar illness.

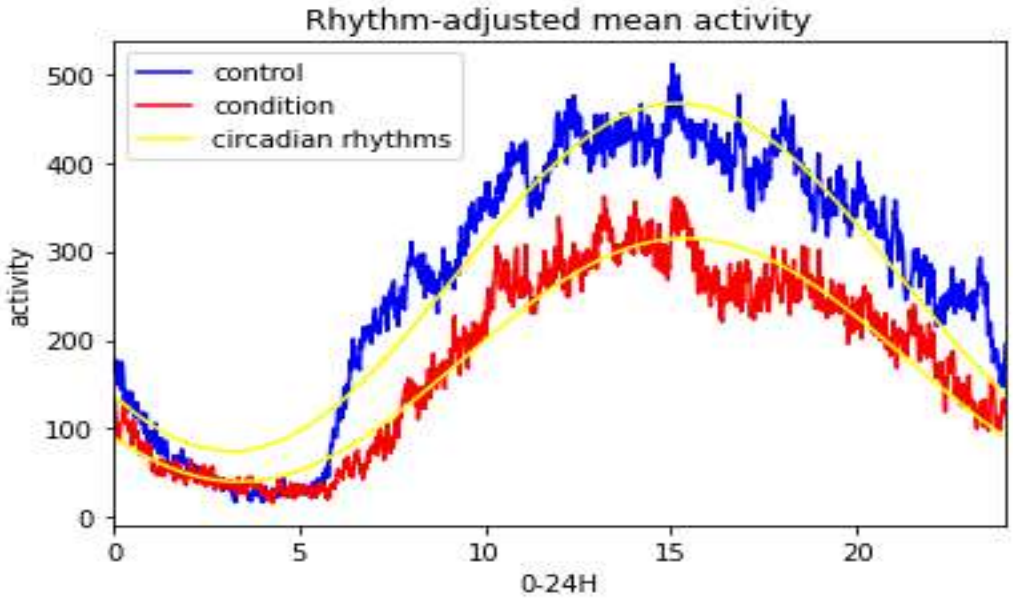

Figure 5: Adjusted mean activity

Adjusted mean activity refers to a statistical method used to estimate the average level of activity or behaviour in a particular group while controlling for the influence of other factors or variables.

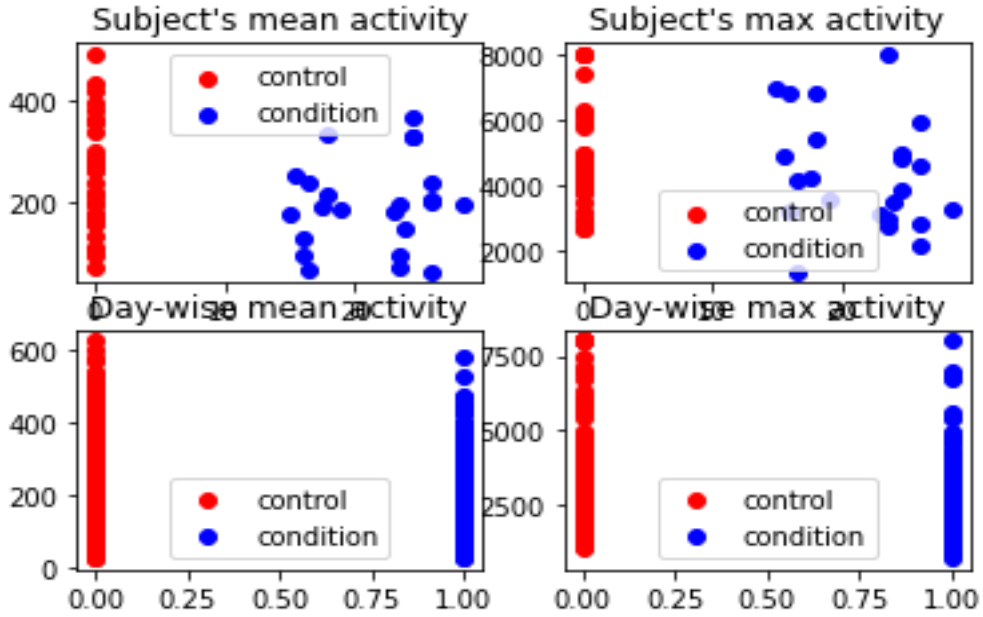

Figure 6: Subject with Day wise mean and max Activity

Figure 6 displays the day-wise mean and maximum activity of a subject. The horizontal axis represents the day of the study, while the vertical axis represents the activity level of the subject. The blue line represents the mean activity level, while the orange line represents the maximum activity level recorded for each day. The graph shows fluctuations in the activity level over the course of the study. The mean activity level shows a general trend of increasing activity over

time, with some fluctuations. On the other hand, the maximum activity level shows a more erratic pattern, with peaks and valleys throughout the study.

Table 1: Clustering accuracy comparison chart

Algorithm	Number of Clusters	Accuracy
K-Means	4	91
K- Medoids	5	90
PSO	5	92
GKFCM	5	93
PSO-GKFCM	5	98.6

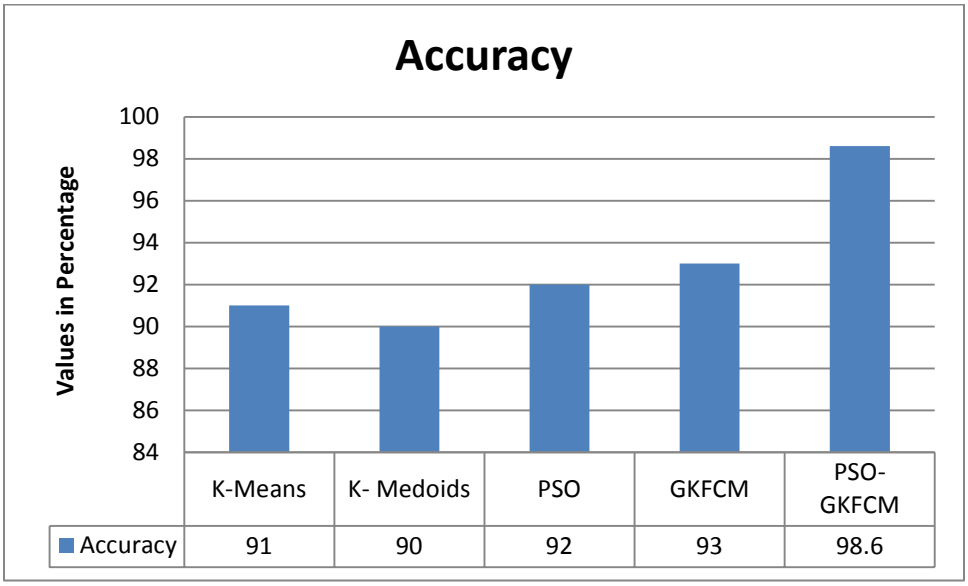

Figure 7: Accuracy Comparison Chart

The table 1 shows the performance of different clustering algorithms for the given dataset. The number of clusters ranges from 4 to 5, and the accuracy of each algorithm is reported. The K-Means algorithm achieved an accuracy of 91%, while the K-Medoids algorithm resulted in an accuracy of 90%. The PSO algorithm produced an accuracy of 92%, and the GKFCM algorithm resulted in an accuracy of 93%. Finally, the proposed hybrid approach, PSO-GKFCM, achieved the highest accuracy of 98.6%. This suggests that the integrated approach is more effective in clustering the data accurately than individual algorithms. The results indicate that the proposed PSO-GKFCM algorithm is a promising technique for clustering bipolar disorder data and can aid in the prediction and diagnosis of the disease with high accuracy and the clustering accuracy is represented at figure 7.

V. CONCLUSION

In conclusion, the proposed method for efficient bipolar disorder prediction using a combination of various algorithms and techniques has demonstrated promising results. The data preprocessing and missing value analysis using an improved decision tree and weighted k-means algorithm helped to enhance the data quality. The weighted binary BAT algorithm was utilized to select essential features that are critical for accurate bipolar disorder prediction. The data were

clustered with a hybrid PSO and GKFCM method to group similar data points. This integrated approach offers a robust and efficient solution for predicting bipolar disorder in patients, enabling early diagnosis and treatment. The proposed method has the potential to be extended and applied to other disease prediction tasks, providing a more comprehensive and accurate patient diagnosis. Python was used for implementing the proposed method. Overall, the proposed method has the potential to significantly enhance the efficiency and accuracy of disease prediction systems. For further to improve the classification accuracy to using the Probability Density Function (PDF) with Recurrent Neural Network (RNN) algorithm.

VI. REFERENCE

1. Borges-Junior, R., Salvini, R., Nierenberg, A. A., Sachs, G. S., Lafer, B., & Dias, R. S. (2018). Forecasting depressive relapse in Bipolar Disorder from clinical data. 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). doi:10.1109/bibm.2018.8621255
2. Chen, Y.-L., Tu, P.-C., Huang, T.-H., Bai, Y.-M., Su, T.-P., Chen, M.-H., & Wu, Y.-T. (2020). Using Minimal-Redundant and Maximal-Relevant Whole-Brain Functional Connectivity to Classify Bipolar Disorder. Frontiers in Neuroscience, 14. doi:10.3389/fnins.2020.563368
3. He, H., Yu, Q., Du, Y., Vergara, V., Victor, T. A., Drevets, W. C., … Calhoun, V. D. (2016). Resting-state functional network connectivity in prefrontal regions differs between unmedicated patients with bipolar and major depressive disorders. Journal of Affective Disorders, 190, 483–493. doi:10.1016/j.jad.2015.10.042
4. Huang, K.-Y., Wu, C.-H., Su, M.-H., & Kuo, Y.-T. (2018). Detecting Unipolar and Bipolar Depressive Disorders from Elicited Speech Responses Using Latent Affective Structure Model. IEEE Transactions on Affective Computing, 1–1. doi:10.1109/taffc.2018.2803178
5. Jie, N.-F., Zhu, M.-H., Ma, X.-Y., Osuch, E. A., Wammes, M., Theberge, J., … Calhoun, V. D. (2015). Discriminating Bipolar Disorder From Major Depression Based on SVM-FoBa: Efficient Feature Selection With Multimodal Brain Imaging Data. IEEE Transactions on Autonomous Mental Development, 7(4), 320–331. doi:10.1109/tamd.2015.2440298
6. Laksshman, S., Bhat, R. R., Viswanath, V., & Li, X. (2017). DeepBipolar: Identifying genomic mutations for bipolar disorder via deep learning. Human Mutation, 38(9), 1217–1224. doi:10.1002/humu.23272
7. Librenza-Garcia, D., Kotzian, B. J., Yang, J., Mwangi, B., Cao, B., Pereira Lima, L. N., … Passos, I. C. (2017). The impact of machine learning techniques in the study of bipolar disorder: A systematic review. Neuroscience & Biobehavioral Reviews, 80, 538–554. doi:10.1016/j.neubiorev.2017.07.004
8. Lobentanzer, S., Hanin, G., Klein, J., & Soreq, H. (2019). Integrative Transcriptomics Reveals Sexually Dimorphic Control of the Cholinergic/Neurokine Interface in

Schizophrenia and Bipolar Disorder. Cell Reports, 29(3), 764–777.e5. doi:10.1016/j.celrep.2019.09.017

9. Mansur, R. B., Lee, Y., McIntyre, R. S., & Brietzke, E. (2020). WHAT IS BIPOLAR DISORDER? A DISEASE MODEL OF DYSREGULATED ENERGY EXPENDITURE. Neuroscience & Biobehavioral Reviews. doi:10.1016/j.neubiorev.2020.04.006
10. Nunes, A., Schnack, H. G., Ching, C. R. K., Agartz, I., Akudjedu, T. N., … Bauer, J. (2018). Using structural MRI to identify bipolar disorders – 13 site machine learning study in 3020 individuals from the ENIGMA Bipolar Disorders Working Group. Molecular Psychiatry. doi:10.1038/s41380-018-0228-9
11. O'Shea, K. S., & McInnis, M. G. (2016). Neurodevelopmental origins of bipolar disorder: iPSC models. Molecular and Cellular Neuroscience, 73, 63–83. doi:10.1016/j.mcn.2015.11.006
12. Rotenberg, L. de S., Borges-Júnior, R. G., Lafer, B., Salvini, R., & Dias, R. da S. (2021). Exploring machine learning to predict depressive relapses of bipolar disorder patients. Journal of Affective Disorders, 295, 681–687. doi:10.1016/j.jad.2021.08.127
13. Schwarz, E., Doan, N. T., Pergola, G., Westlye, L. T., Kaufmann, T., … Meyer-Lindenberg, A. (2019). Reproducible grey matter patterns index a multivariate, global alteration of brain structure in schizophrenia and bipolar disorder. Translational Psychiatry, 9(1). doi:10.1038/s41398-018-0225-4
14. Tekin Erguzel, T., Tas, C., & Cebi, M. (2015). A wrapper-based approach for feature selection and classification of major depressive disorder–bipolar disorders. Computers in Biology and Medicine, 64, 127–137. doi:10.1016/j.compbiomed.2015.06.021
15. Wu, M.-J., Mwangi, B., Bauer, I. E., Passos, I. C., Sanches, M., Zunta-Soares, G. B., … Soares, J. C. (2017). Identification and individualized prediction of clinical phenotypes in bipolar disorders using neurocognitive data, neuroimaging scans and machine learning. NeuroImage, 145, 254–264. doi:10.1016/j.neuroimage.2016.02.016
16. Yatham, L. N., Kennedy, S. H., Parikh, S. V., Schaffer, A., Bond, D. J., Frey, B. N., … Berk, M. (2018). Canadian Network for Mood and Anxiety Treatments (CANMAT) and International Society for Bipolar Disorders (ISBD) 2018 guidelines for the management of patients with bipolar disorder. Bipolar Disorders, 20(2), 97–170. doi:10.1111/bdi.12609