DOI: 10.53992/njns.v10i3.272 NUST Journal of Natural Sciences, Vol. 10, Issue 3, 2025

Research Article

A computational approach for correlational analysis of symptoms of major depressive disorder

Sahar Safdar[1], Abdur Rauf[2], Aneeza Zaheer[1], Zeenat Fatima[1], Syed Haseeb Ahmad Shah[1], Syeda Marriam Bakhtiar[1*]

[1]Genetic and Molecular Epidemiology Research Group, Department of Bioinformatics and Biosciences, Capital University of Science and Technology, Islamabad, Pakistan.
[2] National University of Science and Technology (NUST), Islamabad, Pakistan.

***Corresponding author's email:** marriam@cust.edu.pk

Abstract

Major Depressive Disorder (MDD) is a complicated mental illness that consists of wide range of correlated symptoms. According to DSM-V, it is characterized by pervasive low mood, loss of interest or pleasure in nearly all activities, and additional symptoms that can cause significant distress in social, occupational, and important areas of life. In this study, association rules, decision tree classification and agglomerative clustering are employed to classify MDD symptoms interconnection and their co-occurrence pattern. A combination of key symptoms occurrence, such as aggression with euphoric responses and overthinking with euphoria are identified through association rules that show highest lift values. Decision tree is employed to predict primary node which is mood swing as key predictor of MDD. Then, agglomerative clustering is used to split the dataset into three clusters based on expert diagnosis to identify the range of symptoms that overlap. In this study, computational approach is utilized to unravel the hidden pattern and overlapping of correlated symptoms that will help in improved diagnosis and better personalized treatment plans. This study highlighted the interrelationships of symptoms associated with MDD and their thorough examination for therapeutic approach. Future perspectives should focus on diverse datasets with extension and validation of findings that produce sustainable findings for clinical decision making. By the complex interplay of symptoms of MDD shows the contribution of this research towards advancement of evidence-based diagnostics, ultimately aim to improve clinical outcomes by intergradational symptomatic study.

Keywords: Machine Learning Algorithms, Major Depressive Disorder (MDD), Association Rule, Decision Tree, Agglomerative Clustering.

Article History: **Received:** 06 Dec 2025, **Revised:** 02 Jul 2025, **Accepted**: 11 Sep 2025, **Published:** 18 Nov 2025.

Introduction

With over 300 million cases globally, Major Depressive Disorder (MDD) is a multifaceted mental illness [1]. It is the third most frequent cause of illness worldwide, with significant public health threat and economic burden as well [2].

After Covid-19 findings reveal MDD cases are increasing considerably [3]. Clinical Depression leads to severe mood swings characterized by continuously feeling sadness, hopelessness, and loss of interest from daily activities is also referred to as MDD. It interferes with everyday tasks and psychological processes and is marked by persistent grief, loss of interest or happiness, low energy, worse sleeping and eating habits, and even suicidal thoughts [2]. Individuals suffering from major depressive disorder (MDD) are more susceptible to diabetes, stroke, heart disease, obesity, cancer, cognitive decline, and Alzheimer's disease[3]. Additionally, MDD patients' social lives and quality of life are severely compromised. Morbidity and mortality are increased by MDD's potential for recurrence.

Although the prevalence of MDD varies widely between nations and regions, it is rising worldwide every year. In Chinese adults, the lifetime prevalence of depression is 6.8%, with 3.4% of cases being major depressive disorder (MDD). In adults, the lifetime prevalence of MDD was 11.3% in European adults and 20.6% in American people. According to the WHO, MDD accounted for the third-highest share of the world's illness burden in 2008 and was expected to rise to the top by 2030 [4]. However, despite being prevalent, MDD has an upsetting degree of discrepancy and heterogeneity in its diagnosis and assessment. The current DSM-5 diagnostic criteria include nine symptoms that define MDD. When five or more symptoms are present, they satisfy the DSM-5 diagnostic criteria for major depressive disorders which are as shown in Table 1.

However, there is some level of overlap between the symptoms of other neuropsychiatric disorders, and the lack of unique MDD biomarkers for correct diagnosis and evaluation can result in bias and misinterpretation, which can impact the course and prognosis of the disease. While approximately three-quarters of MDD patients recovered within a year of therapy and over half recovered within six months, 27.8% of MDD patients did not improve and went on to develop chronic depression [4]. Furthermore, the idea that MDD is only the accumulation of its symptoms suggests that the disorder's underlying symptoms develop uniformly. This would imply that each symptom is not significant in and of itself, and that all symptoms are equally meaningful [3].

Table 1. Diagnostic criteria of DSM-V

DSM-V Symptoms of MDD	
S/No	**Symptoms**
1	Depression (mood),
2	Anhedonia (loss of interest or enjoyment),
3	Decrease in appetite or weight,
4	Hypersomnia or Insomnia (sleep),
5	Agitation or retardation of the psychomotor system (psychomotor),
6	Energy loss or exhaustion (fatigue),
7	Depressive or unjustified feelings of shame (worthlessness),
8	Deficit in focus or difficulty making decisions (focus);
9	Suicidal ideas, plans, or thoughts of suicide (suicidal thoughts and acts)

One of the main goals in mental health research is the prediction of depression, as this is a critical field of study that can lead to improved treatment success rates and early intervention. Because self-reported experiences by patients may be portrayed in a lower-dimensional network form, which provides an advantage in visualizing the interacting properties of depression-related feelings, they constitute a promising biomarker for depression prediction[5]. The study provides a comprehensive examination of the existing literature on network analyses related to Major Depressive Disorder (MDD). The studies reveal that depressed mood and fatigue emerged as significant symptoms through network approach. Through robust connections, these symptoms formed pattern that shows mood anhedonia as most frequent symptom [3]. The study also revealed that the life quality of individuals with MDD is also affected by painful somatic symptoms. Patients faced great severity. The relationship between emotional and physical symptoms relationship should be considered by clinicians while treating patients for effective treatment [6]. The literature survey also revealed that MDD is also associated with comorbidities. There is a cognitive decline in patients with an increased incidence of dementia and Alzheimer's disease as well. The chances of cardiovascular diseases and chances of strokes become higher and worsening. There is a notably incidence of obesity and severity of autoimmune diseases along with HIV/AIDS and many more [7].

Artificial Intelligence (AI) has emerged as a powerful tool for predicting co-existing symptoms of Major Depressive Disorder (MDD), enabling early intervention and improved treatment outcomes. Various studies have demonstrated the effectiveness of different ML algorithms in identifying risk factors and predicting MDD in diverse populations. Artificial intelligence (AI) is defined as the computer software that mimics human cognitive processes. It includes a variety of technological advancements that let robots to carry out operations that traditionally require human intelligence, like comprehending natural language, spotting patterns, and making clinical decisions. A subset of AI called machine learning (ML) involves learning abilities of computers by using algorithms to examine data, find patterns and make decisions. There are various forms of ML used in healthcare [8]. Using Pearson's correlation-based feature selection, the study ranked attributes based on their importance for predicting depressive disorders. Key predictors identified for bipolar disorder included temperature, atmospheric pressure, season, and ozone, while factors like wind speed and humidity were ranked lower, indicating weaker correlations [9].

The existing literature on Major Depressive Disorder (MDD) highlights several limitations that create a significant research gap in understanding the disorder's symptomatic relationships. In most studies, diagnosis of MDD is based on isolated symptoms that predict binary outcomes which are presence or absence of the disorder. This overlooks the interrelation between symptoms co-occurrence while with the use of machine learning algorithms these hidden relationships and patterns can be predicted. Machine learning algorithms including association rules, decision tree classification and agglomerative cluttering are very useful in the scenario.

This study addresses the utilization of association rule mining, decision tree classification and agglomerative clustering to classify and analyze the co-relation of symptoms of MDD. Through this approach, this research aims to provide insight based on data exploring interrelated patterns of symptoms that help in early diagnosis and personalized therapeutics strategies.

Methods

This study explores the hidden pattern symptoms for their classification and identification of co-occurrence of multiple symptoms for symptomatic relationships in MDD using association rules, decision tree and agglomerative clustering.

Association rule

In the field of MLA, association rules are termed as the data mining algorithm that is used to identify intricated patterns and relationships within the variables of dataset. "If-then" rule is employed for the identification of items that occur frequently and their association by exploring dependencies and correlation. It is helpful in analyzing categorical and qualitative attributes on the dataset. If it is the antecedent that proceeds, then which is consequent following specific pattern.

The association rule can be validated through support, confidence and lift. Support is the measure of items percentage in given dataset. It can be calculated as:

1. $Support(X \rightarrow Y) = P(X \cup Y)$

Here, in the rule, X→Y, antecedent (If) is X while consequent (Then) is Y.

Confidence is the measure of conditional probability of item X inclusion led to inclusion Y.

2. $Confidence(X \rightarrow Y) = P(Y|X)$

While lift is the measure of association rule performance. It can be calculated as[10].

3. $Lift(X \rightarrow Y) = Confidence(X \rightarrow Y)/ Support\ (X \rightarrow Y$

Decision tree

It is a serial framework that integrate many fundamental test with numerical attribute of each test of specific threshold in coherent and efficient manner [11].

Entropy

For the quantification of randomness and impurity in a dataset, entropy is employed. It always lies between "0" and "1" such that the closer it is to 0, the better quantification will be [11]. The general equation for the *Entropy(S)* of dataset containing c = classes number and p_i = probability of class i in dataset can be calculated as [12].

$$4. Entropy(S) = -\sum_{i=1}^{c} (p_i * \log_2\ p_i)$$

Information gain

One segmentation metric that is frequently employed is information gain, also known as mutual information. It helps in determining the random variable values. According to "equation (5)," which defines entropy, Gain (S, A) is defined as follows.

$$5. Gain\ (S, A) = \sum_{v \in V(A)} \frac{|S_v|}{|S|}\ Entropy\ (S_v)$$

Here, "S_v" is Set S subset to v attribute value while A's range equals $V(A)$ [11].

The Gini index serves as a purity or impurity indicator when constructing a tree with the classification and regression tree (CART) technique. Selecting an attribute with a low Gini index is preferable than one with a high index. Equation 15 can be used for the calculation of Gini index in which number of classes is c and probability is [13].

$$6.\ Gini\ Index = 1 - \sum_{i=1}^{c} (p_i)^2$$

Agglomerative clustering

A bottom-up approach for hierarchical clustering, agglomerative clustering starts with each data point as its own cluster. After that, the algorithm repeatedly combines the closest clusters according to the linkage methods until either a predetermined number of clusters is reached, or all data points are clustered into a single cluster. The hierarchy of cluster mergers is represented by a dendrogram, which is commonly used to depict this process [14].

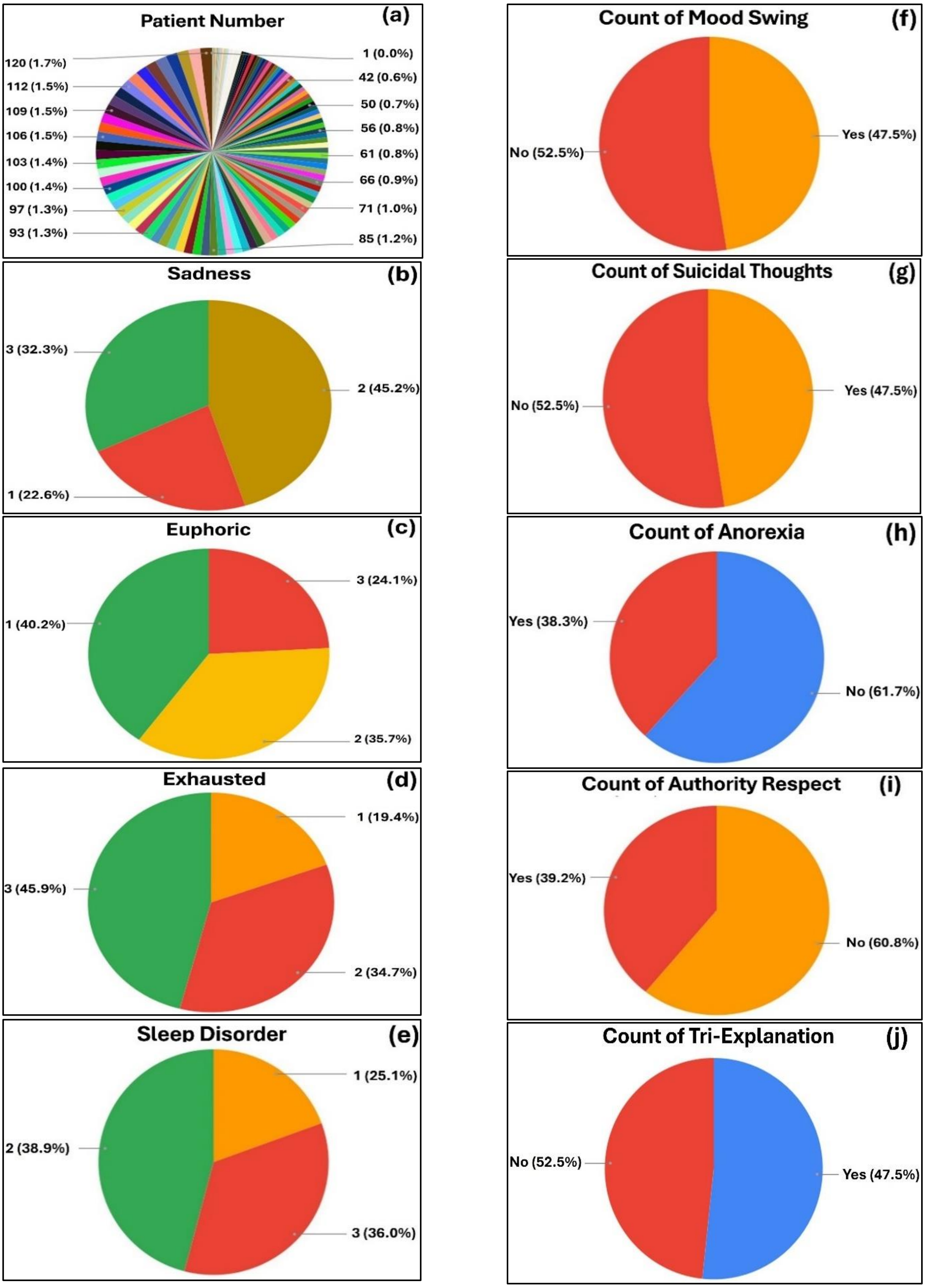
Results

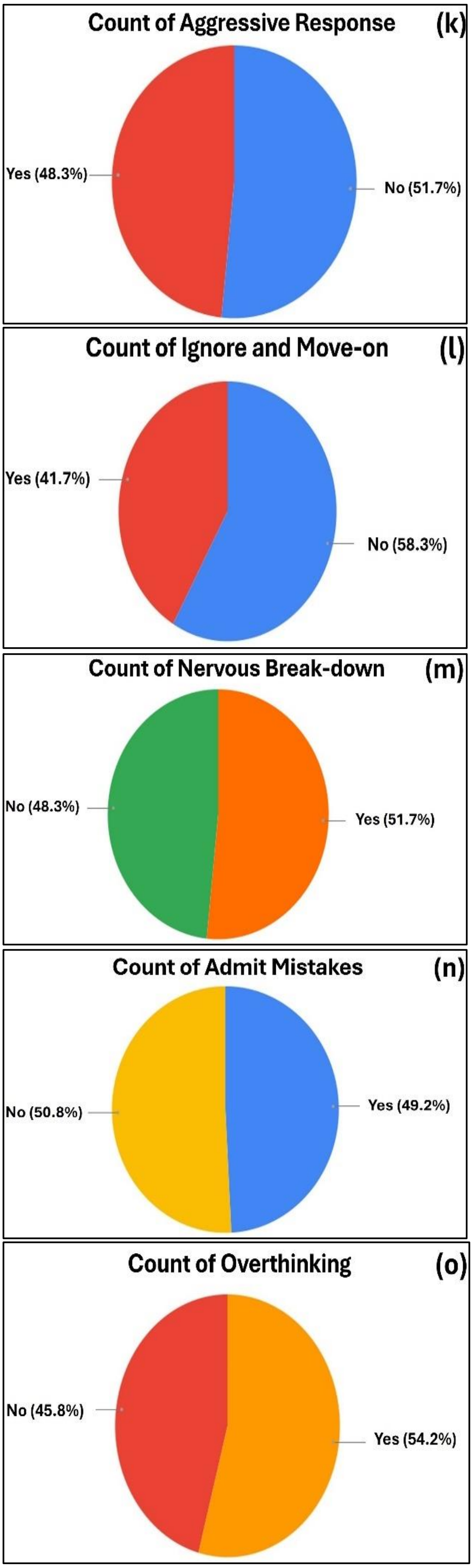
Dataset related to the symptoms of Major Depressive Disorder is retrieved from Kaggle. Initially, the data is downloaded and saved in excel format. The dataset consists of 19 attributes which are Patient Number, Sadness, Euphoric, Exhausted, Sleep disorder, Mood Swing, Suicidal thoughts, Anorexia, Authority Respect, Try-Explanation, Aggressive Response, Ignore & Move-On, Nervous Breakdown, Admit Mistakes, Overthinking, Sexual Activity, Concentration, Optimism and Expert Diagnose of 120 patients with varying values. The data type of dataset is shown in Table 2. Important trends, patterns, and variability in the dataset's features are revealed by the graphical distribution of its constituent variables.

Table 2: Dataset attributes their data type and description.

S/No	Attributes	Data Type	Characteristics
1	Patient Number	Integer	Unique identifier to each patient (1 to 120)
2	Sadness	Continuous	Represent severity of sadness (Seldom=0, Sometimes=1, Usually=2, Most often=3)
3	Euphoric	Continuous	Represent level of euphoria (Seldom=0, Sometimes=1, Usually=2, Most often=3)
4	Exhausted	Continuous	Represent level of exhaustion (Seldom=0, Sometimes=1, Usually=2, Most often=3)
5	Sleep disorder	Continuous	Represent level of sleep disorder (Seldom=0, Sometimes=1, Usually=2, Most often=3)
6	Mood Swing	Boolean	Presence/ Absence of Mood swing (Yes/No)
7	Suicidal thoughts	Boolean	Presence/ Absence of Suicidal thoughts (Yes/No)
8	Anorexia	Boolean	Presence/ Absence of Anorexia (Yes/No)
9	Authority Respect	Boolean	Presence/ Absence of Authority Respect (Yes/No)
10	Try-Explanation	Boolean	Presence/ Absence of Try-Explanation (Yes/No)
11	Aggressive Response	Boolean	Presence/ Absence of Aggressive Response (Yes/No)
12	Ignore & Move-On	Boolean	Presence/ Absence of Ignore & Move-On (Yes/No)
13	Nervous Break-down	Boolean	Presence/ Absence of Nervous Break-down (Yes/No)
14	Admit Mistakes	Boolean	Presence/ Absence of Admit Mistakes (Yes/No)
15	Overthinking	Boolean	Presence/ Absence of Overthinking (Yes/No)
16	Sexual Activity	Integer	Measure the frequency of sexual activity on 1-10 scale
17	Concentration	Integer	Measure the frequency of Concentration on 1-10 scale
18	Optimism	Integer	Measure the frequency of Optimism on 1-10 scale
19	Expert Diagnose	Categorical	Experts classify each patient based on symptoms (Normal, Bipolar type 1, Bipolar type 2 and Depression).

This reveals dataset insights that offer a strong basis for applying machine learning models to accurately forecast cardiac illnesses. The distribution of attributes among datasets is presented as pie charts as shown in Figure 1.





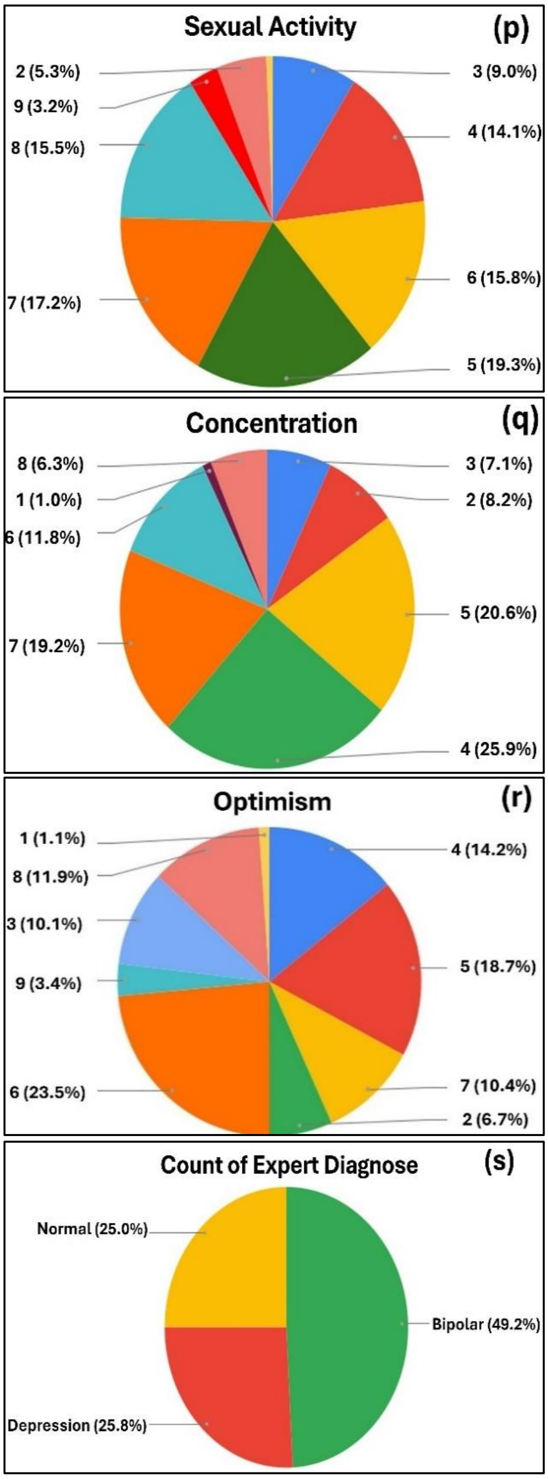


Figure 1: Distribution of attributes of Dataset. (a) Patients number (b) Sadness (c) Euphoric (d) Exhausted (e) sleep disorder (f) Mood swing (g) Suicidal thoughts (h) Anorexia (i) Authority respect (j) Try-Explanation (k) Aggressive response (l) Ignore & move-on (m) Nervous breakdown (n) Admit mistakes (o) Overthinking (p) Sexual activity (q) Concentration (r) Optimism (s) Expert diagnosis.

Through association rules with the highest lift values are used to identify relationships between symptoms and diagnoses related to Major Depressive Disorder (MDD). In association rule mining, the lift value indicates how much more likely the occurrence of the consequent is given the antecedent compared to random chance. A lift value greater than 1 implies a positive association between the items in the rule. Co-symptoms and their lifts as predicted by models are shown in Table 3.

Table 3: Top 5 highest lift with association rule.

S/No	Co-Symptoms	Lift
1	Aggressive Response & Euphoric	≈ 3.0
2	Expert Diagnosis with Overthinking and Euphoric	≈2.8
3	Expert Diagnosis with Mood Swing and Euphoric	≈ 2.7
4	Aggressive Response & Exhausted	≈ 2.6
5	Suicidal Thoughts & Nervous Breakdown	≈ 2.5

Aggressive response & euphoric (Lift ≈ 3.0)

Individuals exhibiting aggressive responses are about three times more likely to also show euphoric symptoms. This might indicate emotional instability, which can be observed in certain subtypes or phases of mood disorders, including bipolar aspects that may coexist with MDD.

Expert diagnosis with overthinking and euphoric (Lift ≈ 2.8)

Patients diagnosed by an expert who show overthinking tendencies are also likely to exhibit euphoric behavior. This combination might reflect underlying thought patterns where rapid or excessive thinking coexists with mood elevation, potentially relevant in mood disorder contexts within MDD.

Expert diagnosis with mood swing and euphoric (Lift ≈ 2.7)

The co-occurrence of mood swings and euphoric symptoms in diagnosed patients highlight mood instability as a significant trait, which can manifest in individuals with severe or atypical MDD.

Aggressive response & exhausted (Lift ≈ 2.6)

The combination of aggressive behavior and exhaustion indicates that individuals with MDD might experience heightened irritability alongside fatigue, both common symptoms in depression.

Suicidal thoughts & nervous breakdown (Lift ≈ 2.5)

This rule indicates a strong association between suicidal thoughts and nervous breakdowns. The co-occurrence of these severe symptoms may highlight critical cases requiring immediate intervention, as they are strongly predictive of severe depression.

The scatter plot represents the relationship between support and confidence for association rules derived from a dataset, aiming to analyze symptom co-occurrence in Major Depressive Disorder (MDD) as shown in Figure 2. The X-axis indicates support, which reflects the frequency of symptom combinations in the dataset, while the Y-axis shows confidence, representing the likelihood that one set of symptoms leads to another. Each point corresponds to a specific association rule, and the intensity of its color denotes the lift, a measure of the rule's significance in terms of how much more likely symptoms occur together than independently.

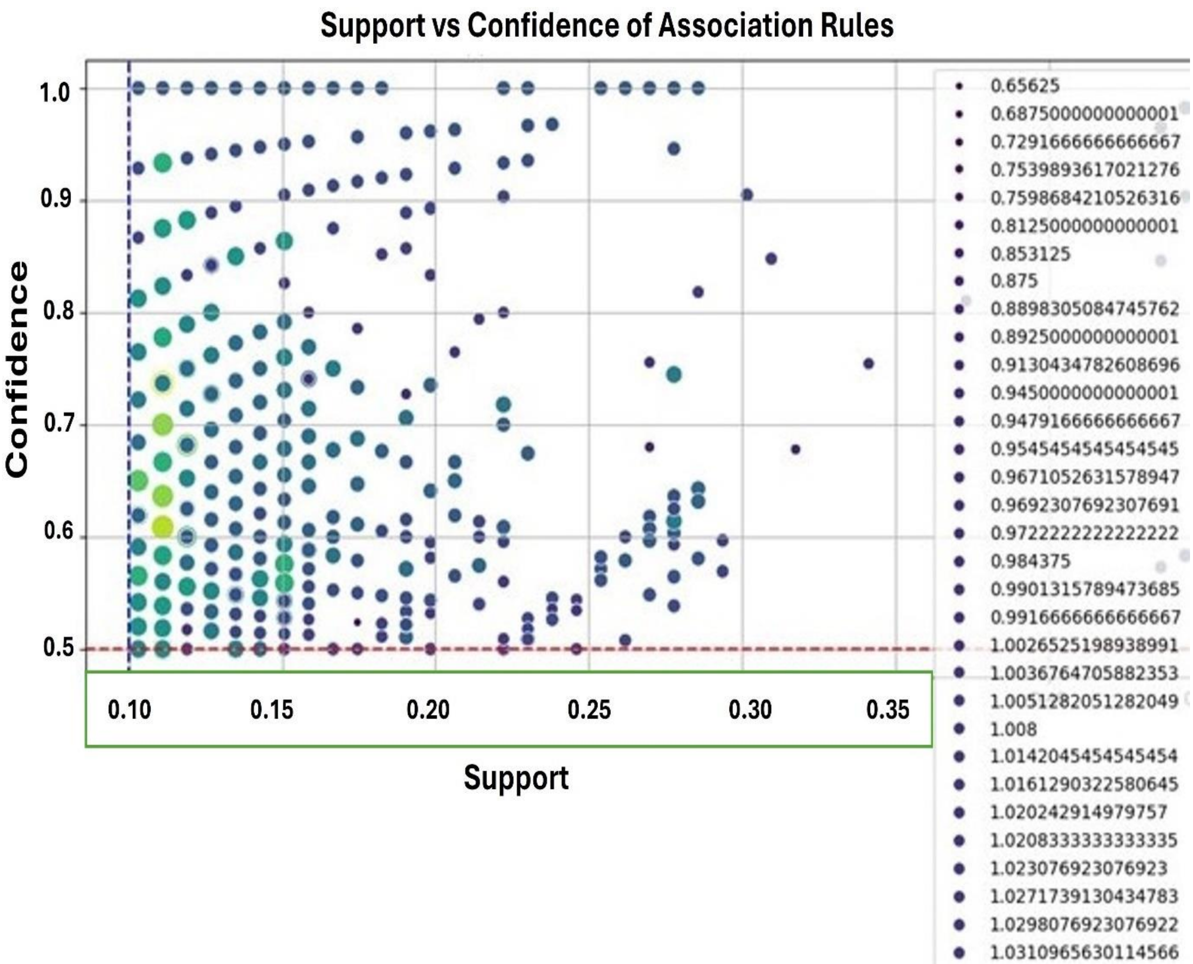


Figure 2: Confidence vs support by association rule

The plot highlights two thresholds: a confidence threshold marked by a red dashed line at 0.5, where rules below this line are considered less reliable, and a minimum support threshold shown as a blue dashed line at approximately 0.1, indicating infrequent symptom combinations that may have limited significance. Most association rules with high confidence cluster within the support range of 0.1 to 0.35, demonstrating that frequently occurring symptoms are more likely to produce reliable predictions. Additionally, darker points with high lift suggest impactful symptom relationships, while lower-support rules exhibit variable confidence, indicating that rare symptom combinations may still be predictive under certain conditions.

The dataset is classified through decision trees which determine the diagnosis probability as shown in Figure. 3. There is a primary differentiator at the root node which is the presence or absence of mood swings which is the significant indicator of MDD. The left path is followed if the mood swing is absent (≤ 0.5), this shows higher certainty in classification because Gini Index is lower. This pathway is further subdivided with individuals with lower or absent sexual activity (≤ 0.5) which is further evaluated based on optimism. When the Gini Index is 0.059, it shows high certainty in which lower level of optimism (≤ 0.5) strongly related with the diagnosis. On the other hand, higher level of optimism (> 0.5) shows higher uncertainty which would lead to further splits into nervous breakdown and euphoria. Cases such as where nervous breakdowns are absent (≤ 0.5) show confirmation of expert diagnosis while no diagnosis is shown when nervous break.

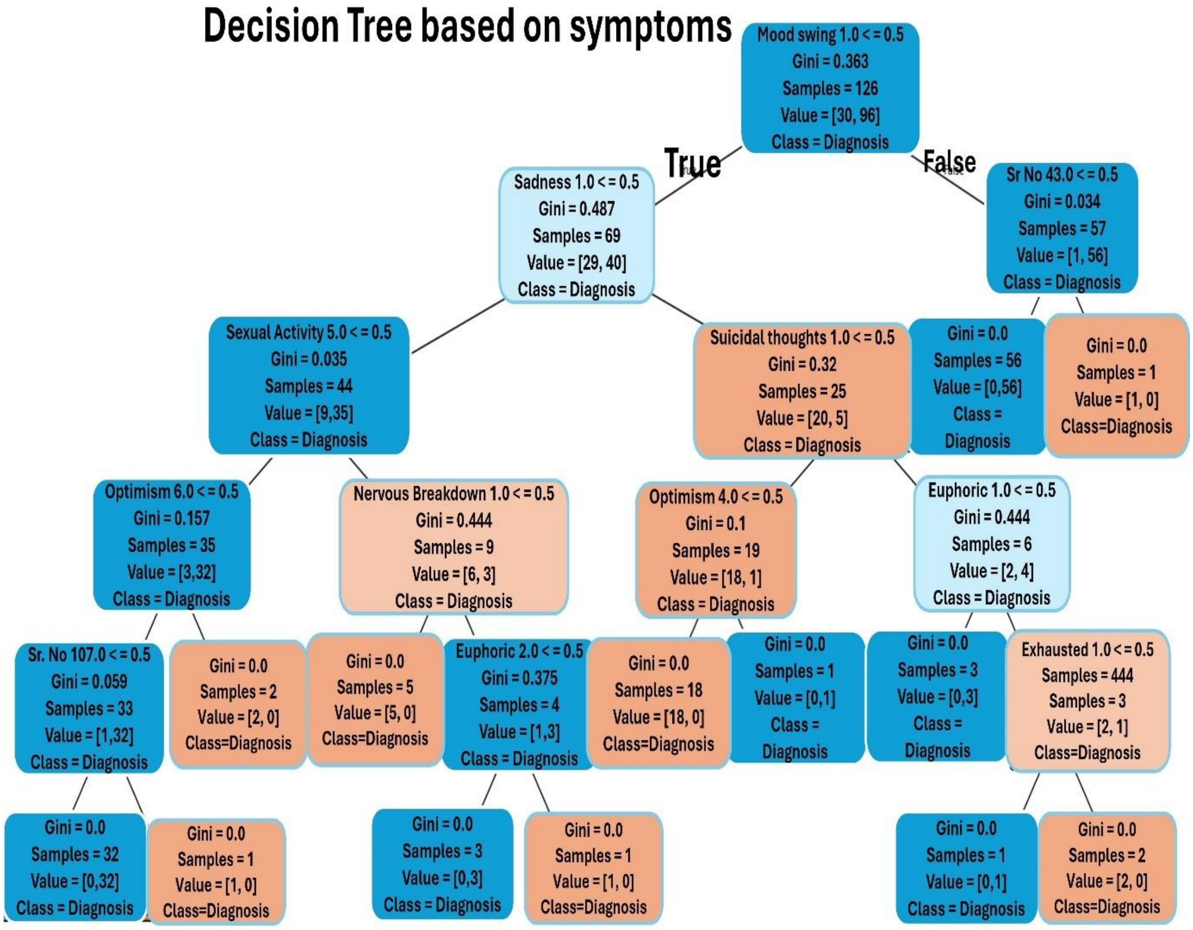


Figure 3. Decision tree based on symptoms.

Conversely, higher levels of optimism (> 0.5) introduce greater uncertainty, leading to further splits based on symptoms like nervous breakdowns and euphoria. In cases where nervous breakdowns are absent (≤ 0.5), a diagnosis is more likely, while their presence (> 0.5) often suggests no diagnosis. Through a decision tree, if mood swing is prominent, further classification relies on sexual activity and sadness. The left branch indicates low level of sexual activity together with reduced optimism while nervous breakdown is strongly related with MDD diagnosis. On the other hand, absence of sadness, suicidal thoughts and euphoria on right branch suggests no diagnosis while classification in influenced by their presence. The Gini index measures nodes' purity. Its low value corresponds to more consistent classification. The mood swings, sadness and optimism are strong predictors for MDD.

Agglomerative clustering splits the dataset into three clusters based on attribute of expert diagnosis from dataset as shown in Table 4.

Table 4: Clusters based on expert Diagnosis through agglomerative clustering.

Cluster	Expert Diagnose	Proportion
0	Bipolar	0.516129
	Depression	0.435484
	Normal	0.048387
1	Normal	0.757576
	Bipolar	0.151515
	Depression	0.090909
2	Bipolar	0.88
	Normal	0.08
	Depression	0.04

People with bipolar disorders, depression, and normal appear to be mixed in Cluster 0. This could suggest that the algorithm had trouble distinguishing between these circumstances in this specific cluster. People with a diagnosis of Normal make up the majority in Cluster 1, indicating that the algorithm successfully identified this group. Since Cluster 2 has a strong correlation with bipolar disorder, the algorithm was able to correctly detect this condition.

The dendrogram shows the outcomes of hierarchical agglomerative clustering, in which Euclidean distances are used to categorize patients according to similarity as shown in Figure 4. Each patient is initially grouped into a separate cluster in the hierarchical structure, which then gradually combines them into larger clusters according to how similar they are until all the patients are part of a single cluster at the top. The dissimilarity threshold at which clusters merge is represented by the y-axis (Euclidean distances), where increased similarity is indicated by smaller distances. As seen by the color coding (orange, green, and red), the diagram indicates that the dataset naturally forms three different clusters. The ideal number of clusters is found by cutting the dendrogram at a height of about 10 to 15. These clusters are subsets of patients with comparable patterns. Clusters' dissimilarity is reflected in the height at which they merge Clusters that merge at higher elevations (larger distances) are less comparable than those that combine at lower elevations. Potential patterns or groups in the data that could support a diagnosis, treatment planning, or additional exploratory examination of patient features can be found with the use of this technique.

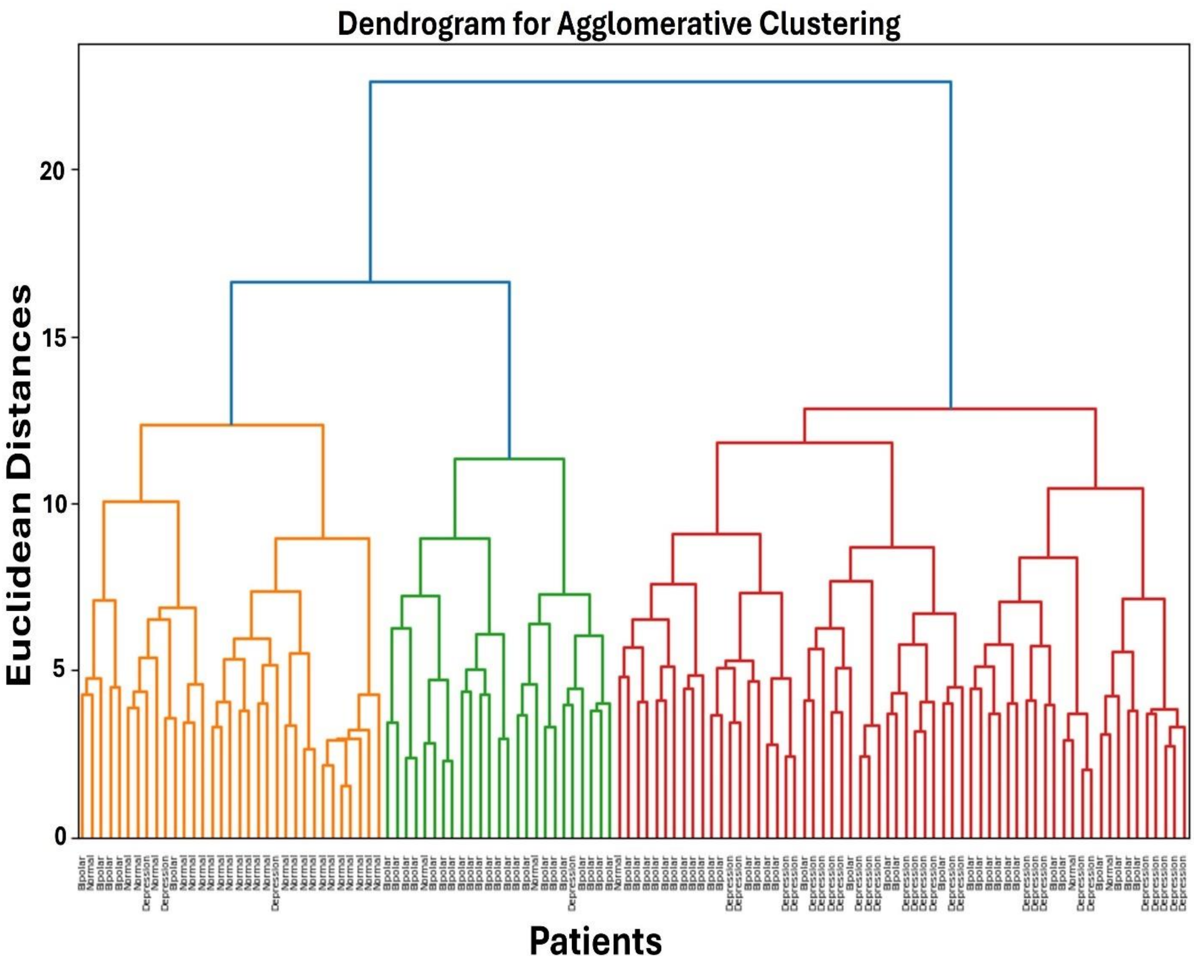


Figure 4: Dendrogram for hierarchical clustering

Discussion

In this study, association rule mining, decision tree analysis and agglomerative clustering have been used to highlight critical links between clinical symptoms, behavioral patterns and diagnostic categories in an in-depth study of the dataset. The high discriminatory power shown in the decision tree analysis about major depressive disorder reveals that mood swing is a primary predictor [15]. Further, the co–occurrence relationships between mood swing aggressive response euphoria overthinking and suicidal thoughts and nervous breakdown exhausted are identified as coinciding with clinical practice which is also explained by Cui *et al.* [16]. Results confirm previous research showing a large role in diagnosing MDD of mood and cognitive function symptoms [17]. In other words, the finding of new groups of symptoms such as combine sex reduction with an increase of nervous breaking down episodes, stress the power of a data driven discovery over newfound dependencies which multidisciplinary expert knowledge would not yield spontaneously [18]. Given that data set was based on expert diagnosis features, it was clustered into three distinct and separate clusters using agglomerative clustering [19]. The fact that the patients admitted in Cluster 0 included patients with bipolar disorder, depression, and other healthy patients, suggests that the model had difficulty distinguishing between the feature related overlap. In addition, association rule mining for finding cooccurrence between Euphoria, Overthinking, Suicidal Thoughts, Nervous Breakdown, and being Exhausted with Mood Swing and Aggressive Response was found to agree with clinical observations [20]. These results are consistent with prior reports suggesting that mood and cognitive symptoms have a dominant role in the diagnosis of MDD. Although, the discovery of striking disease symptoms coupled with the discovery of new striking disease symptoms such as those that jumped out were reduced sexual activity and increased incidence of nervous breakdown proves that data analytics is necessary to discover implicit patterns hidden from classical medical observations. Agglomerative clustering was performed to cluster the dataset into three separate clusters. Those diagnosed with bipolar disorder and depression and other common conditions in Cluster 0 have the most potential to suffer from the symptom overlap that proved difficult for our algorithm to tease apart. The proof of the algorithm's ability to do this is shown in Cluster 1 where the majority are diagnosed as healthy individuals. Cluster 2 is very well associated with bipolar disorder; hence it is possible to recognize this disease with high accuracy [19].

The relationship between symptomatic expressions in Major Depressive Disorder (MDD) and their impact on psychosocial functioning is multifaceted, involving various symptom domains and their interactions. Research indicates that patients often report a broader range of symptoms and greater functional impairment compared to healthcare providers assessments, highlighting discrepancies in perception [21]. Symptom networks reveal that a denser interconnection among symptoms is associated with a poorer prognosis, suggesting that specific symptoms like fatigue and guilt are critical in understanding the course of MDD. Some paper identifies that depressive symptoms, functional impairment, and quality of life are interrelated in major depressive disorder. Depressed mood significantly influences functional impairment and anxiety, while functional disability mediates the relationship between depressive symptoms and quality of life [22]. In some studies, it is also discussed the relationship between symptom severity and well-being in major depressive disorder, finding moderate to strong correlations. It highlights that while symptoms improved significantly, subjective well-being showed less improvement, indicating they are related but distinct concepts [23].

This study employs machine learning techniques such as association rule mining and decision tree classification to fill a significant gap in MDD research by concentrating on symptom interrelationships, as opposed to isolated symptom analysis [24]. These findings also offer a foundation upon which more holistic diagnostic frameworks and personalized treatment strategies for several conditions affecting genome organization and its contribution to disease and aging can be developed. Taking this study as heed, this research provides insights to doctors that based on this study, the patient presenting mood swings with sadness and fatigue may be flagged for early intervention to avoid worsening of the disorder [25]. However, due to this dataset's cross-sectional approach, this study is limited in the ability to draw conclusions regarding causality. Furthermore, data imbalance in symptom severity distribution of the dataset might have affected the generalizability of the decision tree model.

As future research, longitudinal data should be used to identify the progression of symptoms over time. However, future studies should build on this work by investigating symptom interrelationships in different populations that reflect cultural and environmental influences. Moreover, additional insights gain strength from other machine learning methods, for example neural networks which extract a more sophisticated and complex logic of nonlinear symptom patterns in MDD. This study shows in general that machine learning techniques can identify disease symptom interrelationships, thereby suggesting the opportunity to make better diagnoses and treat MDD more precisely. They may also inform the design of data aware mental health care platforms to promote better outcomes for this disabling disorder.

Conclusion

In this study, association rule mining, decision tree classification and agglomerative clustering techniques were used to investigate symptomatic relationships in Major Depressive Disorder (MDD). Identified the main predictors of MDD as mood swings and sadness, and uncovered co-occurrence patterns such as aggressive Response along with the euphoric or expert diagnosis, co-occurring with overthinking and euphoric which have high lift. Group level patterns and possible overlaps in diagnostic categories are revealed by clustering. This research fills gaps in traditional diagnostic methods which frequently neglect the complexity in symptom interactions, by concentrating on interrelationships of symptoms amongst themselves. Machine learning showed practical use of their machine learning techniques to reveal latent patterns, serving as a foundation for improved diagnostic accuracy and the research of personalized treatment. The drawback of the study is that it relies on a cross-sectional dataset not allowing causal inferences, and an ambiguous dataset imbalance impact on the model performance. The future work includes validation of these findings using longitudinal data and more diverse populations to explore dynamic changes in symptoms over time. The ability of computational techniques to support future advances in diagnostics of MDD and intervention tailored to individual patients is demonstrated in this study.

References

1. Sandmire HF, Austin SD, Bechtel RC. Experience with 40,000 Papanicolaou smears. Obstetrics and Gynecology. 1976 Jul 1;48(1):56-60.
2. Li Z, Ruan M, Chen J, Fang Y. Major depressive disorder: advances in

neuroscience research and translational applications. Neuroscience bulletin. 2021 Jun;37(6):863-880.
3. M. Malgaroli, A. Calderon, and G. A. Bonanno, "Networks of major depressive disorder: A systematic review," *Clin. Psychol. Rev.*, vol. 85, no. July 2020, p. 102000, 2021, doi: 10.1016/j.cpr.2021.102000.
4. Li Y, Gui Y, Zhao M, Chen X, Li H, Tian C, Zhao H, Jiang C, Xu P, Zhang S, Ye S. The roles of extracellular vesicles in major depressive disorder. Frontiers in psychiatry. 2023 Mar 9;14:1138110.
5. Yun M, Jeon M, Yang H. A novel machine learning-based prediction method for patients at risk of developing depressive symptoms using a small data. Plos one. 2024 May 22;19(5):e0303889.
6. Muñoz RA, McBride ME, Brnabic AJ, López CJ, Hetem LA, Secin R, Dueñas HJ. Major depressive disorder in Latin America: the relationship between depression severity, painful somatic symptoms, and quality of life. Journal of affective disorders. 2005 May 1;86(1):93-98.
7. Arnaud AM, Brister TS, Duckworth K, Foxworth P, Fulwider T, Suthoff ED, Werneburg B, Aleksanderek I, Reinhart ML. Impact of major depressive disorder on comorbidities: a systematic literature review. The Journal of clinical psychiatry. 2022 Oct 19;83(6):43390.
8. Tekkeşin Aİ. Artificial intelligence in healthcare: past, present and future. Anatol J Cardiol. 2019 Oct 1;22(Suppl 2):8-9.
9. Kumar S, Chong I. Correlation analysis to identify the effective data in machine learning: Prediction of depressive disorder and emotion states. International journal of environmental research and public health. 2018 Dec;15(12):2907.
10. Rai VK, Chakraborty S, Chakraborty S. Association rule mining for prediction of COVID-19. Decision Making: Applications in Management and Engineering. 2023 Apr 8;6(1):365-378.
11. Charbuty B, Abdulazeez A. Classification based on decision tree algorithm for machine learning. Journal of applied science and technology trends. 2021 Mar 24;2(01):20-28.
12. Barile S, Bassano C, Lettieri M, Piciocchi P, Saviano M. Intelligence augmentation (IA) in complex decision making: a new view of the VSA concept of relevance. InInternational Conference on Applied Human Factors and Ergonomics. Cham: Springer International Publishing. 2020 Jul 10; 251-258.
13. Rauf A, Ammar M, Azhar M, Noor N, Bakhtiar SM. Unveiling the Secrets of Obesity with Machine Learning (ML) Techniques/Algorithms. Current Trends in OMICS. 2024;4(1):55-76.
14. Tokuda EK, Comin CH, Costa LD. Revisiting agglomerative clustering. Physica A: Statistical mechanics and its applications. 2022 Jan 1;585:126433.
15. Colledani D, Anselmi P, Robusto E. Machine learning-decision tree classifiers in psychiatric assessment: An application to the diagnosis of major depressive disorder. Psychiatry Research. 2023 Apr 1;322:115127.
16. Cui L, Li S, Wang S, Wu X, Liu Y, Yu W, Wang Y, Tang Y, Xia M, Li B. Major depressive disorder: hypothesis, mechanism, prevention and treatment. Signal transduction and targeted therapy. 2024 Feb 9;9(1):30.
17. Krogh J, Benros ME, Jørgensen MB, Vesterager L, Elfving B, Nordentoft M. The association between depressive symptoms, cognitive function, and inflammation in major depression. Brain, behavior, and immunity. 2014 Jan 1;35:70-76.
18. Kost R, Littenberg B, Chen ES. Exploring generalized association rule mining for disease co-occurrences. InAMIA Annual Symposium Proceedings 2012;Vol. 2012: 1284.
19. Corponi F, Anmella G, Pacchiarotti I,

Samalin L, Verdolini N, Popovic D, Azorin JM, Angst J, Bowden CL, Mosolov S, Young AH. Deconstructing major depressive episodes across unipolar and bipolar depression by severity and duration: a cross-diagnostic cluster analysis on a large, international, observational study. Translational psychiatry. 2020 Jul 19;10(1):241.

20. Holdaway AS, Luebbe AM, Becker SP. Rumination in relation to suicide risk, ideation, and attempts: Exacerbation by poor sleep quality. Journal of affective disorders. 2018 Aug 15; 236:6-13
21. Christensen MC, Wong CM, Baune BT. Symptoms of major depressive disorder and their impact on psychosocial functioning in the different phases of the disease: do the perspectives of patients and healthcare providers differ?. Frontiers in Psychiatry. 2020 Apr 24;11:280.
22. Zhou J, Zhou J, Feng L, Feng Y, XiaoL, Chen X, Yang J, Wang G. The associations between depressive symptoms, functional impairment, and quality of life, in patients with major depression: undirected and Bayesian network analyses. Psychological Medicine. 2023 Oct;53(14):6446-6458.
23. Weijers A, Rasing S, Creemers D, Vermulst A, Schellekens AF, Westerhof GJ. The relationship between depressive symptoms, general psychopathology, and well-being in patients with major depressive disorder. Journal of Clinical Psychology. 2021 Jun;77(6):1472-1486.
24. Lee KS, Ham BJ. Machine learning on early diagnosis of depression. Psychiatry Investigation. 2022 Aug 24;19(8):597.
25. Lam RW, Kennedy SH, McIntyre RS, Khullar A. Cognitive dysfunction in major depressive disorder: effects on psychosocial functioning and implications for treatment. The Canadian Journal of Psychiatry. 2014 Dec;59(12):649-654.