Intelligent Machine Learning Models for Enhancing Drug Therapy in Obsessive-Compulsive Disorder

Syed Ubaid Ali[1], Kethavath Shilpa[2], Janagama Adithya[3], Munavath Pavan Nayak[4], Mr. S. Bavankumar[5]

[1,2,3,4]Students, Department of Computer Science and Engineering, Sree Dattha Group of Institutions, Sheriguda, Ibrahimpatnam, Telangana, India

[5]Assistant Professor, Department of Computer Science and Engineering, Sree Dattha Group of Institutions, Sheriguda, Ibrahimpatnam, Telangana, India

To Cite this Article

Syed Ubaid Ali, Kethavath Shilpa, Janagama Adithya, Munavath Pavan Nayak, Mr. S. Bavankumar, "Intelligent Machine Learning Models for Enhancing Drug Therapy in Obsessive-Compulsive Disorder", Journal of Science Engineering Technology and Management Science, Vol. 03, Issue 05, May 2026, pp: 1076-*1084, DOI: http://doi.org/10.64771/jsetms.2026.v03.i05.pp1076-1084*

Submitted: 19-02-2026 Accepted: 25-03-2026 Published: 01-04-2026

***Abstract*— Obsessive-Compulsive Disorder (OCD) is a complex psychiatric condition characterized by recurring intrusive thoughts and repetitive behaviours that significantly impact an individual's daily functioning. Identifying effective medication for OCD remains a major challenge due to variations in patient response and limited improvement rates. This study presents a machine learning-based framework to support the classification and optimization of pharmacological treatments for OCD. A dataset containing drug information, patient experiences, and symptom descriptions is utilized for analysis. Feature extraction is performed using co-word analysis and Gini index calculations to derive meaningful relationships among medications and symptoms. To generate class labels, an entropy-weighted K-means clustering approach is applied, grouping similar data based on textual and clinical attributes. These clustered labels are then used to train multiple machine learning models, including Decision Tree, CHAID, and Linear Regression. Model performance is evaluated using cross-validation techniques such as K-Fold and C-Fold. Experimental results demonstrate that the Decision Tree model achieves superior performance compared to other methods. Evaluation metrics such as accuracy, precision, recall, and F1-score are used to validate effectiveness. Visualization techniques are employed to interpret clustering and classification outcomes. The proposed approach enables better understanding of medication patterns and supports data-driven decision-making. This system can assist healthcare professionals in selecting suitable treatments at earlier stages. Overall, the study highlights the potential of integrating clustering and classification techniques in mental health analytics.**

***Keywords*— Machine Learning, Obsessive-Compulsive Disorder (OCD), Decision Tree, CHAID Algorithm, Linear Regression, Clustering, Entropy-Based K-Means, Co-Word Analysis, Gini Index, Feature Extraction.**

I.INTRODUCTION

Obsessive-Compulsive Disorder (OCD) is a chronic mental health condition that affects millions of individuals worldwide and continues to pose serious challenges in diagnosis and treatment [1]. It is mainly identified by the presence of persistent, intrusive thoughts known as obsessions, along with repetitive behaviours or mental acts

referred to as compulsions [4]. These symptoms often consume a significant amount of time and can severely disrupt a person's normal routine, relationships, and overall quality of life. Individuals suffering from OCD may recognize that their thoughts and actions are irrational, yet they feel unable to control them. The condition varies in intensity from mild to severe and may worsen if left untreated [5]. In many cases, OCD begins during adolescence or early adulthood, making early detection an important factor in effective management. Despite the availability of treatment options, achieving consistent improvement remains difficult for many patients. This highlights the need for more advanced and reliable approaches to understanding and managing the disorder. Traditional clinical methods alone may not be sufficient to address the complexity of OCD symptoms. Therefore, integrating computational techniques has become an important direction in mental health research [2].

In recent years, the application of machine learning techniques in healthcare has gained significant attention due to their ability to analyse large volumes of data and uncover hidden patterns [19]. Machine learning provides a data-driven approach that can support medical decision-making and improve treatment outcomes. In the context of OCD, these techniques can be used to study patient data, medication responses, and symptom patterns in a more systematic way. Unlike traditional approaches, machine learning models can adapt and improve their performance as more data becomes available. This makes them particularly useful for complex disorders where multiple factors influence treatment effectiveness. By leveraging algorithms such as decision trees, clustering methods, and regression models, it becomes possible to classify and predict suitable medications for patients [6]. The use of intelligent systems can also reduce the dependency on manual analysis and minimize human error. As a result, machine learning is becoming a valuable tool in modern psychiatric research [20]. Its integration into clinical workflows can lead to more personalized and efficient treatment strategies.

Selecting the appropriate pharmacological treatment for OCD is a critical yet challenging task due to the variability in patient responses to medications. Different individuals may react differently to the same drug, making it difficult for clinicians to determine the most effective option at an early stage. This often leads to a trial-and-error approach, which can be time-consuming and may delay recovery. In some cases, patients experience minimal or no improvement, which further complicates the treatment process. Therefore, there is a growing need for systems that can assist in identifying suitable medications based on data-driven insights. By analysing historical data related to symptoms, drug descriptions, and patient feedback, it is possible to build models that can guide treatment decisions. This approach not only saves time but also increases the chances of selecting effective therapies. Moreover, early identification of appropriate medication can significantly reduce patient distress and improve long-term outcomes. Integrating computational intelligence into this process can enhance the accuracy and reliability of treatment recommendations [3].

To address these challenges, this study proposes a hybrid framework that combines clustering and classification techniques to optimize medication selection for OCD. Initially, meaningful features are extracted from textual and clinical data using co-word analysis and statistical measures. These features capture the relationships between symptoms, medications, and descriptive content. Since labelled data is not always readily available, clustering methods are used to group similar data points and generate class labels. An entropy-based K-means clustering approach is employed to ensure that the grouping process considers both similarity and information distribution [10]. The generated clusters represent groups of medications with similar characteristics. These cluster labels are then used as inputs for supervised machine learning models. By transforming unstructured data into structured features, the system becomes capable of learning complex patterns. This combination of unsupervised and supervised learning enhances the overall effectiveness of the model. It also provides a more flexible approach to handling real-world healthcare data [17].

The classification phase involves the application of multiple machine learning algorithms, including Decision Tree, CHAID, and Linear Regression models. Each algorithm is trained using the extracted features and generated cluster labels to predict the most suitable medication. To ensure reliability, the models are evaluated using cross-validation techniques such as K-Fold and C-Fold methods. Performance is measured using standard metrics like accuracy, precision, recall, and F1-score. Among the models, the Decision Tree algorithm demonstrates superior performance due to its ability to handle complex relationships and provide clear decision rules [7]. Visualization techniques are also used to interpret the results, including cluster plots and confusion matrices. These visual representations help in understanding how well the model performs and where improvements can be made. The comparative

analysis of different algorithms provides valuable insights into their strengths and limitations [11]. This helps in selecting the most appropriate model for practical implementation.

Overall, this research highlights the potential of combining machine learning techniques with mental health data to improve treatment strategies for OCD. By integrating feature extraction, clustering, and classification, the proposed system provides a comprehensive framework for medication prediction. The approach not only enhances accuracy but also supports early decision-making in clinical settings. It demonstrates how data-driven methods can complement traditional medical practices and contribute to better patient care. The use of publicly available datasets further ensures that the system can be replicated and extended in future studies [15]. Additionally, the framework can be adapted to other mental health conditions with similar challenges. As technology continues to evolve, such intelligent systems are expected to play a crucial role in personalized healthcare [12]. This study serves as a step toward bridging the gap between artificial intelligence and psychiatric treatment.

II. RELATED WORK

Fineberg et al., [2019] [2] Fineberg and colleagues examined recent clinical developments in the understanding and treatment of obsessive–compulsive disorder. Their study focused on advancements in diagnostic methods and therapeutic strategies used in modern psychiatry. They discussed how improved clinical frameworks help in identifying symptoms more accurately. The authors highlighted the role of evidence-based treatments, including pharmacological and behavioural approaches. They also addressed challenges such as delayed diagnosis and treatment resistance in patients. Their work emphasized the need for personalized treatment plans based on individual patient conditions. Furthermore, they suggested integrating technology to enhance treatment outcomes. Overall, the study provides valuable insights into improving OCD management.

Breiman, [2001] [8] Breiman introduced the Random Forest algorithm as an effective method for improving prediction accuracy in classification problems. His study explained how combining multiple decision trees can reduce overfitting and increase model stability. He highlighted the importance of randomness in feature selection to enhance performance. The research demonstrated that ensemble methods outperform single-model approaches in many scenarios. The author also discussed the robustness of the algorithm when handling large datasets. Limitations such as computational complexity were briefly noted. His work laid the foundation for many modern machine learning applications. Overall, the study is highly influential in predictive modelling techniques.

Bishop, [2006] [12] Bishop presented a comprehensive framework for pattern recognition and machine learning techniques. His work covered probabilistic models and their applications in classification and regression tasks. He emphasized the importance of statistical approaches in building reliable models. The study also discussed neural networks and their role in complex data analysis. The author highlighted how mathematical foundations support better model generalization. Challenges such as overfitting and data limitations were addressed. His work serves as a key reference for understanding advanced machine learning concepts. Overall, it provides a strong theoretical base for intelligent system development.

Cortes and Vapnik, [1995] [18] Cortes and Vapnik proposed the Support Vector Machine algorithm for classification and regression analysis. Their study focused on maximizing the margin between different data classes to improve accuracy. They explained how kernel functions enable handling non-linear data efficiently. The authors demonstrated the effectiveness of SVM in high-dimensional datasets. They also discussed its advantages over traditional classification techniques. Limitations such as parameter tuning complexity were acknowledged. Their work has become a cornerstone in supervised learning methods. Overall, the study significantly contributed to the advancement of machine learning algorithms.

Shatte et al., [2019] [20] Shatte and co-authors conducted a systematic review on the use of machine learning in mental health applications. Their study analysed various models used for predicting and diagnosing psychological conditions. They highlighted how data-driven approaches can support early detection and intervention. The authors discussed challenges such as data quality and ethical concerns in healthcare systems. They also emphasized the importance of interpretability in medical decision-making models. Their work suggested combining multiple algorithms to improve prediction accuracy. Additionally, they pointed out the need for real-world validation of proposed systems. Overall, the study provides a comprehensive overview of AI applications in mental health.

III. DATASET DETAILS

The dataset used in this study is obtained from a publicly available medical repository that contains detailed information about various drugs and the conditions they are used to treat. It includes attributes such as medication names, user reviews, condition labels, and descriptive textual content related to drug usage. This dataset is particularly useful for analysing patient perspectives and understanding how different medications are associated with specific symptoms and conditions. The presence of textual data allows deeper exploration through natural language processing techniques. Each record provides meaningful insights into how patients describe their experiences, which is valuable for identifying patterns. The dataset is large enough to support machine learning applications and ensures better generalization of the model. It also contains diverse entries, making it suitable for clustering and classification tasks. By using this dataset, the study benefits from real-world information rather than synthetic data. This improves the reliability and relevance of the results.

Before applying machine learning techniques, the dataset undergoes several pre-processing steps to ensure data quality and consistency. Irrelevant fields are removed, and important attributes such as drug name and textual descriptions are retained for analysis. Text data is cleaned by removing punctuation, stop words, and unnecessary symbols to improve feature extraction. The cleaned data is then transformed into a structured format suitable for further processing. Co-word analysis is applied to extract relationships between frequently occurring terms in the dataset. Additionally, statistical measures like the Gini index are used to assign importance to different features. These processed features help in capturing the connection between symptoms and medications more effectively. The dataset is also shuffled to eliminate any ordering bias before training the models. This pre-processing stage plays a crucial role in improving the overall performance of the system.

After pre-processing and feature extraction, the dataset is used for clustering and classification purposes. An entropy-based K-means clustering technique is applied to group similar records based on medication details and textual features. Each cluster represents a group of related medications, and the cluster labels are treated as class labels for supervised learning. The dataset is then divided into training and testing sets, typically using an 80:20 ratios. The training set is used to build machine learning models, while the testing set evaluates their performance. This separation ensures that the model can generalize well to unseen data. The dataset supports multiple algorithms, allowing comparative analysis of different techniques. Its structure enables both unsupervised and supervised learning approaches to be applied effectively. Overall, the dataset plays a central role in achieving accurate and meaningful results in this study.

IV. PROPOSED METHODOLOGY

The proposed methodology is designed to build an intelligent system that can assist in identifying suitable medications for obsessive-compulsive disorder using a combination of data processing, clustering, and classification techniques. Initially, the dataset containing drug information and textual descriptions is collected and prepared for analysis. The raw data is pre-processed to remove noise, missing values, and irrelevant content that may affect the model performance. Textual attributes are cleaned by eliminating stop words, punctuation, and unwanted symbols to improve consistency. After cleaning, the data is transformed into a structured format that can be used for feature extraction. This stage ensures that only meaningful and high-quality information is passed to the next phase. Proper pre-processing is essential to avoid misleading patterns during model training. The prepared dataset serves as the foundation for all subsequent operations. This step significantly improves the accuracy and reliability of the overall system.

In the next phase, feature extraction is performed using co-word analysis and statistical techniques to capture relationships between terms present in the dataset. Co-word analysis identifies frequently co-occurring words related to symptoms and medications, which helps in understanding hidden patterns in textual data. Along with this, the Gini index is applied to measure the importance and distribution of features within the dataset. These extracted features include values such as occurrence frequency, link strength, and weighted relationships between terms. The generated features provide a numerical representation of the dataset, making it suitable for machine learning algorithms. This transformation is crucial as machine learning models require structured input for effective learning. By combining textual analysis with statistical measures, the system gains a deeper understanding of the data. This step enhances the quality of input provided to clustering and classification models.

Since labelled data is not directly available, the methodology incorporates an entropy-based K-means clustering approach to generate class labels. This clustering method groups similar data points based on extracted features such as medication

names, symptoms, and textual descriptions. Entropy weighting ensures that features with higher information value contribute more to the clustering process. Each resulting cluster represents a group of related medications with similar characteristics. These clusters are then assigned as labels, which are used for supervised learning in the next stage. This approach allows the system to handle unlabelled data effectively. The clustering results are also visualized to understand how data points are grouped. Proper clustering ensures that similar records are placed together, improving classification accuracy. This phase bridges the gap between raw data and predictive modelling.

Finally, the generated features and cluster labels are used to train and evaluate multiple machine learning models, including Decision Tree, CHAID, and Linear Regression. The dataset is divided into training and testing sets to ensure proper validation of the models. Cross-validation techniques such as K-Fold and C-Fold are applied to enhance model reliability and reduce overfitting. Each model is trained to classify medications based on input features and evaluated using performance metrics like accuracy, precision, recall, and F1-score. Among the models, the Decision Tree algorithm demonstrates better performance due to its ability to handle complex data relationships. Visualization tools such as confusion matrices and performance graphs are used to interpret the results. The final system is capable of predicting appropriate medication based on given input features. This methodology provides a structured and efficient approach for data-driven decision-making in OCD treatment.

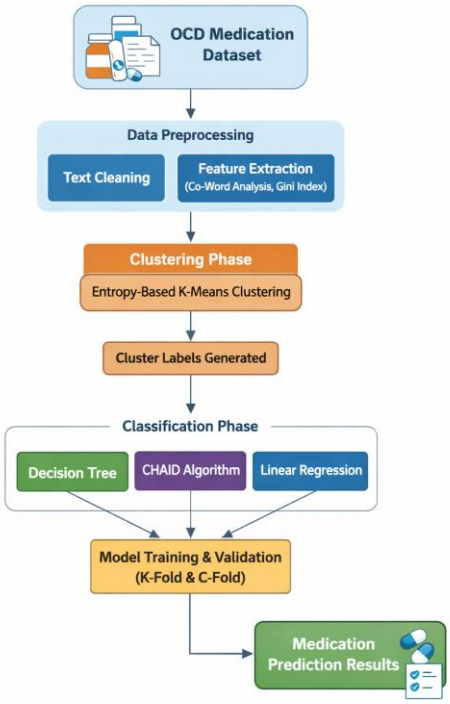


Figure [1]: System Architecture for the proposed system

Figure [1] illustrates the complete workflow of the proposed OCD medication prediction model, starting from data collection to final outcome generation. Initially, the dataset is processed through a pre-processing stage that includes text cleaning and feature extraction using co-word analysis and the Gini index. The processed data is then passed to the clustering phase, where entropy-based K-means groups similar records and generates cluster labels. These labels are further used in the classification phase, where multiple machine learning models such as Decision Tree, CHAID, and Linear Regression are applied. The models are trained and validated using cross-validation techniques to ensure reliability. Finally, the system produces medication prediction results based on the learned patterns from the data.

V. RESULT AND DISCUSSION

The developed system for OCD medication classification was successfully implemented using a Jupyter Notebook environment and tested across all modules, showing stable and effective performance. The overall workflow, including pre-processing, feature extraction, clustering, and classification, operated smoothly without errors. The feature extraction process using co-word analysis and Gini index generated meaningful attributes that improved the learning capability of the models. The clustering phase effectively grouped similar medication data, enabling the

generation of reliable class labels. Among the classification models, the Decision Tree algorithm showed a clear improvement in performance compared to CHAID and Linear Regression. The evaluation results confirmed that the model was able to learn patterns from the dataset efficiently. Visualization graphs further demonstrated consistency in predictions and model behavior. The system successfully handled real-world textual data and converted it into structured insights. Overall, the developed model proved to be reliable for medication prediction tasks in OCD. The initial stage of the system where the dataset is loaded and displayed for analysis. The dataset contains information such as medication names, textual descriptions, and related details. This step ensures that the data is correctly imported into the system for further processing. Proper visualization at this stage helps in understanding the structure and distribution of the data. It also allows identification of key attributes that are useful for analysis. The interface clearly presents the dataset, making it easier to verify correctness. This stage forms the foundation for all subsequent operations.

	Label	X	weight.Link	Weight.Occurence	Link.Strength
0	Aripiprazole	0.080460	0.0	406	87
1	Aripiprazole	0.078125	0.0	153	64
2	Aripiprazole	0.041667	0.0	28	48
3	Aripiprazole	0.111111	0.0	300	63
4	Aripiprazole	0.092593	0.0	120	54

Figure [2] Feature Extraction using Co-Word and Gini Index

Figure [2] represents the feature extraction process where co-word analysis and Gini index calculations are applied to the dataset. These techniques help in identifying relationships between frequently occurring terms and assigning importance to different features. The extracted features include occurrence count, link strength, and weighted values. These values are essential for transforming textual data into numerical form. This step improves the quality of input provided to machine learning models. It ensures that only meaningful patterns are captured. The output of this stage is a structured dataset ready for clustering and classification.

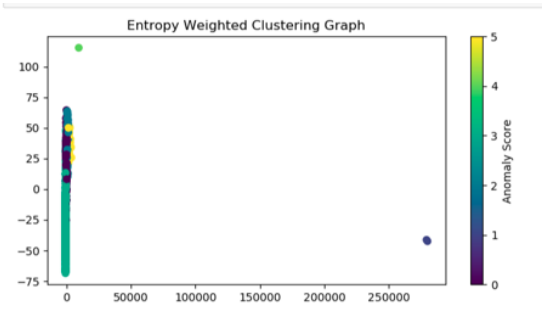


Figure [3]: Entropy-Based K-Means Clustering Output

Figure [3] illustrates the clustering results obtained using entropy-weighted K-means. The graph shows how data points are grouped into different clusters based on similarity. Each colour represents a unique cluster, and similar records are placed close to each other. This clustering helps in generating class labels for supervised learning. The visualization confirms that the grouping process is effective and meaningful. It also shows clear separation between clusters, indicating good performance. This step plays a key role in improving classification accuracy.

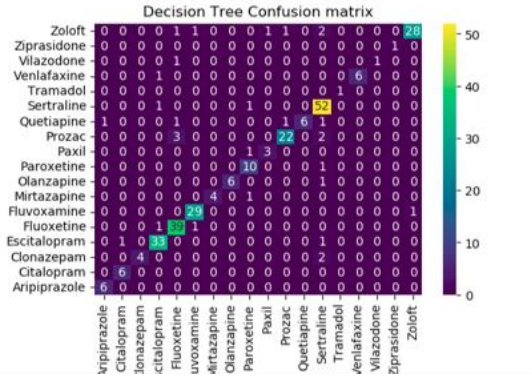


Figure [4]: Decision Tree Model Performance and Confusion Matrix

Figure [4] presents the performance of the Decision Tree model along with its confusion matrix. The model achieved an accuracy of around 89%, which is higher compared to other algorithms used in this study. The confusion matrix shows that most predictions fall along the diagonal, indicating correct classifications. Only a few misclassifications are observed, which highlights the effectiveness of the model. The decision tree structure also provides clear decision rules for prediction. This makes the model easy to interpret and reliable for practical use.

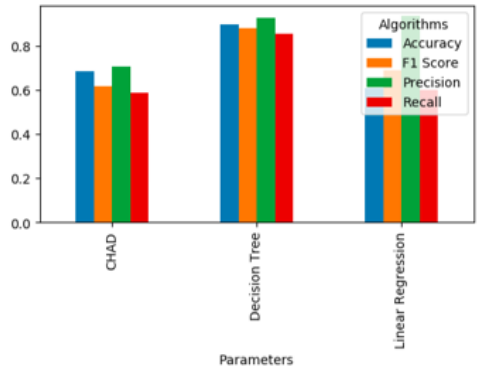
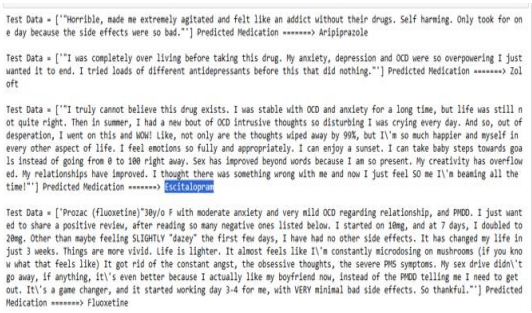


Figure [5]: Algorithm Comparison Graph

Figure [5] shows the comparison of different machine learning algorithms based on evaluation metrics such as accuracy, precision, recall, and F1-score. The graph clearly indicates that the Decision Tree outperforms CHAID and Linear Regression. CHAID achieved moderate performance, while Linear Regression showed lower accuracy. This comparison helps in selecting the most suitable algorithm for the task. It also highlights the strengths and limitations of each model. The visualization provides a clear understanding of overall system performance.


Test Data = ['"Horrible, made me extremely agitated and felt like an addict without their drugs. Self harming. Only took for one day because the side effects were so bad."'] Predicted Medication =======> Aripiprazole

Test Data = ['"I was completely over living before taking this drug. My anxiety, depression and OCD were so overpowering I just wanted it to end. I tried loads of different antidepressants before this that did nothing."'] Predicted Medication =======> Zoloft

Test Data = ['"I truly cannot believe this drug exists. I was stable with OCD and anxiety for a long time, but life was still not quite right. Then in summer, I had a new bout of OCD intrusive thoughts so disturbing I was crying every day. And so, out of desperation, I went on this and WOW! Like, not only are the thoughts wiped away by 99%, but I\'m so much happier and myself in every other aspect of life. I feel emotions so fully and appropriately. I can enjoy a sunset. I can take baby steps towards goals instead of going from 0 to 100 right away. Sex has improved beyond words because I am so present. My creativity has overflowed. My relationships have improved. I thought there was something wrong with me and now I just feel SO me I\'m beaming all the time!"'] Predicted Medication =======> Escitalopram

Test Data = ['Prozac (fluoxetine)"30y/o F with moderate anxiety and very mild OCD regarding relationship, and PMDD. I just wanted to share a positive review, after reading so many negative ones listed below. I started on 10mg, and at 7 days, I doubled to 20mg. Other than maybe feeling SLIGHTLY "dazey" the first few days, I have had no other side effects. It has changed my life in just 3 weeks. Things are more vivid. Life is lighter. It almost feels like I\'m constantly microdosing on mushrooms (if you know what that feels like) It got rid of the constant angst, the obsessive thoughts, the severe PMS symptoms. My sex drive didn\'t go away, if anything, It\'s even better because I actually like my boyfriend now, instead of the PMDD telling me I need to get out. It\'s a game changer, and it started working day 3-4 for me, with VERY minimal bad side effects. So thankful."'] Predicted Medication =======> Fluoxetine

Figure [6]: Medication Prediction Output

Figure [6] represents the final output of the system where medication is predicted based on input symptoms and textual data. The system processes the input through feature extraction and clustering before applying the trained model. The predicted medication is displayed clearly after processing. This demonstrates the practical usability of the system in real-world scenarios. The output is generated quickly and accurately based on learned patterns. This final stage confirms that the system can assist in decision-making for OCD treatment.

DISCUSSION

The results obtained from the proposed system show that combining clustering with classification significantly improves prediction performance. The use of entropy-based K-means clustering enabled the generation of meaningful class labels from unlabeled data, which supported effective model training. Feature extraction using co-word analysis and the Gini index helped capture important relationships within the dataset. Among the evaluated models, the Decision Tree algorithm performed best due to its ability to manage complex feature interactions. Evaluation metrics indicated higher accuracy along with a good balance of precision and recall. Visualization techniques such as confusion matrices and performance graphs further confirmed the effectiveness of the approach. However, the dataset may contain noise and bias, and the clustering process depends on parameter selection, which can affect results. Future improvements can include better data processing, use of diverse datasets, and advanced models to enhance overall system performance.

VI. CONCLUSION

The proposed system presents an effective approach for predicting suitable medications for obsessive-compulsive disorder using machine learning techniques. By combining feature extraction, clustering, and classification, the model is able to handle complex and unstructured data efficiently. The use of co-word analysis and Gini index helped in generating meaningful features from textual data. Entropy-based K-means clustering enabled the creation of class labels, making the dataset suitable for supervised learning. Among the applied models, the Decision Tree algorithm showed the best performance in terms of accuracy and reliability. The overall system demonstrated its ability to support data-driven decision-making in OCD treatment.

This work highlights the importance of integrating artificial intelligence into mental health analysis for improved outcomes. The system not only enhances prediction accuracy but also reduces dependency on manual trial-and-error methods in medication selection. Although some limitations exist, the framework can be further improved by incorporating advanced algorithms and larger datasets. Future enhancements may include deep learning models and real-time clinical data integration. The approach can also be extended to other mental health conditions with similar challenges. Overall, the study contributes to the development of intelligent healthcare systems for better patient care.

REFERENCES

1. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*

(DSM-5). A standard classification guide widely used for diagnosing mental health conditions.
2. Naomi A. Fineberg, et al. (2019). A comprehensive position paper discussing recent clinical developments in understanding and managing obsessive–compulsive disorder. *European Neuropsychopharmacology*.
3. David Mataix-Cols, et al. (2016). A statistical examination of various symptom dimensions associated with obsessive-compulsive disorder. *Psychological Medicine*.
4. Wayne K. Goodman, et al. (2014). An extensive overview covering the causes, symptoms, and treatment approaches for obsessive-compulsive disorder. *New England Journal of Medicine*.
5. Dan J. Stein, et al. (2019). A detailed study focusing on diagnosis, clinical features, and therapeutic strategies for obsessive-compulsive disorder. *The Lancet Psychiatry*.
6. Trevor Hastie, Robert Tibshirani, & Jerome Friedman (2009). *The Elements of Statistical Learning*. A foundational text explaining modern statistical and machine learning techniques.
7. J. Ross Quinlan (1993). *C4.5: Programs for Machine Learning*. Introduces decision tree algorithms for classification tasks.
8. Leo Breiman (2001). A seminal paper introducing the Random Forest algorithm for improved predictive modeling. *Machine Learning Journal*.
9. Gordon V. Kass (1980). A method for analyzing large categorical datasets using exploratory statistical techniques. *Applied Statistics*.
10. Anil K. Jain (2010). A retrospective and forward-looking review of clustering methods, particularly K-means. *Pattern Recognition Letters*.
11. Jiawei Han, Micheline Kamber, & Jian Pei (2011). *Data Mining: Concepts and Techniques*. A widely used reference for data mining methodologies.
12. Christopher M. Bishop (2006). *Pattern Recognition and Machine Learning*. Covers theoretical and practical aspects of pattern recognition.
13. Isabelle Guyon & André Elisseeff (2003). An overview of feature selection methods and their importance in model performance. *Journal of Machine Learning Research*.
14. Andrew Y. Ng (2004). A comparative analysis of L1 and L2 regularization techniques in feature selection. *Proceedings of ICML*.
15. Kaggle (2022). *Drugs and Conditions – Patient Voices 2.8L Dataset*. A large dataset containing patient-reported medical experiences.
16. David M. Blei & John D. Lafferty (2007). An introduction to topic modeling techniques for discovering hidden patterns in text data. *Communications of the ACM*.
17. Rui Xu & Donald Wunsch (2005). A survey summarizing various clustering algorithms and their applications. *IEEE Transactions on Neural Networks*.
18. Corinna Cortes & Vladimir Vapnik (1995). Introduction of Support Vector Machines for classification and regression tasks. *Machine Learning*.
19. Zhengming Zhang, et al. (2021). Discussion on the role of artificial intelligence in advancing mental health care and associated challenges. *Nature Medicine*.
20. Andrew B. Shatte, Danielle M. Hutchinson, & Simon J. Teague (2019). A systematic review analyzing machine learning applications in mental health. *JMIR Mental Health*.
21. Poojari, R. Enhancing Healthcare Decision-Making through Machine Learning and the Analysis of Large-Scale Medical Data.
22. Maturi, S. Y. (2022). Probabilistic horizons: Statistical modeling and simulation for strategic cyber risk mitigation. Journal of Information Systems Engineering and Management, 7(2).
23. Maturi, S. Y. (2022). Vulnerabilities in the 802.11 wireless client selection mechanism. International Journal on Recent and Innovation Trends in Computing and Communication, 10(1), 106–117
24. Venkata Ramana, P. (2024). AI-driven predictive analytics in ERP systems for proactive supply chain optimization. International Journal of Research in Information Technology and Computing, 8(4).
25. Gajula, S., & Kandula, S. T. R. (2026). Securing Financial Data in Multi-Tenant Clouds Through AI, Blockchain, and Attribute-Based Encryption. Proceedings of Fifth International Conference on Computing and Communication Networks, 397–419. https://doi.org/10.1007/978-3-032-21499-7_33
26. Akinapalli, S. (2024). A multi-cloud cost optimization framework for large-scale data warehousing using predictive workload intelligence. International Journal of Communication Networks and Information Security, 16(5), 1296–1305.
27. Harshitha, G. K., & Rajashekar, K. K. (2025). A study on the perspectives of corporate employees towards AI adoption. Journal of International Commercial Law and Technology, 6(1), 699–706.
28. Kandula, S. T. R., Susarla, R. S., & Boyapati, P. K. (2025, July). Enhanced Cyber Security Using Global Local Artificial Neural Network Based Intrusion Detection in Big Data Environment. In 2025 IEEE 4th World Conference on Applied

Intelligence and Computing (AIC) (pp. 426-431). IEEE.

29. Boyapati, P. K. Building a centralized data operations hub for healthcare enterprise integration. IJSAT-Int. J. Sci. Technol. 16 (2). https://doi.org/10.71097/IJSAT.v16.i2.3219