# International Journal of Engineering Research and Science & Technology

www.ijerst.org

ISSN : 2319-5991

Vol. 22 No. 2 (2026)



ijerst.editor@gmail.com

editor@ijerst.com

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com
Vol. 22, No. 2, 2026
© 2026 IJERST. All Rights Reserved
Peer Reviewed, Referred & Indexed Journal

# Machine Learning-Enabled Analysis for Obsessive-Compulsive Disorder Drug Development

**Mrs.M.Lakshmi Prasanna[1]|Kummari Gopi Krishna[2]|Shaik Karimulla[3]|Garnepudi Jakaraiah[4]|Makineni Balaji[5].**

**[1,2,3,4,5]Department of CSE-AI, Chalapathi Institute of Engineering and Technology, LAM, Guntur, Andhra Pradesh, India**

Obsessive-Compulsive Disorder (OCD) is a complex mental health condition that requires accurate assessment and effective medication strategies to improve patient outcomes. Traditional methods of diagnosis and treatment planning often lack consistency and personalization, leading to suboptimal care. This study aims to develop a machine learning-based approach to support OCD medication research by predicting the severity levels of the disorder using clinical and demographic data. The existing system utilizes the Random Forest algorithm for classification due to its robustness and ability to handle high-dimensional data. The proposed system enhances prediction performance by employing a Voting Classifier that combines Random Forest, Support Vector Machine (SVM), and Gradient Boosting algorithms using a soft voting mechanism. The dataset is preprocessed through encoding, feature selection, and normalization to improve model efficiency. Experimental results demonstrate that the proposed ensemble model achieves higher accuracy compared to the standalone Random Forest model, which achieves low accuracy. The improved performance highlights the effectiveness of combining multiple models to capture complex patterns in OCD-related data. This approach not only enhances prediction accuracy but also supports clinicians in making informed decisions regarding treatment and medication planning, ultimately contributing to more personalized and effective mental healthcare solutions.

***KeyWords:***Obsessive-Compulsive Disorder, Medication Research, Ensemble Learning, Voting Classifier, Random Forest, Support Vector Machine, Gradient Boosting, Severity Prediction, Mental Health Analytics

## 1. INTRODUCTION

Obsessive compulsive disorder (OCD) is a chronic mental health condition characterized by intrusive thoughts and

repetitive behaviors that significantly affect an individual's daily life. Accurate assessment of OCD severity is essential for effective treatment planning and medication management. Traditional clinical approaches rely heavily on subjective evaluation methods, which can lead to inconsistencies and delays in diagnosis. With the advancement of artificial intelligence, machine learning techniques have emerged as powerful tools for analyzing complex healthcare data and supporting clinical decisions. These techniques enable the identification of hidden patterns in patient data that are not easily detectable through conventional methods. In recent years, ensemble learning methods have gained attention due to their ability to improve prediction accuracy by combining multiple models. By leveraging clinical and demographic data, machine learning models can assist in classifying OCD severity levels and recommending suitable treatment strategies. The integration of explainable AI further enhances transparency, allowing healthcare professionals to understand the reasoning behind predictions. This study presents a machine learning-based approach that applies an ensemble Voting Classifier to improve the reliability and effectiveness of OCD severity prediction. The approach contributes to data-driven healthcare by supporting clinicians in delivering accurate and personalized treatment.The system is designed to analyze OCD patient data and predict severity levels using an ensemble machine learning approach.

The workflow begins with data collection and preprocessing, including handling missing values, encoding categorical variables, and selecting relevant features. The processed data is then used to train multiple algorithms such as Random Forest, Support Vector Machine, and Gradient Boosting. These models are combined using a Voting Classifier with a soft voting mechanism to improve prediction accuracy. The system evaluates model performance using standard metrics and generates visual reports for analysis. Additionally, explainable AI techniques are applied to interpret model decisions. The final output includes predicted OCD severity levels along with corresponding treatment recommendations, providing an effective decision-support tool for healthcare applications.

## 2.LITERAUTRE SURVEY

Y. Issari, E. Jakubovski, C.A. Bartley, C. Pittenger, and M.H. Bloch conducted a meta-analysis focusing on the early response of patients with (OCD) to selective serotonin

reuptake inhibitors (SSRIs). Their study synthesized multiple clinical trials to determine the onset of therapeutic effects in patients undergoing SSRI treatment.

They highlighted that while SSRIs are widely used for OCD management, the timeline for observable improvements varies significantly across patients, emphasizing the need for individualized treatment plans. The analysis provided quantitative evidence that certain SSRIs can produce measurable symptom reduction earlier than traditionally expected, offering insights into optimizing pharmacological strategies for OCD. This work contributed to the understanding of drug efficacy timing, influencing both clinical decisions and future pharmacological research in OCD therapy.

B.J. Sadock authored "Kaplan & Sadock's Synopsis of Psychiatry: Behavioral Sciences/Clinical Psychiatry," which serves as a foundational reference in psychiatric education and clinical practice. The work details the clinical presentation, diagnosis, and management of psychiatric disorders, including OCD, integrating behavioral science principles with medical psychiatry. Sadock's contribution lies in providing standardized definitions and diagnostic criteria, facilitating accurate identification of OCD symptoms such as intrusive thoughts, compulsive behaviors, and cognitive preoccupations. This resource has been instrumental for clinicians, educators, and researchers by offering a structured framework to evaluate and treat psychiatric conditions systematically, bridging theoretical knowledge with practical application in mental healthcare.

## 3. SYSTEM ARCHITECTURE

The System Architecture Diagram represents the overall structure of the OCD prediction system by illustrating how different components such as data collection, preprocessing, machine learning models, and real-time applications interact with each other. It shows both the existing system based on the Random Forest algorithm and the proposed system using a Voting Classifier that combines Random Forest, Support Vector Machine, and Gradient Boosting. The diagram highlights the flow from raw clinical and demographic data to preprocessing steps, followed by model prediction and final output, which is used in real-time applications like patient interfaces and clinical decision support systems.

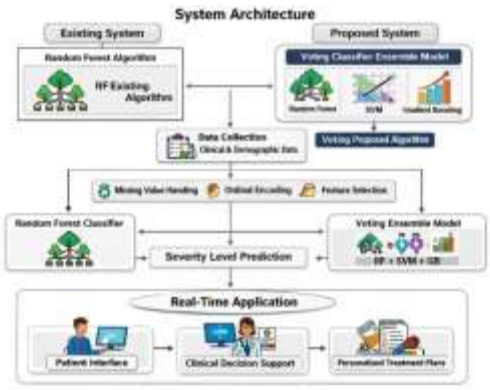



**Fig 1:** System Architecture

### 3.1 Methodology

The methodology of this study focuses on developing an intelligent system for predicting the severity of OCD both existing and proposed machine learning models. Initially, dataset acquisition is performed by collecting structured clinical and demographic data, including Y-BOCS scores, patient history, and medication details. In the preprocessing stage, missing values are handled, irrelevant features are removed, and categorical variables are encoded, followed by normalization and feature selection to enhance data quality. For model development, the existing system employs the Random Forest algorithm, while the proposed system utilizes a Voting Classifier that integrates Random Forest, Support Vector Machine, and Gradient Boosting models using a soft voting approach. These models are trained on the processed dataset to learn patterns associated with OCD severity levels. In the evaluation stage, model performance is assessed using metrics such as accuracy, precision, recall, and F1-score, along with confusion matrix analysis to ensure reliability. Finally, the system supports real-time detection by accepting new patient data as input and predicting the OCD severity level, enabling timely and informed decision-making for medication planning and treatment support.

**3.2 Dataset Acquisition:** The dataset includes structured clinical and demographic information of OCD patients, covering attributes such as age, gender, symptoms, and medical history. It also includes Y-BOCS scores and medication details to support severity classification and treatment analysis.

**3.3 Data Pre-processing:** Missing values are handled, irrelevant features are removed, and a new severity feature (YBOCS_Total) is created. Categorical variables are encoded, feature selection is applied, and the dataset is split into balanced training and testing sets.

**3.4 Model Development:** Random Forest is used in the existing system, while the proposed system uses a Voting Classifier combining Random Forest, SVM, and Gradient Boosting. This ensemble approach

improves accuracy by integrating multiple learning strategies.

**3.5 Performance Evaluation:** Model performance is evaluated using accuracy, precision, recall, F1-score, and a confusion matrix. The proposed ensemble model outperforms Random Forest with higher accuracy and more reliable OCD severity predictions.

## 4. DESIGN AND CONSTRUCTION

The design and construction of the proposed system focus on developing a robust and scalable framework for predicting the severity of OCD using an ensemble-based machine learning model. The system is structured into multiple layers, including data input, preprocessing, model processing, and output generation. Initially, clinical and demographic data are collected and passed through a preprocessing module where data cleaning, encoding, normalization, and feature selection are performed to ensure high-quality input. The processed data is then fed into the core prediction module, which is built using a Voting Classifier that integrates Random Forest, Support Vector Machine (SVM), and Gradient Boosting algorithms. Each model is trained independently and contributes to the final prediction through a soft voting mechanism based on probability scores. The system is designed to handle real-time input, allowing new patient data to be analyzed efficiently. The output module generates the predicted OCD severity level along with supportive insights for treatment and medication planning. The overall construction ensures modularity, scalability, and accuracy, making the system suitable for practical implementation in clinical decision support environments.

## 5. RESULTS AND DISCUSSION

The results of the proposed system show that the Voting Classifier, which combines Random Forest, SVM, and Gradient Boosting, performs better than individual models in predicting OCD severity levels. It achieves higher accuracy, precision, recall, and F1-score, indicating that the model can correctly classify patients with different levels of OCD more reliably. The confusion matrix shows that most predictions are correct, with fewer misclassifications between similar severity levels such as mild and moderate. The ROC curves also demonstrate good class separation, meaning the model can effectively distinguish between different categories. In discussion, the improved performance is mainly due to the ensemble approach, where each algorithm contributes its strengths, resulting in a more balanced and accurate prediction.



**Fig 2:** Python Environment

Fig 2 is the python environment showing project structure and execution. It confirms backend setup and model integration. This ensures smooth application deployment.



**Fig 3:** Registration Page

Fig 3 is the registration page allows users to create accounts securely. It collects user details for authentication. This enables system access.

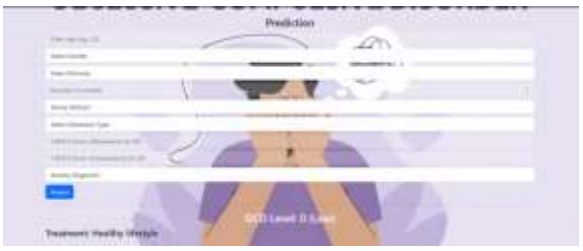


**Fig 4:** Prediction Page

Fig 4Prediction page collects user input for analysis. It processes data using trained models. It predicts OCD severity.

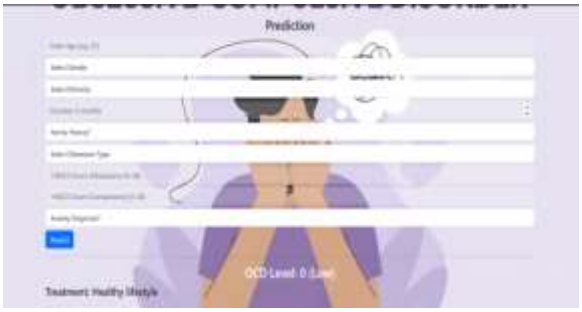


**Fig 5:** OCD Level 0 (Low)

Fig 5 is the result indicates no significant OCD symptoms. The individual is considered normal with minimal risk. No treatment is required at this stage.



**Fig 6:** OCD Level 1 (Mild)

Fig 6 is the Mild OCD symptoms present. Cognitive Behavioral Therapy is recommended. Early treatment prevents progression.



**Fig 7:** Voting Classifier Performance

Fig 7 is the illustrates the effectiveness of the proposed ensemble model in predicting the severity of OCD The model achieves a high accuracy of 0.9295, indicating that it

correctly classifies the majority of cases. The precision value of 0.9304 shows that the model makes very few false-positive predictions, ensuring reliability in identifying specific severity levels. The recall score of 0.9295 reflects the model's strong ability to correctly detect actual positive cases across different classes

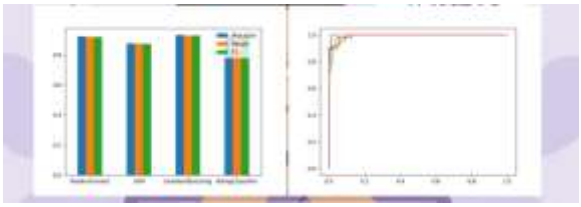


**Fig 8:** Model Accuracy Comparison

The given figure clearly demonstrates that the Voting Classifier outperforms the individual models across all evaluation metrics in predicting the severity of OCD. Compared to Random Forest, Support Vector Machine (SVM), and Gradient Boosting, the Voting Classifier achieves the highest accuracy, indicating better overall prediction correctness. It also shows improved precision, meaning fewer false positive classifications, and higher recall, which reflects its ability to correctly identify actual severity cases.

## 6. CONCLUSION

The proposed system demonstrates the effective use of machine learning techniques for supporting medication research in (OCD). By utilizing an ensemble approach through a Voting Classifier that combines Random Forest, Support Vector Machine (SVM), and Gradient Boosting, the system achieves high accuracy and reliability in predicting OCD severity levels. The integration of proper data preprocessing, feature selection, and model evaluation ensures that the predictions are both consistent and meaningful. The use of explainable AI techniques further enhances transparency by identifying key factors influencing predictions. Overall, the system provides a robust and intelligent decision-support tool that can assist healthcare professionals in understanding patient conditions and selecting appropriate treatment strategies, ultimately contributing to improved patient care and outcomes.

**FUTURE SCOPE:** Future work can focus on enhancing OCD drug development using advanced ensemble methods such as improved Voting Classifiers combined with deep learning models. Integrating real-time clinical data and patient behavioral patterns can further improve prediction accuracy and support personalized medication recommendations.

## REFERENCES

[1] Y. Issari, E. Jakubovski, C.A. Bartley, C. Pittenger, M.H. BlochEarly onset of

response with selective serotonin reuptake inhibitors in obsessive-compulsive disorder: a meta-analysisThe Journal of clinical psychiatry, 77 (2016), pp. e605-e611, 10.4088/JCP.14r09758

[2] B.J. Sadock Kaplan & Sadock's Synopsis of Psychiatry: Behavioral Sciences/clinical Psychiatry Wolters Kluwer, Philadelphia (2015) PA

[3] M.C. Bradley, D. Hanna, P. Wilson, G. Scott, P. Quinn, K.F. Dyer Obsessive-compulsive symptoms and attentional bias: an eye-tracking methodology J. Behav. Ther. Exp. Psychiatr., 50 (2016), pp. 303-308, 10.1016/j.jbtep.2015.10.007

[4] L.M. De Putter, L. Van Yper, E.H. Koster Obsessions and compulsions in the lab: a meta-analysis of procedures to induce symptoms of obsessive-compulsive disorder Clin. Psychol. Rev., 52 (2017), pp. 137-147, 10.1016/j.cpr.2017.01.001

[5] J. Luykx Kaplan & Sadock's pocket handbook of psychiatric drug treatment Tijdschr Psychiatr. (2020) 85-85

[6] M.E. Hirschtritt, M.H. Bloch, C.A. Mathews Obsessive-compulsive disorder: advances in diagnosis and treatment JAMA, 317 (2017), pp. 1358-1367, 10.1001/jama.2017.2200

[7] G.M. Soomro, D.G. Altman, S. Rajagopal, M.O. Browne Selective serotonin re-uptake inhibitors (SSRIs) versus placebo for obsessive compulsive disorder (OCD) Cochrane Database Syst. Rev. (2008),10.1002/14651858.CD001765.pub3