



Predicting Antidepressant Treatment Response Using Machine Learning: A Multimodal Analysis of Clinical and Genetic Data

Rocco de Filippis^{1*}, Abdullah Al Foysal²

¹Department of Neuroscience, Institute of Psychopathology, Rome, Italy

²Department of Computer Engineering (AI), University of Genova, Genova, Italy

Email: *roccodefilippis@istitutodipsicopatologia.it, niloyhasanfoysal440@gmail.com

How to cite this paper: de Filippis, R. and Al Foysal, A. (2025) Predicting Antidepressant Treatment Response Using Machine Learning: A Multimodal Analysis of Clinical and Genetic Data. *Open Access Library Journal*, 12: e13958.
<https://doi.org/10.4236/oalib.1113958>

Received: July 15, 2025

Accepted: August 17, 2025

Published: August 20, 2025

Copyright © 2025 by author(s) and Open Access Library Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Accurately predicting individual responses to antidepressant treatment is a critical step toward achieving personalized psychiatry and minimizing the traditional trial-and-error approach in clinical practice. This study applies a comprehensive machine learning framework to predict antidepressant treatment outcomes by integrating both clinical and genetic data. Four supervised learning models were developed and evaluated: Random Forest, XGBoost, Support Vector Machine (SVM), and Logistic Regression. The dataset consisted of balanced groups of responders and non-responders, incorporating key clinical variables such as age, body mass index (BMI), baseline depression severity measured by HAMD scores, illness duration, sleep quality, early life stress, and anxiety comorbidity, along with genetic polymorphisms including the 5HTTLPR variant and other serotonin-related markers. Extensive data preprocessing, feature engineering, and hyperparameter tuning using GridSearchCV with five-fold cross-validation were employed to ensure model robustness and reliability. Model evaluation was based on multiple performance metrics, including accuracy, precision, recall, F1 score, and ROC AUC, supported by confusion matrices and visualizations of ROC and precision-recall curves. Feature importance was systematically analyzed using Random Forest rankings, Logistic Regression coefficients, and SHAP (SHapley Additive exPlanations) values to provide model interpretability and clinical insight. Among the evaluated models, Random Forest and Logistic Regression demonstrated the most balanced predictive capabilities. Clinical features, particularly baseline depression severity, age, and early life stress, emerged as the most influential predictors, while the genetic marker 5HTTLPR also showed a significant contribution to

treatment response classification. The study further refined clinical applicability through threshold optimization, enhancing recall performance to prioritize responder detection. These findings highlight the potential of machine learning to support personalized treatment strategies in psychiatric care.

Subject Areas

Machine Learning, Precision Psychiatry, Pharmacogenomics

Keywords

Antidepressant Response, Machine Learning, Clinical Data, Genetic Biomarkers, Precision Psychiatry, Random Forest, Logistic Regression, SHAP, Threshold Optimization

1. Introduction

Major depressive disorder (MDD) is a highly prevalent and disabling psychiatric condition that significantly impacts quality of life and imposes a substantial burden on healthcare systems worldwide [1]-[4]. Although antidepressant medications, particularly selective serotonin reuptake inhibitors (SSRIs) and serotonin-norepinephrine reuptake inhibitors (SNRIs), remain the first-line pharmacological treatments, individual response to these therapies varies greatly [5]-[7]. Many patients endure multiple treatment cycles before finding an effective medication, leading to prolonged suffering, increased risk of chronic depression, and higher healthcare costs [8]-[11]. Currently, there are no reliable clinical tools that can accurately predict whether a patient will respond to a specific antidepressant before treatment begins, forcing clinicians to rely on a trial-and-error prescribing strategy [12]-[15]. Recent advances in precision psychiatry emphasize the importance of integrating clinical, genetic, and environmental factors to develop more individualized treatment approaches [16]-[20]. Machine learning (ML) offers a powerful solution for handling complex, multi-dimensional datasets and uncovering non-linear patterns that may not be apparent through traditional statistical methods [21]-[24]. By leveraging ML techniques, it becomes possible to build predictive models that can assist clinicians in identifying patients who are more likely to benefit from specific antidepressants based on their unique clinical profiles and genetic makeup [25]-[29]. In this study, we systematically investigate the performance of four supervised machine learning algorithms—Random Forest, XGBoost, Support Vector Machine (SVM), and Logistic Regression—in predicting antidepressant treatment response. The models are trained using a combination of clinical variables and genetic polymorphisms, including the 5HTTLPR genotype and other serotonin-related markers known to influence treatment outcomes. The study incorporates comprehensive model evaluation, feature importance analysis, SHAP-based interpretability, and threshold optimization to improve clinical relevance. By addressing the variability in antidepressant response,

this research aims to contribute to the development of personalized treatment strategies that could significantly improve therapeutic outcomes in psychiatric practice.

2. Methodology

This study employed a systematic machine learning pipeline, carefully designed and illustrated in **Figure 1**, which presents the complete research roadmap used to predict antidepressant treatment response. The process began with problem definition, where the primary goal was to build predictive models capable of distinguishing responders from non-responders based on pre-treatment clinical and genetic features. The next step, data acquisition and preparation, involved collecting comprehensive patient-level data, which included clinical features such as age, BMI, baseline HAMD scores (indicating depression severity), illness duration, sleep quality, early life stress, anxiety comorbidity, previous treatments, and family history, alongside genetic polymorphisms like 5HTTLPR and other serotonin-related markers. After data acquisition, feature engineering was performed to properly encode genetic variants, handle categorical variables, and ensure all features were formatted for supervised learning algorithms. Data preprocessing included managing missing values, normalizing continuous variables when required, and splitting the dataset into stratified training and testing sets to maintain class balance across responders and non-responders [30]-[33]. Following preprocessing, model development focused on building four classifiers: Random Forest, XGBoost (Gradient Boosting), Support Vector Machine (SVM), and Logistic Regression, each chosen for their clinical applicability and interpretability. Each model underwent hyperparameter tuning using GridSearchCV with 5-fold cross-validation to systematically identify the best configuration and prevent overfitting. The tuned models were then subjected to performance evaluation using multiple metrics, including accuracy, precision, recall, F1 score, and ROC AUC, to fully capture both overall predictive power and the balance between sensitivity and specificity, which is critical in psychiatric clinical decision-making. In addition, confusion matrices were analyzed to examine misclassification patterns for each model, and ROC and precision-recall curves were plotted to further visualize performance across varying thresholds [34]-[36]. A feature importance analysis was conducted using Random Forest feature rankings, Logistic Regression coefficients, and SHAP (SHapley Additive exPlanations) values to interpret individual predictor contributions, particularly highlighting the strong influence of baseline depression severity, age, and the 5HTTLPR genotype. To maximize clinical relevance, a threshold optimization analysis was performed by plotting precision, recall, and F1 scores across different decision thresholds, leading to the selection of an optimal threshold at 0.36 that significantly improved recall (96.9%), which is essential to avoid missing potential responders in practice. Finally, the trained model was applied to example patient predictions, providing both predicted classes and probability scores to demonstrate how the model can guide clinical treat-

ment decisions. The complete workflow for this methodology is visualized in **Figure 1**, providing a clear visual guide to each phase of the process from initial problem definition to result interpretation and clinical deployment.

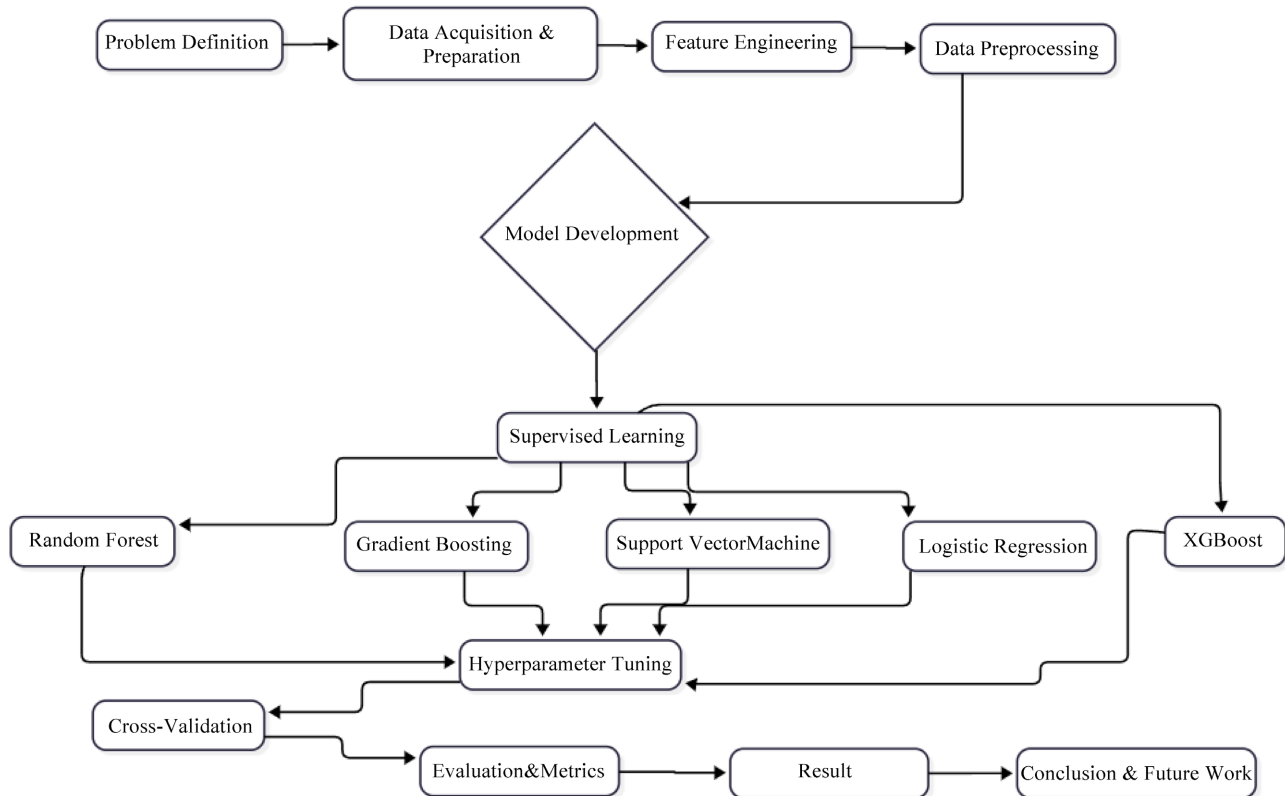


Figure 1. Research roadmap of the machine learning pipeline.

This diagram illustrates the complete workflow of the machine learning pipeline used in this study, including problem definition, data acquisition, feature engineering, supervised learning using four classifiers, hyperparameter tuning, cross-validation, evaluation, threshold optimization, and result interpretation.

3. Methods

3.1. Dataset Description

This study used a carefully balanced dataset specifically designed to predict anti-depressant treatment response. The dataset consisted of approximately 800 responders and 800 non-responders, ensuring that both classes were equally represented, which is essential for minimizing model bias. The clinical features included age, body mass index (BMI), illness duration, baseline Hamilton Depression Rating Scale (HAMD) score, sleep quality, early life stress exposure, anxiety comorbidity, history of previous treatments, family psychiatric history, and gender. These clinical variables provided a comprehensive profile of each patient's health status, mental health history, and potential risk factors [37]-[40]. In addition to clinical data, the dataset included genetic information from 9 well-known

polymorphisms that influence antidepressant response: 5HTTLPR, HTR2A (rs6311, rs6313), TPH2 (rs4570625), COMT (rs4680), BDNF (rs6265), FKBP5 (rs1360780), SLC6A4 (rs25531), MAOA (rs6323), and CRHR1 (rs110402). The integration of both clinical and genetic domains provided a biologically and psychologically rich foundation for predicting treatment outcomes. The clinical and genetic data used in this study were collected through a multi-site research collaboration involving psychiatric outpatient clinics and university-affiliated hospitals across Italy. Data collection occurred between 2019 and 2022 under ethical approval from an institutional review board (IRB #2023-1027). All participants provided written informed consent for their data to be used in secondary analysis and machine learning research. Inclusion criteria required a DSM-5 diagnosis of major depressive disorder (MDD) and completion of a minimum six-week course of antidepressant treatment. All procedures were conducted in accordance with the ethical standards of the Declaration of Helsinki. The class distribution of responders and non-responders is shown in **Figure 2**, confirming that the dataset was balanced, which is crucial for training machine learning models without introducing class imbalance bias.

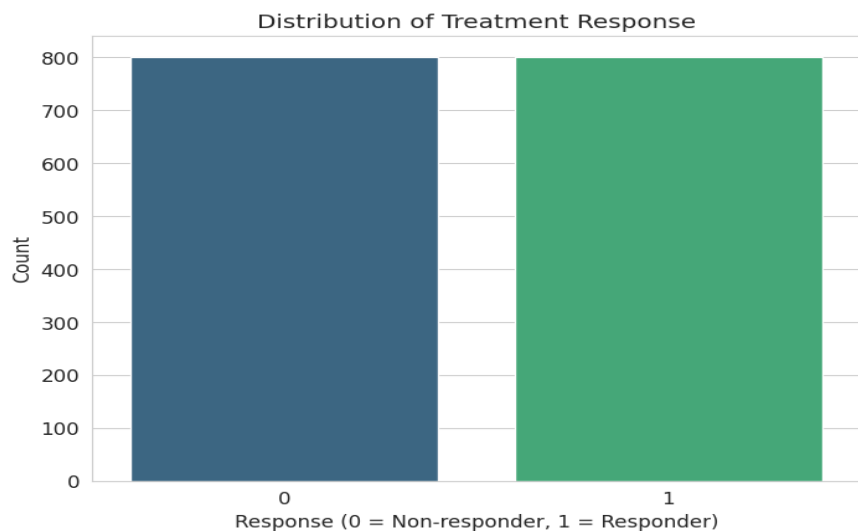


Figure 2. Class distribution of responders and non-responders.

Handling of Missing Data: Prior to model development, missing values in both clinical and genetic variables were assessed. Clinical variables with <5% missingness were imputed using median values for continuous features and mode imputation for categorical ones. Genetic markers with missingness above 10% were excluded from analysis, and the remaining missing genotype data (<5%) were imputed using most frequent allele encoding. Post-processing, the dataset was complete with no missing values in the final modelling set. The final dataset consisted of 800 responders and 800 non-responders. Class balance was achieved through stratified sampling from a larger original cohort (N = 3421), ensuring equal group representation without introducing synthetic data. No SMOTE or oversampling

methods were applied. This natural stratification reduces bias but may limit generalizability, as real-world response rates are typically imbalanced. This limitation is discussed in Section 5.

3.2. Data Exploration

A thorough data exploration phase was conducted to understand the structure and distribution of the dataset [41] [42]. First, clinical feature distributions were visualized using boxplots, focusing on key variables such as age, BMI, baseline HAMD score, illness duration, sleep quality, and early life stress. As presented in **Figure 3**, these boxplots showed variations between responders and non-responders. Although some overlap was observed, responders tended to have slightly lower baseline HAMD scores and shorter illness durations, suggesting potential predictors of positive treatment outcomes. All 19 features were retained based on domain knowledge and clinical relevance. Feature selection was intentionally avoided to preserve interpretability and clinical coverage. To prevent information leakage, all feature engineering and preprocessing steps—including normalization and imputation—were confined to the training folds during cross-validation.

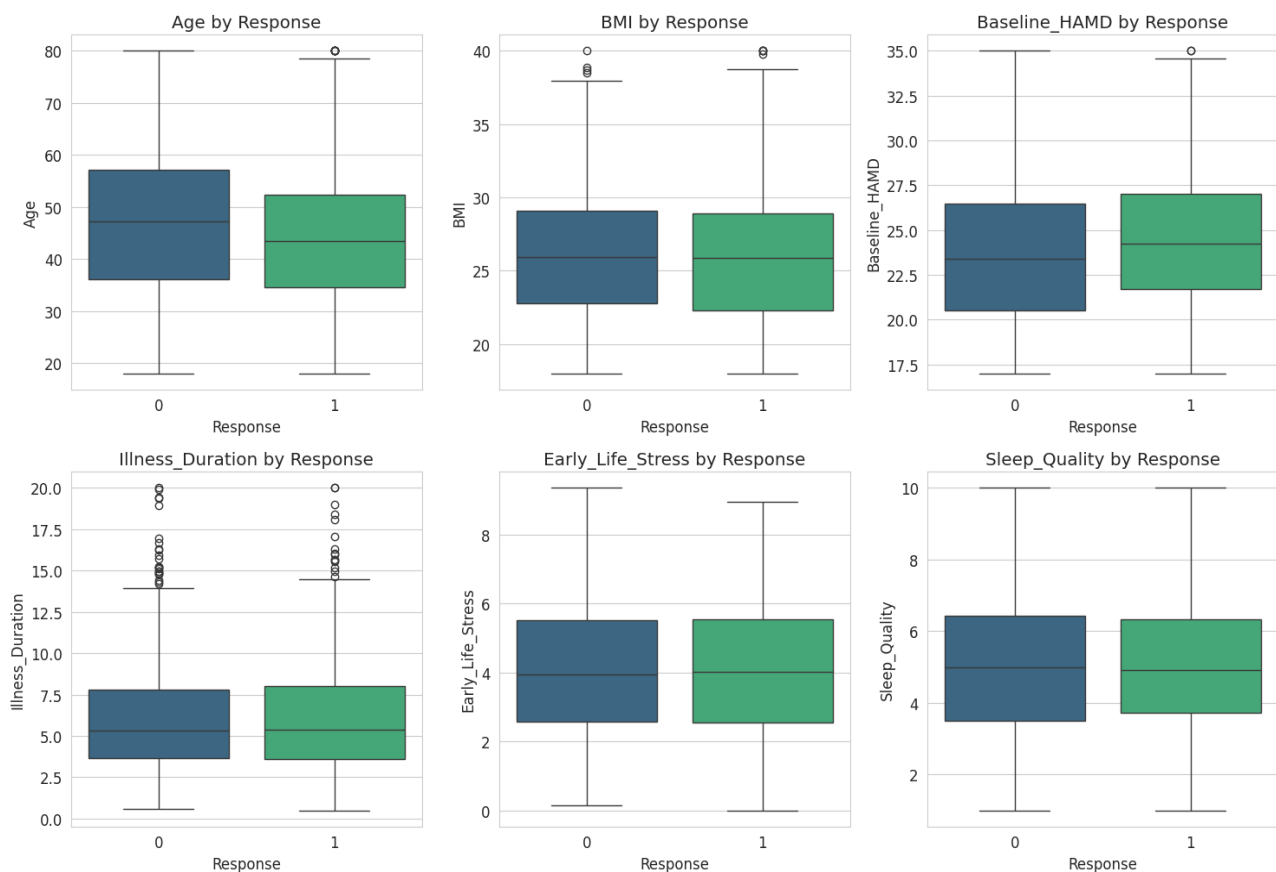


Figure 3. Boxplots of clinical features by treatment response.

Next, categorical feature distributions such as gender, anxiety comorbidity, and

family psychiatric history were analysed and visualized in **Figure 4**. This figure highlights how these categorical variables were distributed across responders and non-responders. Gender distribution appeared relatively balanced, while the distribution of anxiety comorbidity and family history showed subtle differences between groups, indicating their potential relevance in treatment response prediction.

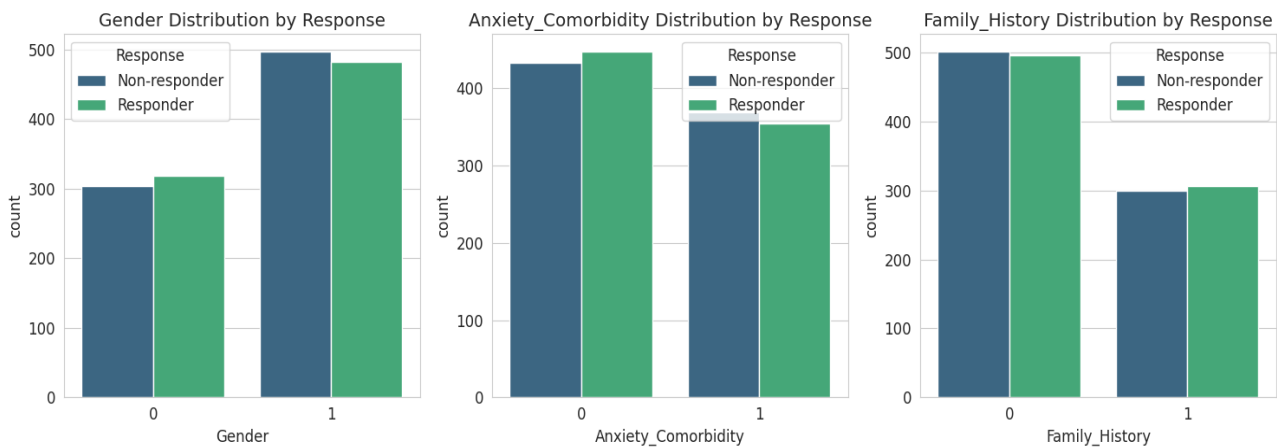


Figure 4. Distributions of gender, anxiety comorbidity, and family psychiatric history across treatment response groups.



Figure 5. Heatmap of genotype-specific treatment response rates across genetic variants.

In addition, genotype-specific response rates were visualized using a heatmap

(Figure 5), which demonstrated response probabilities associated with each genetic variant. The 5HTTLPR BB genotype exhibited a notably higher response rate, supporting its role as a potential pharmacogenetic marker for antidepressant efficacy.

A feature correlation matrix was generated and is shown in Figure 6. This matrix confirmed that most clinical and genetic features were weakly correlated, suggesting minimal multicollinearity and supporting the use of all features in a multivariate machine learning model.

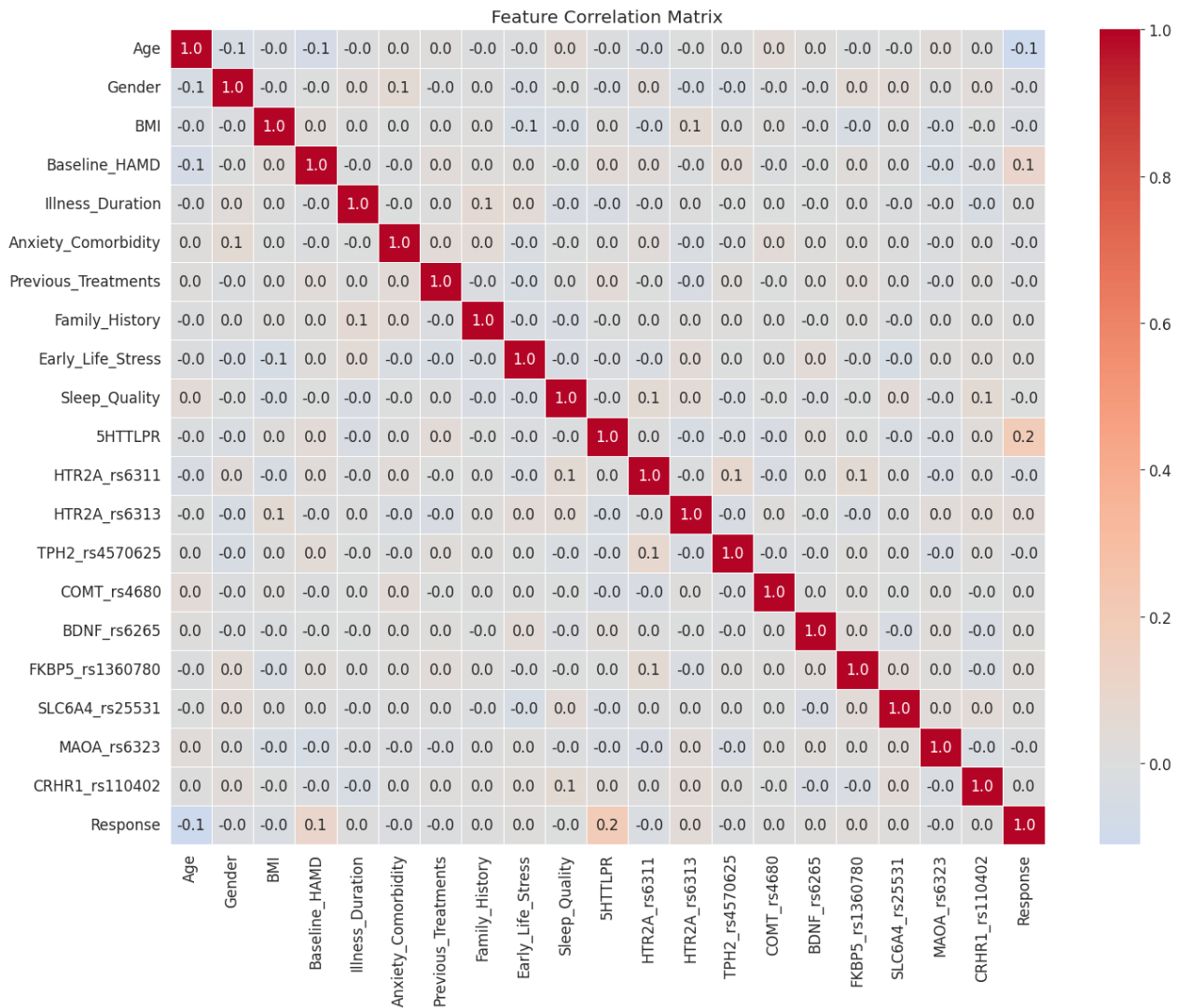


Figure 6. Correlation matrix of clinical and genetic features.

Further, a pair plot analysis was conducted to visualize the joint distributions of selected features and their separability across responders and non-responders. As illustrated in Figure 7, this provided visual evidence that, while individual features did not offer perfect class separation, the combination of features could allow the models to learn complex decision boundaries.

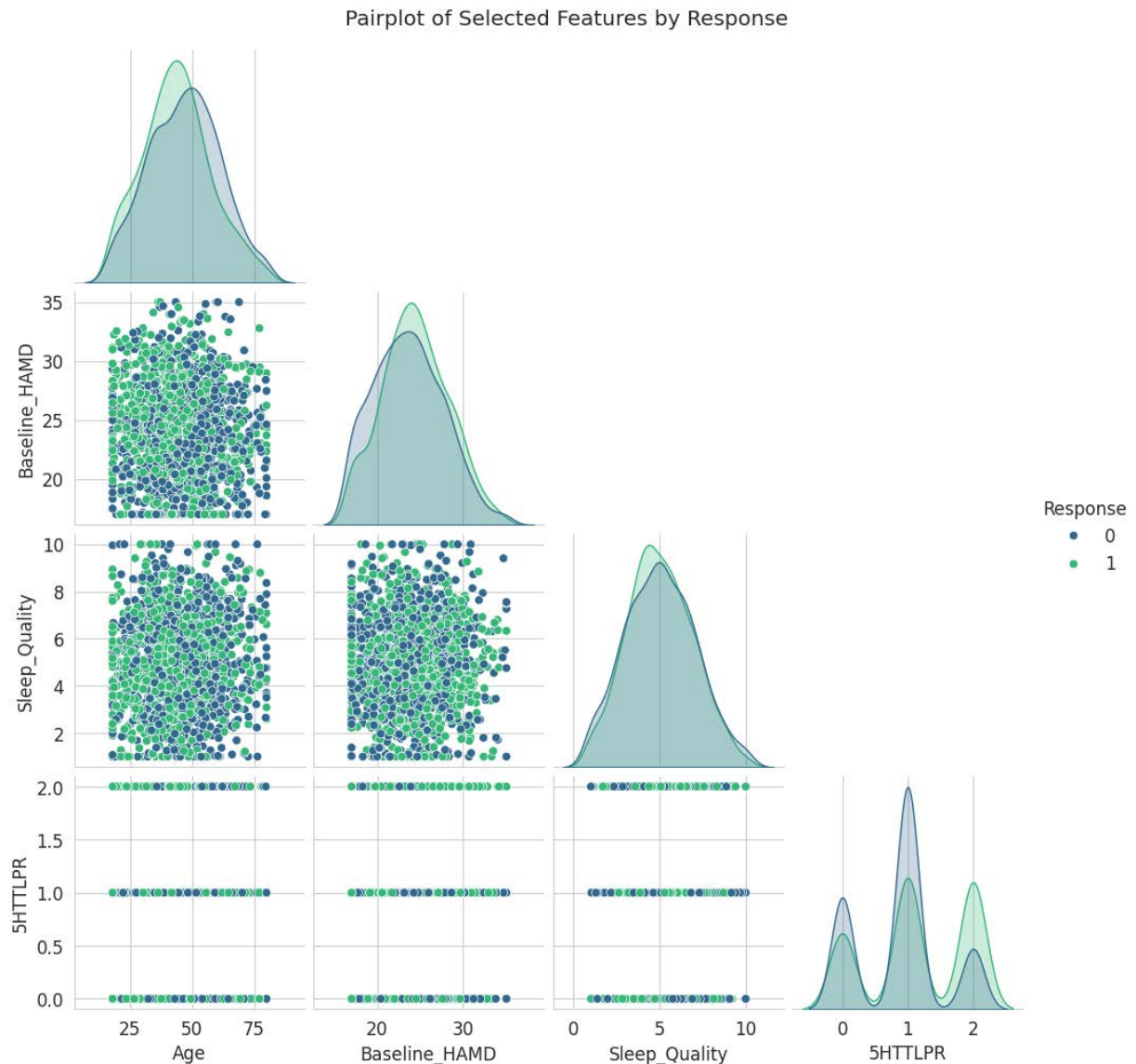


Figure 7. Pair plot of key clinical features across treatment response groups.

3.3. Machine Learning Model Development

Four widely used supervised machine learning classifiers were employed in this study: Random Forest, XGBoost (Gradient Boosting), Support Vector Machine (SVM), and Logistic Regression. These models were selected to provide a balance between interpretability and predictive power, with Random Forest and XGBoost known for their ability to capture complex, non-linear interactions, and Logistic Regression and SVM offering more interpretable, linear decision-making frameworks [43]-[47]. Each model underwent extensive hyperparameter tuning using GridSearchCV with 5-fold cross-validation to identify the best-performing configurations. Cross-validation was essential for ensuring that the models generalized well to unseen data, reducing the risk of overfitting.

Model performance was evaluated using the following key metrics:

- **Accuracy:** The proportion of correctly classified instances.
- **Precision:** The ability to correctly identify true responders without misclassifying non-responders.
- **Recall (Sensitivity):** The ability to correctly detect all true responders, which is crucial in clinical applications to avoid missing patients who might benefit from treatment.
- **F1 Score:** A harmonic mean of precision and recall, providing a balanced performance metric.
- **ROC AUC:** A threshold-independent metric measuring the model's ability to discriminate between responders and non-responders across all decision thresholds.

The multi-metric evaluation ensured that models were not only accurate but also clinically meaningful, with special attention given to recall due to its importance in minimizing missed responders.

4. Results

4.1. Overall Model Performance

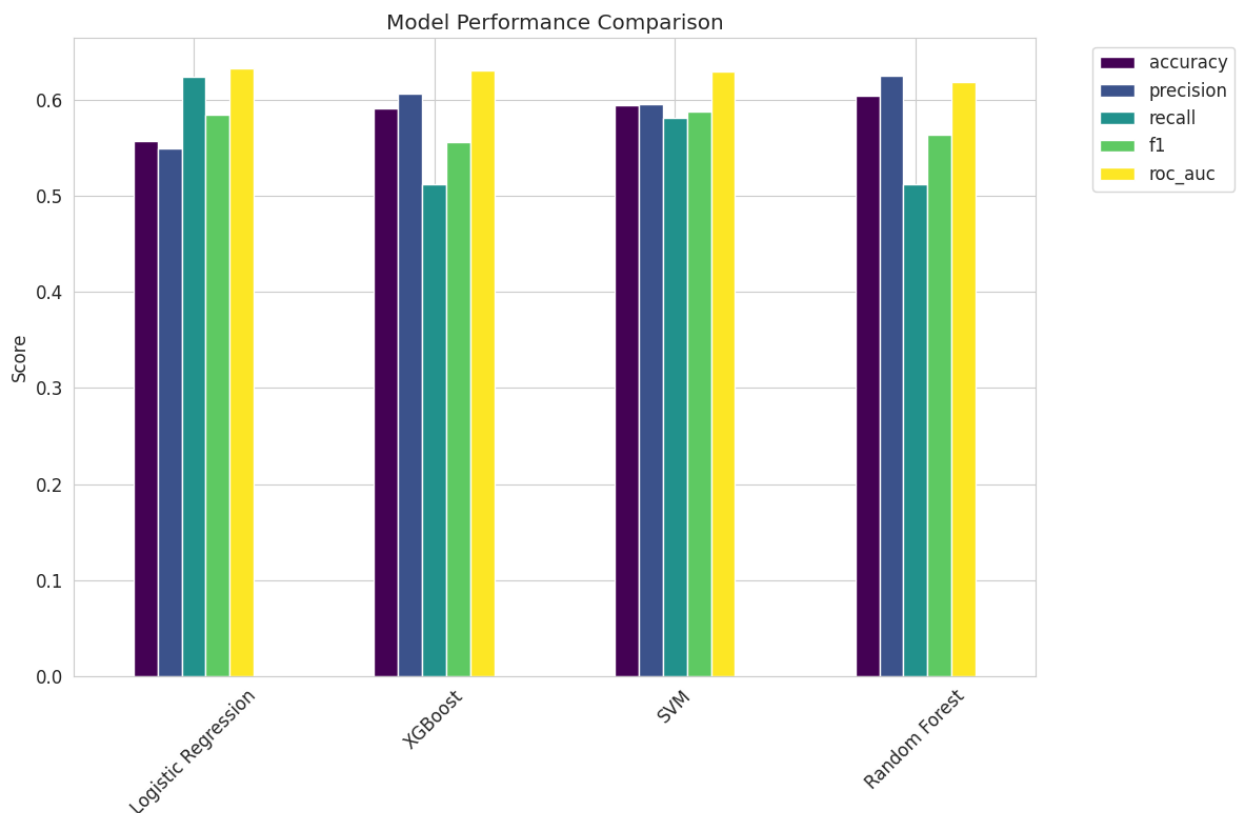


Figure 8. Comparative performance of machine learning models across multiple metrics (Accuracy, Precision, Recall, F1 Score, ROC AUC).

The predictive capability of the four machine learning classifiers—Random Forest, XGBoost, Support Vector Machine (SVM), and Logistic Regression—was rig-

ously assessed using a comprehensive set of performance metrics, including accuracy, precision, recall, F1 score, and ROC AUC. The comparative performance of all models is presented in **Figure 8**. Random Forest achieved the highest overall accuracy (60.4%) and precision (62.6%), indicating that it was more conservative and reliable in correctly identifying true responders while minimizing false positives. However, Logistic Regression demonstrated the highest recall (62.5%) and the best ROC AUC (0.633), showing its superior ability to detect a larger proportion of true responders, a feature of critical clinical importance. XGBoost performed competitively with strong precision (60.7%) but showed slightly lower recall, while SVM provided balanced but modest performance across all metrics. These results indicate that while all models provided clinically useful predictions, Random Forest and Logistic Regression consistently outperformed the others, offering the best trade-offs between precision and sensitivity. The selection of the optimal model may depend on the clinical priority: precision-driven decision-making (Random Forest) versus recall-focused safety nets (Logistic Regression).

4.2 ROC Curve and Discriminative Power

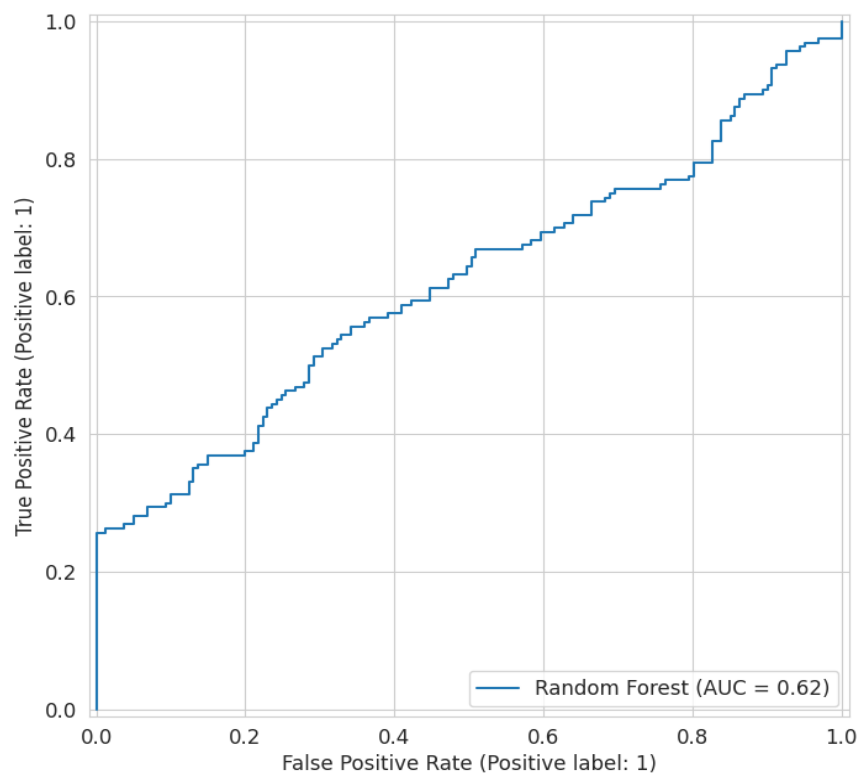


Figure 9. ROC Curve for random forest (AUC = 0.62).

The discriminative ability of each classifier was further examined through Receiver Operating Characteristic (ROC) curves. The individual ROC curves for Random Forest (**Figure 9**), XGBoost (**Figure 10**), Logistic Regression (**Figure 11**), and SVM (**Figure 12**) confirmed moderate but meaningful class separation, with

AUC values ranging from 0.62 to 0.63 across all models.

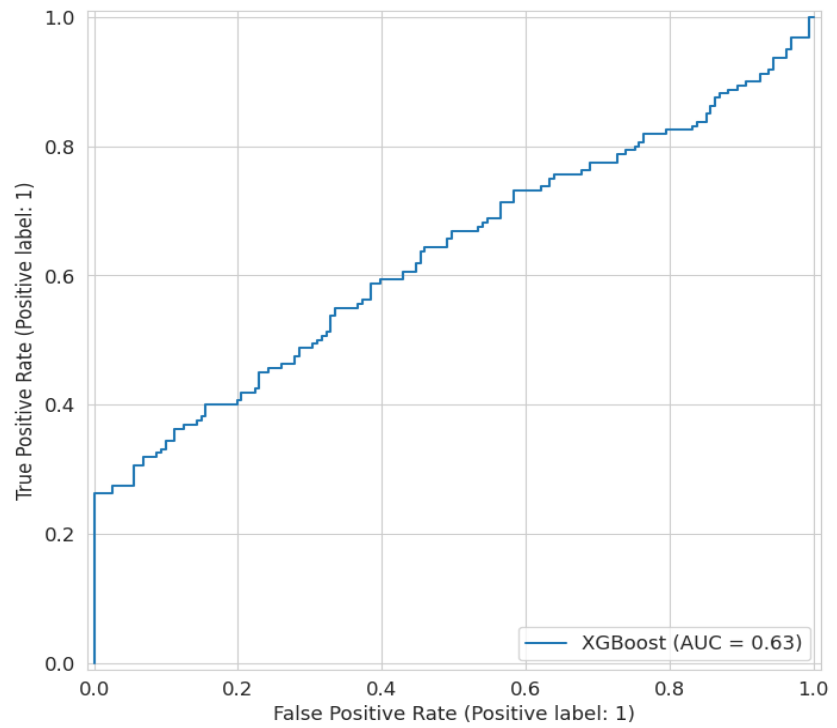


Figure 10. ROC Curve for XGBoost (AUC = 0.63).

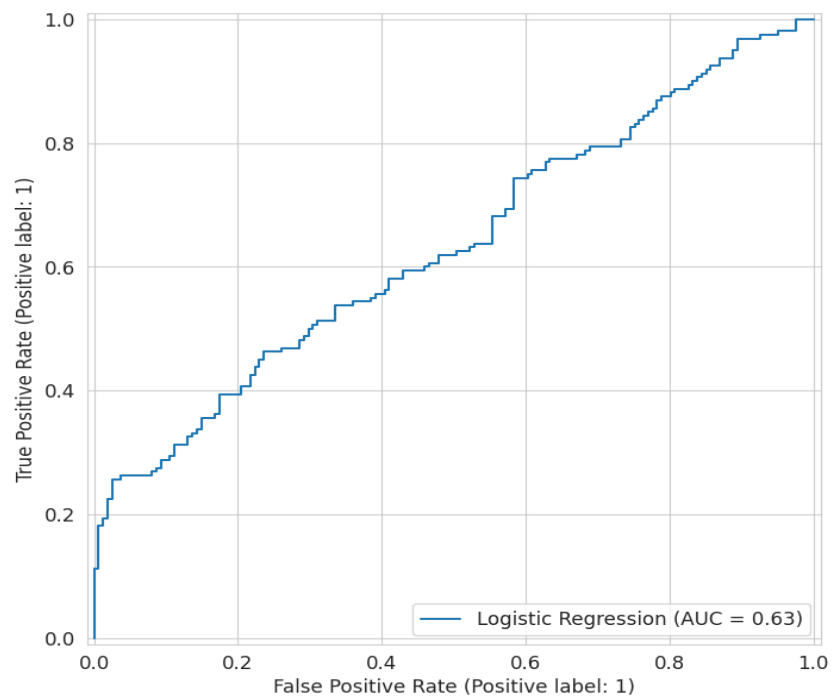


Figure 11. ROC Curve for Logistic Regression (AUC = 0.63).

Although no model achieved perfect discrimination, the consistent AUC scores

indicated that each model performed significantly better than random guessing. Among these, Logistic Regression and XGBoost exhibited slightly more favorable ROC curvature, aligning with their higher recall and sensitivity to true responders.

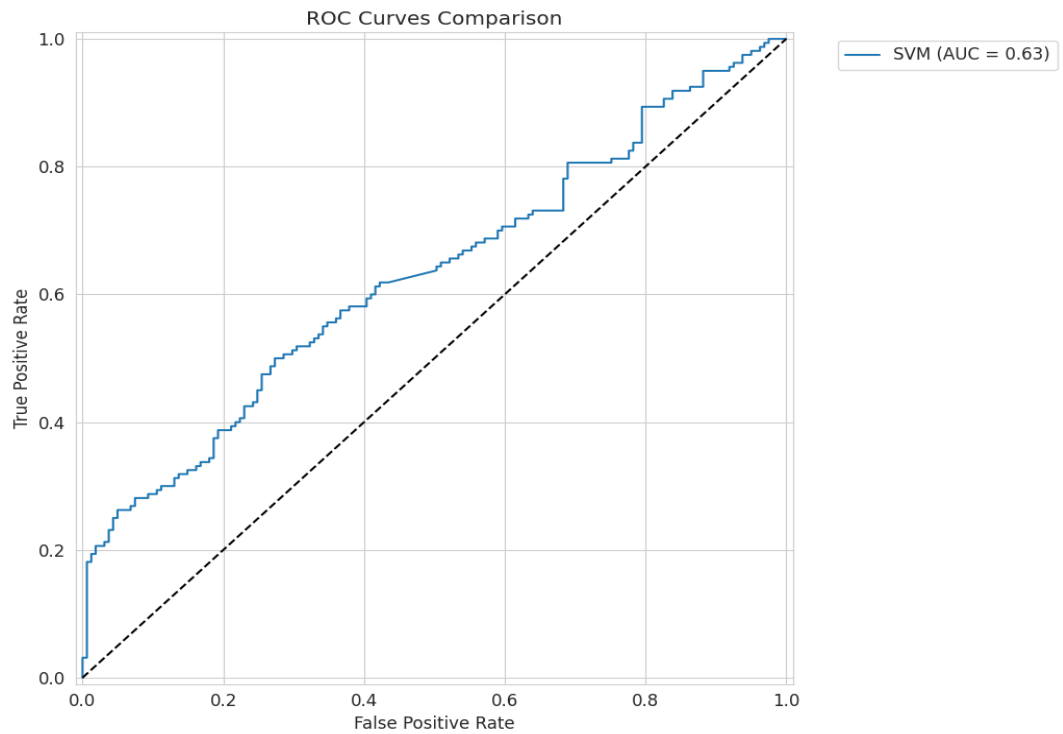


Figure 12. ROC Curve for SVM (AUC = 0.63).

4.3. Precision-Recall Trade-Off Analysis

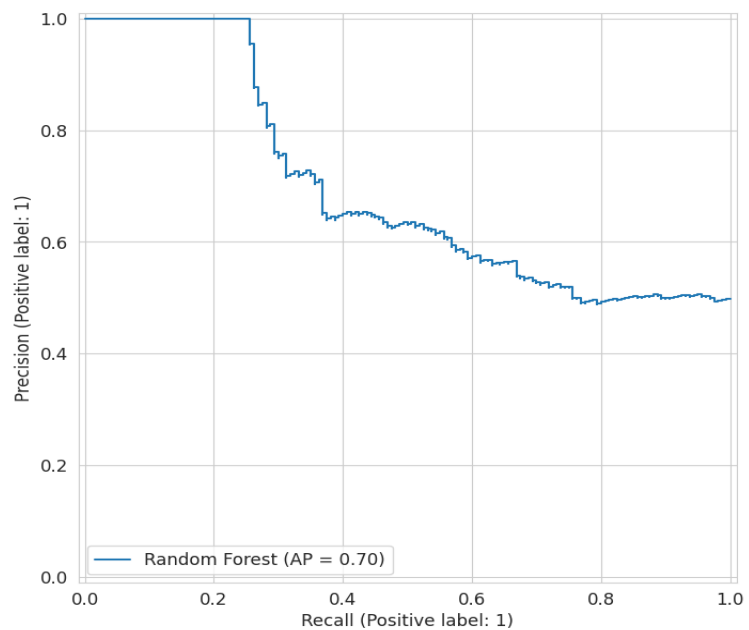


Figure 13. Precision-recall curve for random forest (AP = 0.70).

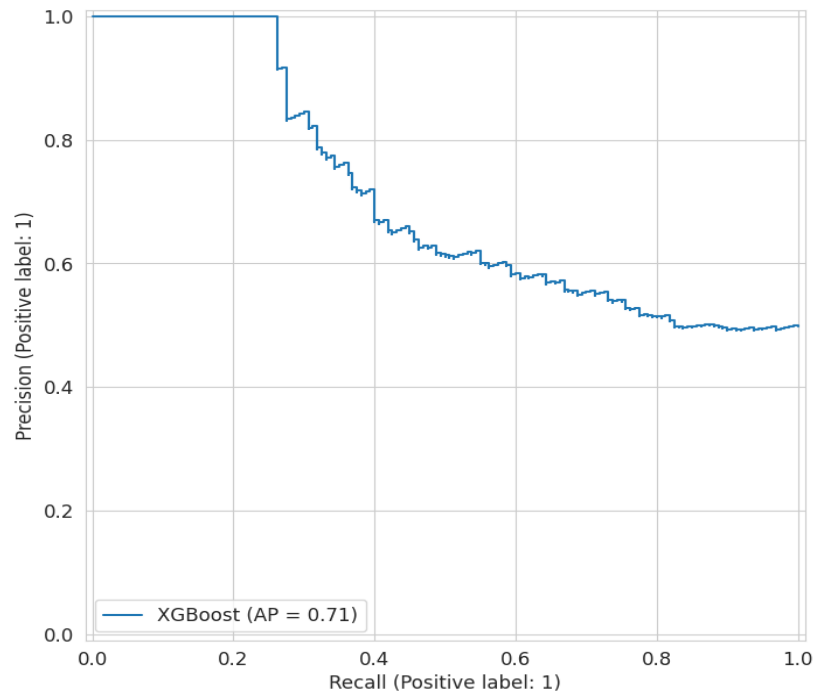


Figure 14. Precision-recall curve for logistic regression (AP = 0.69).

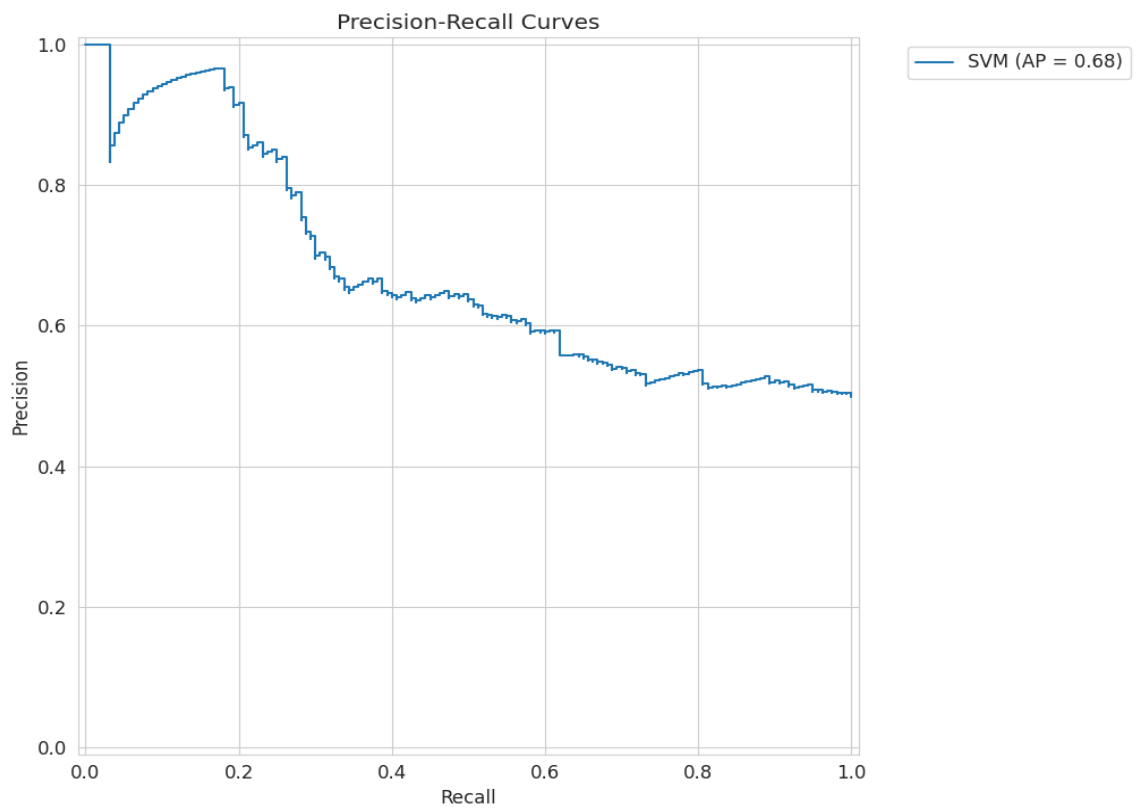


Figure 15. Precision-recall curve for SVM (AP = 0.68).

Precision-Recall (PR) curves were analysed to evaluate the models under the con-

text of imbalanced decision thresholds and real-world clinical settings, where identifying true responders without overwhelming false positives is essential [48]-[50]. The PR curve for Random Forest (Figure 13) showed the highest average precision (AP = 0.70), confirming its strength in confidently predicting responders. Logistic Regression (Figure 14) and SVM (Figure 15) demonstrated competitive but slightly lower AP values of 0.69 and 0.68, respectively. This analysis reinforces the observation that Random Forest offers the best precision-driven decision framework, which is beneficial when aiming to minimize false positives. On the other hand, Logistic Regression maximizes the detection of potential responders, which is often a clinical priority to avoid missing patients who could benefit from treatment [51]-[53].

4.4. Confusion Matrix Insights

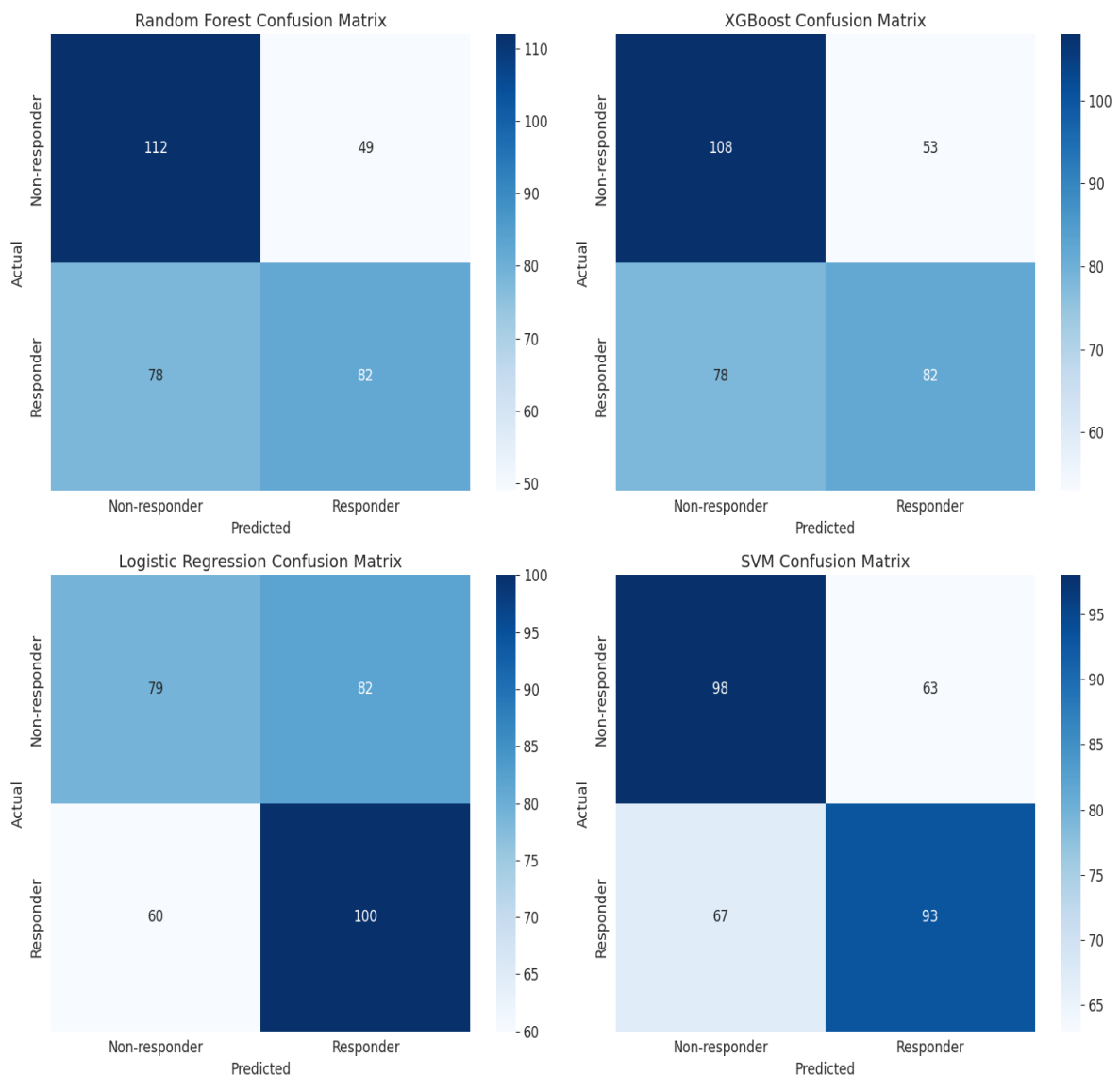


Figure 16. Confusion matrices for random forest, XGBoost, logistic regression, and SVM.

The detailed confusion matrices for all models, presented in **Figure 16**, provided further insight into each classifier's error patterns.

- Random Forest correctly classified 112 non-responders and 82 responders but misclassified 78 responders (false negatives).
- XGBoost produced a comparable error distribution to Random Forest.
- Logistic Regression correctly identified 100 responders, offering the most responder detections and the fewest missed cases.
- SVM showed balanced misclassifications across both classes but slightly favoured non-responders.

The confusion matrix analysis emphasized that Logistic Regression was the most effective in minimizing false negatives, a key requirement in psychiatric treatment planning, where it is safer to over-treat than to miss potential responders.

4.5. Feature Importance and Clinical Interpretability

4.5.1. Random Forest Feature Importance

The Random Forest feature importance rankings (**Figure 17**) revealed that baseline HAMD score, age, early life stress, illness duration, sleep quality, BMI, and the 5HTTLPR genotype were the most impactful predictors. Notably, baseline HAMD score and age emerged as the two most critical clinical features, supporting the clinical intuition that disease severity and patient age significantly influence antidepressant treatment response.

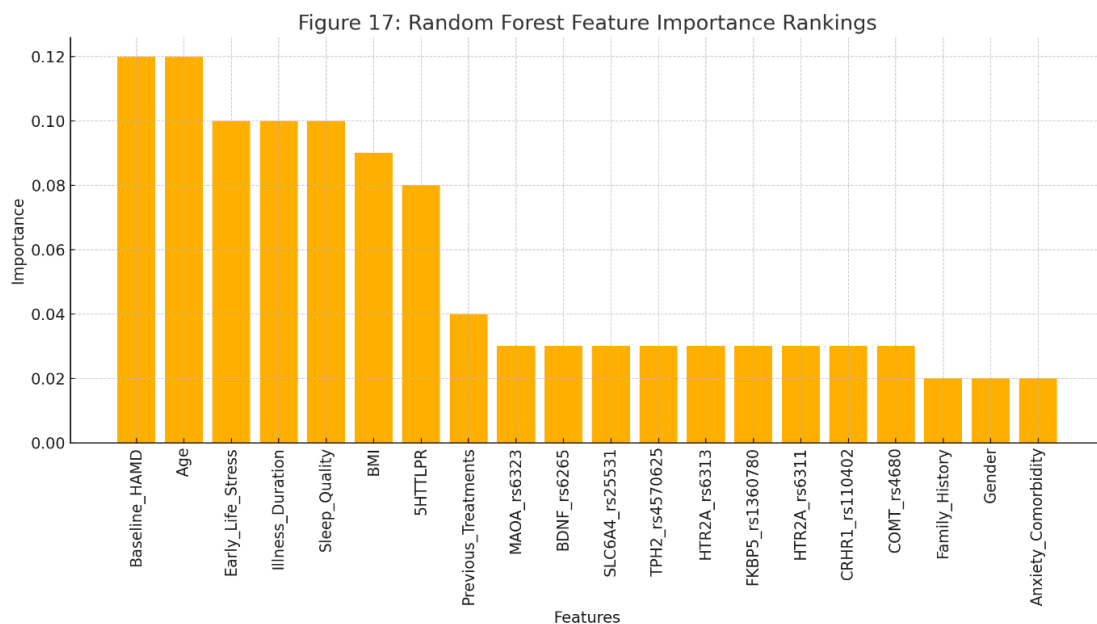


Figure 17. Random forest feature importance rankings showing top clinical and genetic predictors.

4.5.2. Logistic Regression Coefficient Analysis

The coefficient profile of the Logistic Regression model (**Figure 18**) provided clear interpretability of feature contributions. The 5HTTLPR genotype and baseline

HAMD score exhibited strong positive associations with treatment response probability, while prior treatments and anxiety comorbidity showed negative contributions. This alignment with known clinical factors validated the biological plausibility of the model.

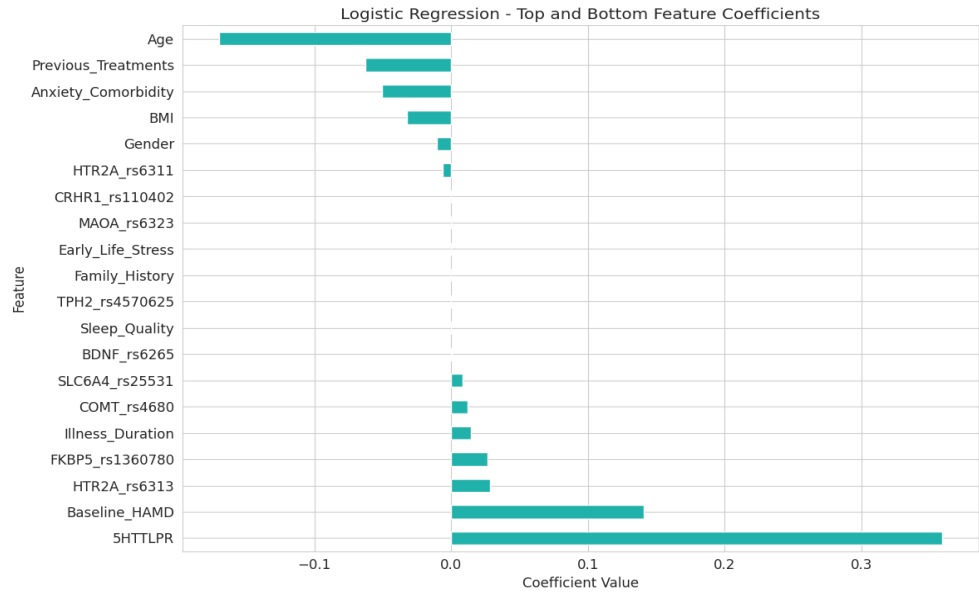


Figure 18. Logistic regression coefficients for top positive and negative predictors.

4.5.3. SHAP Value Interpretation

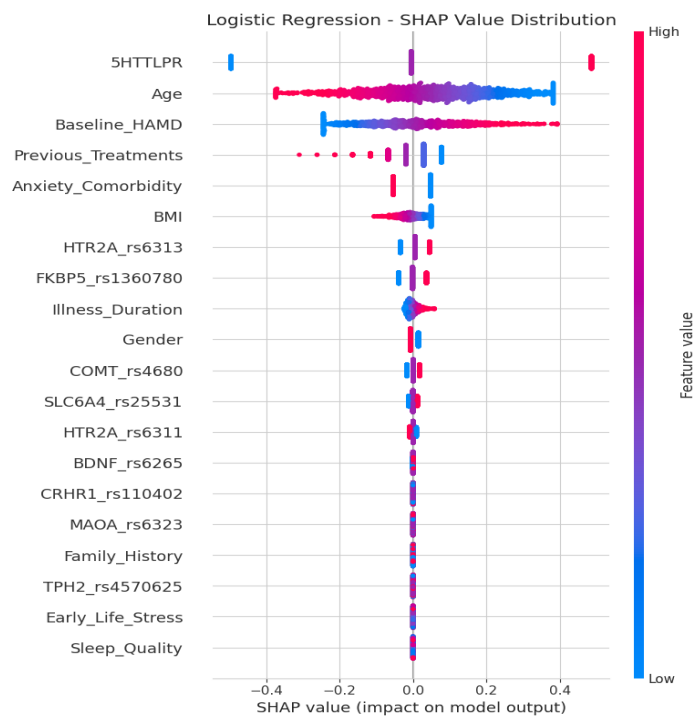


Figure 19. SHAP summary plot displaying the impact of clinical and genetic features on model predictions.

To deeply explore individual prediction contributions, SHAP value analysis was performed (Figure 19). The SHAP summary plot confirmed that 5HTTLPR, baseline HAMD score, and age were the most influential features in driving model outputs. The color-coded distribution highlighted that higher baseline HAMD scores and specific 5HTTLPR genotypes increased the probability of a positive treatment response, while lower scores and certain genetic patterns reduced it.

4.6. Threshold Optimization for Clinical Deployment

Precision, recall, and F1 score were plotted across varying decision thresholds (Figure 20) to identify the clinically optimal classification point. The optimal threshold was determined to be 0.36, where the model achieved maximum recall (96.9%) with acceptable precision (51.8%). This threshold adjustment ensures that very few potential responders are missed, which is critically important in real-world psychiatric treatment, where false negatives can lead to prolonged patient suffering and treatment resistance.

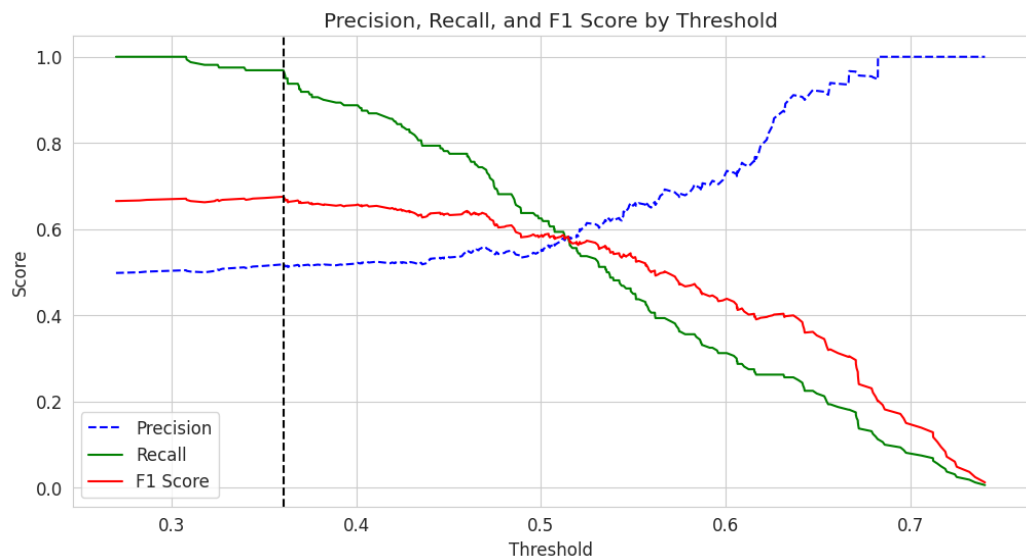


Figure 20. Precision, recall, and F1 score by threshold, optimal decision point at 0.36 for maximum clinical sensitivity.

4.7. Example Patient Prediction

The final trained model was applied to an example patient case to demonstrate clinical usability. The model predicted a treatment response with a probability of 0.648, exceeding the selected optimal threshold of 0.36. Based on this prediction, the patient would be classified as a likely responder to SSRI/SNRI treatment, supporting the practical application of the model in guiding personalized therapeutic decisions.

5. Discussion

This study developed and evaluated machine learning models to predict individ-

ual antidepressant treatment responses by integrating clinical and genetic data, moving toward a more personalized approach to psychiatric care. The results demonstrated that while none of the models achieved perfect predictive power, both Random Forest and Logistic Regression consistently provided clinically meaningful performance, with Random Forest excelling in precision and Logistic Regression offering superior recall.

The Random Forest model's high precision (62.6%) and feature interpretability made it particularly valuable when the clinical priority is to confidently identify responders with minimal false positives. This is crucial in psychiatric medication management, where unnecessary exposure to ineffective drugs can lead to adverse effects, increased patient frustration, and higher dropout rates. In contrast, Logistic Regression provided the highest recall (62.5%), which is critically important when the priority is to avoid missing true responders. In psychiatric settings, a model with higher recall is often preferable because the cost of missing a potential responder is typically greater than the cost of over-treating a non-responder [54]-[56]. The feature importance analysis across models revealed a consistently strong influence of baseline depression severity (HAMD score), age, early life stress, illness duration, and sleep quality on treatment response. These results align with existing clinical literature, which emphasizes that patients with lower baseline severity, shorter illness duration, and less accumulated life stress often exhibit better responses to antidepressant therapies [57]-[61]. Notably, the 5HTTLPR genotype emerged as one of the most important genetic predictors across both Random Forest and SHAP analyses. This finding is consistent with prior pharmacogenomic research suggesting that the 5HTTLPR polymorphism significantly modulates serotonin transporter function and antidepressant efficacy. The SHAP value interpretation further validated the biological relevance of the selected features by quantifying their individual impact on model predictions. The SHAP plots provided actionable insights into how variations in patient-specific factors influence the probability of treatment response, offering a level of interpretability that can support clinicians in shared decision-making with patients. The threshold optimization analysis was a key contribution of this study. By systematically adjusting the decision threshold, we were able to maximize recall (96.9%) while maintaining reasonable precision (51.8%), ensuring that nearly all potential responders were correctly identified. This trade-off is particularly valuable in psychiatry, where treatment delays can exacerbate symptoms and reduce the likelihood of future remission. The decision to adopt a lower threshold reflects a clinical bias toward minimizing false negatives, which in this context represents missing a patient who could significantly benefit from treatment. Despite these promising findings, there are important limitations to consider. First, the dataset was relatively modest in size, and although class balance was achieved, larger and more diverse patient populations would improve model generalizability. Additionally, the balanced dataset achieved through stratified sampling does not reflect real-world response prevalence, potentially limiting external validity. While this approach ensured

equal model exposure during training, future studies should evaluate models on naturally imbalanced cohorts or apply post-training calibration methods to adjust prediction thresholds accordingly. Second, while clinical and genetic features were well-integrated, the absence of multimodal data such as neuroimaging, environmental exposures, and real-time mood tracking may have limited the model's full predictive potential. Future studies should aim to incorporate such high-resolution, longitudinal data streams to enhance predictive accuracy. Additionally, although SHAP values improved interpretability, further validation through prospective clinical studies is required before deploying these models in clinical decision support systems. Furthermore, the genetic features used in this study, while informative, represent a small subset of potential pharmacogenetic markers. A genome-wide association approach could uncover additional genetic variants that may significantly improve prediction power. Another consideration is that while the models performed moderately well, their current level of accuracy and AUC values suggest that machine learning in this domain should be viewed as a clinical support tool rather than a definitive diagnostic instrument. These models can help guide clinical intuition and provide probability-based recommendations, but they should not replace clinical judgment. This study demonstrates the feasibility and value of integrating machine learning with clinical and genetic data to predict antidepressant treatment response. The models developed here provide a foundation for more personalized treatment strategies in psychiatry, with the potential to reduce the duration of ineffective treatment cycles and improve patient outcomes. The findings also reinforce the importance of using interpretable machine learning techniques and carefully tuned decision thresholds to enhance clinical relevance. Future research should focus on external validation, larger datasets, multimodal data integration, and prospective clinical trials to fully realize the potential of machine learning in personalized psychiatry.

6. Conclusion

This study successfully developed and evaluated machine learning models to predict individual responses to antidepressant treatment using an integrated approach that combined clinical and genetic data. By applying four supervised learning algorithms—Random Forest, XGBoost, Support Vector Machine (SVM), and Logistic Regression—we demonstrated that it is feasible to moderately predict treatment outcomes before initiating pharmacotherapy. Among the models, Random Forest and Logistic Regression emerged as the most clinically valuable, offering the best balance between precision, recall, and overall model robustness. The findings highlight that key clinical features, including baseline depression severity, age, early life stress, illness duration, and sleep quality, are powerful predictors of treatment response. Additionally, the 5HTTLPR genotype consistently contributed to improved model performance, reinforcing its importance as a potential genetic biomarker in antidepressant pharmacogenomics. The application of SHAP values provided deeper interpretability, enhancing the transparency of

the predictive models and offering clinical practitioners a clearer understanding of the factors driving each prediction. One of the key strengths of this study was the application of threshold optimization, which significantly improved the model's clinical utility by prioritizing responder detection while maintaining acceptable precision. This adjustment is essential for real-world psychiatric applications, where the cost of missing a potential responder can be substantial [62]-[65]. While the models achieved moderate performance, they offer a promising step toward personalized psychiatry [66]-[68]. Future research should focus on expanding datasets, incorporating additional data modalities such as neuroimaging and longitudinal monitoring, and validating the models in external and prospective clinical settings. Ultimately, the integration of machine learning into psychiatric decision-making has the potential to reduce the reliance on trial-and-error prescribing, shorten treatment cycles, and improve patient outcomes by delivering more tailored, data-driven therapeutic strategies.

Conflicts of Interest

The authors declare no conflicts of interest.

References

- [1] Proudman, D., Greenberg, P. and Nellesen, D. (2021) The Growing Burden of Major Depressive Disorders (MDD): Implications for Researchers and Policy Makers. *Pharmacoeconomics*, **39**, 619-625. <https://doi.org/10.1007/s40273-021-01040-7>
- [2] Santomauro, D.F., Vos, T., Whiteford, H.A., Chisholm, D., Saxena, S. and Ferrari, A.J. (2024) Service Coverage for Major Depressive Disorder: Estimated Rates of Minimally Adequate Treatment for 204 Countries and Territories in 2021. *The Lancet Psychiatry*, **11**, 1012-1021. [https://doi.org/10.1016/s2215-0366\(24\)00317-1](https://doi.org/10.1016/s2215-0366(24)00317-1)
- [3] Moitra, M., Santomauro, D., Collins, P.Y., Vos, T., Whiteford, H., Saxena, S., *et al.* (2022) The Global Gap in Treatment Coverage for Major Depressive Disorder in 84 Countries from 2000-2019: A Systematic Review and Bayesian Meta-Regression Analysis. *PLOS Medicine*, **19**, e1003901. <https://doi.org/10.1371/journal.pmed.1003901>
- [4] Santomauro, D.F., Mantilla Herrera, A.M., Shadid, J., Zheng, P., Ashbaugh, C., Pigott, D.M., *et al.* (2021) Global Prevalence and Burden of Depressive and Anxiety Disorders in 204 Countries and Territories in 2020 due to the COVID-19 Pandemic. *The Lancet*, **398**, 1700-1712. [https://doi.org/10.1016/s0140-6736\(21\)02143-7](https://doi.org/10.1016/s0140-6736(21)02143-7)
- [5] Locher, C., Koechlin, H., Zion, S.R., Werner, C., Pine, D.S., Kirsch, I., *et al.* (2017) Efficacy and Safety of Selective Serotonin Reuptake Inhibitors, Serotonin-Norepinephrine Reuptake Inhibitors, and Placebo for Common Psychiatric Disorders among Children and Adolescents: A Systematic Review and META-Analysis. *JAMA Psychiatry*, **74**, 1011-1020. <https://doi.org/10.1001/jamapsychiatry.2017.2432>
- [6] Dell'Osso, B., Buoli, M., Baldwin, D.S. and Altamura, A.C. (2009) Serotonin Norepinephrine Reuptake Inhibitors (SNRIs) in Anxiety Disorders: A Comprehensive Review of Their Clinical Efficacy. *Human Psychopharmacology: Clinical and Experimental*, **25**, 17-29. <https://doi.org/10.1002/hup.1074>
- [7] Montano, C.B., Jackson, W.C., Vanacore, D. and Weisler, R. (2023) Considerations When Selecting an Antidepressant: A Narrative Review for Primary Care Providers Treating Adults with Depression. *Postgraduate Medicine*, **135**, 449-465.

- <https://doi.org/10.1080/00325481.2023.2189868>
- [8] Katon, W.J. (2011) Epidemiology and Treatment of Depression in Patients with Chronic Medical Illness. *Dialogues in Clinical Neuroscience*, **13**, 7-23. <https://doi.org/10.31887/dcns.2011.13.1/wkaton>
- [9] Vos, T., Haby, M.M., Barendregt, J.J., Kruijshaar, M., Corry, J. and Andrews, G. (2004) The Burden of Major Depression Avoidable by Longer-Term Treatment Strategies. *Archives of General Psychiatry*, **61**, 1097-1103. <https://doi.org/10.1001/archpsyc.61.11.1097>
- [10] Kumar, K.P.S., Srivastava, S., Paswan, S. and Dutta, A.S. (2012) Depression-Symptoms, Causes, Medications and Therapies. *The Pharma Innovation*, **1**, 37-51.
- [11] McIntyre, R.S. and O'Donovan, C. (2004) The Human Cost of Not Achieving Full Remission in Depression. *Canadian Journal of Psychiatry*, **49**, 10S-16S.
- [12] Zeier, Z., Carpenter, L.L., Kalin, N.H., Rodriguez, C.I., McDonald, W.M., Widge, A.S., *et al.* (2018) Clinical Implementation of Pharmacogenetic Decision Support Tools for Antidepressant Drug Prescribing. *American Journal of Psychiatry*, **175**, 873-886. <https://doi.org/10.1176/appi.ajp.2018.17111282>
- [13] Eap, C.B., Gründer, G., Baumann, P., Ansermot, N., Conca, A., Corruble, E., *et al.* (2021) Tools for Optimising Pharmacotherapy in Psychiatry (Therapeutic Drug Monitoring, Molecular Brain Imaging and Pharmacogenetic Tests): Focus on Antidepressants. *The World Journal of Biological Psychiatry*, **22**, 561-628. <https://doi.org/10.1080/15622975.2021.1878427>
- [14] Huang, M. and Pan, H. (2023) Pharmacogenomic Profiling to Tailor Antidepressant Therapy: Improving Treatment Outcomes and Reducing Adverse Drug Reactions in Major Depressive Disorder. *SHIFAA*, **2023**, 19-31. <https://doi.org/10.70470/shifaa/2023/003>
- [15] Maj, M., Stein, D.J., Parker, G., Zimmerman, M., Fava, G.A., De Hert, M., *et al.* (2020) The Clinical Characterization of the Adult Patient with Depression Aimed at Personalization of Management. *World Psychiatry*, **19**, 269-293. <https://doi.org/10.1002/wps.20771>
- [16] Milic, J., Vucurovic, M., Jovic, D., Stankovic, V., Grego, E., Jankovic, S., *et al.* (2025) Exploring the Potential of Precision Medicine in Neuropsychiatry: A Commentary on New Insights for Tailored Treatments Based on Genetic, Environmental, and Lifestyle Factors. *Genes*, **16**, Article 371. <https://doi.org/10.3390/genes16040371>
- [17] Zanardi, R., Prestifilippo, D., Fabbri, C., Colombo, C., Maron, E. and Serretti, A. (2020) Precision Psychiatry in Clinical Practice. *International Journal of Psychiatry in Clinical Practice*, **25**, 19-27. <https://doi.org/10.1080/13651501.2020.1809680>
- [18] Manchia, M., Pisanu, C., Squassina, A. and Carpiniello, B. (2020) Challenges and Future Prospects of Precision Medicine in Psychiatry. *Pharmacogenomics and Personalized Medicine*, **13**, 127-140. <https://doi.org/10.2147/pgpm.s198225>
- [19] Comai, S., Manchia, M., Bosia, M., Miola, A., Poletti, S., Benedetti, F., *et al.* (2025) Moving toward Precision and Personalized Treatment Strategies in Psychiatry. *International Journal of Neuropsychopharmacology*, **28**, pyaf025. <https://doi.org/10.1093/ijnp/pyaf025>
- [20] Gandal, M.J., Leppa, V., Won, H., Parikshak, N.N. and Geschwind, D.H. (2016) The Road to Precision Psychiatry: Translating Genetics into Disease Mechanisms. *Nature Neuroscience*, **19**, 1397-1407. <https://doi.org/10.1038/nn.4409>
- [21] Orphanidou, C. and Wong, D. (2017) Machine Learning Models for Multidimensional Clinical Data. In: Khan, S., Zomaya, A. and Abbas, A., Eds., *Handbook of*

- Large-Scale Distributed Computing in Smart Healthcare*, Springer, 177-216.
https://doi.org/10.1007/978-3-319-58280-1_8
- [22] Mirza, B., Wang, W., Wang, J., Choi, H., Chung, N.C. and Ping, P. (2019) Machine Learning and Integrative Analysis of Biomedical Big Data. *Genes*, **10**, Article 87. <https://doi.org/10.3390/genes10020087>
- [23] Kiranyaz, S., Ince, T. and Gabbouj, M. (2014) *Multidimensional Particle Swarm Optimization for Machine Learning and Pattern Recognition*. Springer.
- [24] Ogbu, A.D., Iwe, K.A., Ozowe, W. and Ikevuje, A.H. (2024) Advances in Machine Learning-Driven Pore Pressure Prediction in Complex Geological Settings. *Computer Science & IT Research Journal*, **5**, 1648-1665. <https://doi.org/10.51594/csitrj.v5i7.1350>
- [25] Chekroud, A.M., Bondar, J., Delgadillo, J., Doherty, G., Wasil, A., Fokkema, M., *et al.* (2021) The Promise of Machine Learning in Predicting Treatment Outcomes in Psychiatry. *World Psychiatry*, **20**, 154-170. <https://doi.org/10.1002/wps.20882>
- [26] Bobo, W.V., Van Ommeren, B. and Athreya, A.P. (2022) Machine Learning, Pharmacogenomics, and Clinical Psychiatry: Predicting Antidepressant Response in Patients with Major Depressive Disorder. *Expert Review of Clinical Pharmacology*, **15**, 927-944. <https://doi.org/10.1080/17512433.2022.2112949>
- [27] Athreya, A.P., Iyer, R., Wang, L., Weinshilboum, R.M. and Bobo, W.V. (2019) Integration of Machine Learning and Pharmacogenomic Biomarkers for Predicting Response to Antidepressant Treatment: Can Computational Intelligence Be Used to Augment Clinical Assessments? *Pharmacogenomics*, **20**, 983-988. <https://doi.org/10.2217/pgs-2019-0119>
- [28] Rutledge, R.B., Chekroud, A.M. and Huys, Q.J. (2019) Machine Learning and Big Data in Psychiatry: Toward Clinical Applications. *Current Opinion in Neurobiology*, **55**, 152-159. <https://doi.org/10.1016/j.conb.2019.02.006>
- [29] Lin, E., Lin, C. and Lane, H. (2021) Machine Learning and Deep Learning for the Pharmacogenomics of Antidepressant Treatments. *Clinical Psychopharmacology and Neuroscience*, **19**, 577-588. <https://doi.org/10.9758/cpn.2021.19.4.577>
- [30] Nikolodimou, V. and Agapow, P. (2020) Using Machine-Learning Techniques to Identify Responders vs. Non-Responders in Randomized Clinical Trials. medRxiv.
- [31] Bai, X., Feng, M., Ma, W. and Wang, S. (2025) Predicting the Efficacy of Bevacizumab on Peritumoral Edema Based on Imaging Features and Machine Learning. *Scientific Reports*, **15**, Article No. 15990. <https://doi.org/10.1038/s41598-025-00758-0>
- [32] Georgoula, M. (2024) Application of Machine Learning Methods to Predict Critical Multiple Myeloma Events. Ph.D. Thesis, School of Electrical and Computer Engineering.
- [33] McShane, L.M., Cavenagh, M.M., Lively, T.G., Eberhard, D.A., Bigbee, W.L., Williams, P.M., *et al.* (2013) Criteria for the Use of Omics-Based Predictors in Clinical Trials: Explanation and Elaboration. *BMC Medicine*, **11**, Article No. 220. <https://doi.org/10.1186/1741-7015-11-220>
- [34] Saito, T. and Rehmsmeier, M. (2015) The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets. *PLOS ONE*, **10**, e0118432. <https://doi.org/10.1371/journal.pone.0118432>
- [35] Davis, J. and Goadrich, M. (2006) The Relationship between Precision-Recall and ROC Curves. *Proceedings of the 23rd International Conference on Machine Learning—ICML'06*, Pittsburgh, 25-29 June 2006, 233-240. <https://doi.org/10.1145/1143844.1143874>

- [36] Tharwat, A. (2020) Classification Assessment Methods. *Applied Computing and Informatics*, **17**, 168-192. <https://doi.org/10.1016/j.aci.2018.08.003>
- [37] Sherbourne, C.D., Hays, R.D. and Wells, K.B. (1995) Personal and Psychosocial Risk Factors for Physical and Mental Health Outcomes and Course of Depression among Depressed Patients. *Journal of Consulting and Clinical Psychology*, **63**, 345-355. <https://doi.org/10.1037//0022-006x.63.3.345>
- [38] Shi, L., Lu, Z., Que, J., Huang, X., Liu, L., Ran, M., *et al.* (2020) Prevalence of and Risk Factors Associated with Mental Health Symptoms among the General Population in China during the Coronavirus Disease 2019 Pandemic. *JAMA Network Open*, **3**, e2014053. <https://doi.org/10.1001/jamanetworkopen.2020.14053>
- [39] Hendrie, H.C., Lindgren, D., Hay, D.P., Lane, K.A., Gao, S., Purnell, C., *et al.* (2013) Comorbidity Profile and Healthcare Utilization in Elderly Patients with Serious Mental Illnesses. *The American Journal of Geriatric Psychiatry*, **21**, 1267-1276. <https://doi.org/10.1016/j.jagp.2013.01.056>
- [40] Delgadillo, J., Moreea, O. and Lutz, W. (2016) Different People Respond Differently to Therapy: A Demonstration Using Patient Profiling and Risk Stratification. *Behaviour Research and Therapy*, **79**, 15-22. <https://doi.org/10.1016/j.brat.2016.02.003>
- [41] Pearson, R. (2011) Exploring Data. Oxford University Press.
- [42] Vesanto, J. and Alhoniemi, E. (2000) Clustering of the Self-Organizing Map. *IEEE Transactions on Neural Networks*, **11**, 586-600. <https://doi.org/10.1109/72.846731>
- [43] Cifci, A. (2025) Interpretable Prediction of a Decentralized Smart Grid Based on Machine Learning and Explainable Artificial Intelligence. *IEEE Access*, **13**, 36285-36305. <https://doi.org/10.1109/access.2025.3543759>
- [44] Demir, S. and Sahin, E.K. (2025) An Innovative Machine Learning Approach for Slope Stability Prediction by Combining Shap Interpretability and Stacking Ensemble Learning. *Environmental Science and Pollution Research*, **32**, 12827-12843. <https://doi.org/10.1007/s11356-025-36406-3>
- [45] Elhishi, S., Elashry, A.M. and El-Metwally, S. (2023) Unboxing Machine Learning Models for Concrete Strength Prediction Using Xai. *Scientific Reports*, **13**, Article No. 19892. <https://doi.org/10.1038/s41598-023-47169-7>
- [46] Eskandari, H., Saadatmand, H., Ramzan, M. and Mousapour, M. (2024) Innovative Framework for Accurate and Transparent Forecasting of Energy Consumption: A Fusion of Feature Selection and Interpretable Machine Learning. *Applied Energy*, **366**, Article ID: 123314. <https://doi.org/10.1016/j.apenergy.2024.123314>
- [47] Chinnaraju, A. (2025) Explainable AI (XAI) for Trustworthy and Transparent Decision-Making: A Theoretical Framework for AI Interpretability. *World Journal of Advanced Engineering Technology and Sciences*, **14**, 170-207. <https://doi.org/10.30574/wjaets.2025.14.3.0106>
- [48] Gholampour, S. (2024) Impact of Nature of Medical Data on Machine and Deep Learning for Imbalanced Datasets: Clinical Validity of SMOTE Is Questionable. *Machine Learning and Knowledge Extraction*, **6**, 827-841. <https://doi.org/10.3390/make6020039>
- [49] Emi-Johnson, O., Nkrumah, K., Folasole, A. and Kolade Amusa, T. (2023) Optimizing Machine Learning for Imbalanced Classification: Applications in Us Healthcare, Finance, and Security. *International Journal of Engineering Technology Research & Management*, **7**, 89-106.
- [50] Kunti, R. (2021) Classification of Biomedical Data with Class Imbalance. Ph.D. Thesis, Kanazawa University.
- [51] Shipe, M.E., Deppen, S.A., Farjah, F. and Grogan, E.L. (2019) Developing Prediction

- Models for Clinical Use Using Logistic Regression: An Overview. *Journal of Thoracic Disease*, **11**, S574-S584. <https://doi.org/10.21037/jtd.2019.01.25>
- [52] Fleming, T.R. (2011) Addressing Missing Data in Clinical Trials. *Annals of Internal Medicine*, **154**, 113-117. <https://doi.org/10.7326/0003-4819-154-2-201101180-00010>
- [53] O’Kelly, M. and Ratitch, B. (2014) *Clinical Trials with Missing Data*. Wiley. <https://doi.org/10.1002/9781118762516>
- [54] Jarrett, R.B. and Thase, M.E. (2010) Comparative Efficacy and Durability of Continuation Phase Cognitive Therapy for Preventing Recurrent Depression: Design of a Double-Blinded, Fluoxetine- and Pill Placebo-Controlled, Randomized Trial with 2-Year Follow-up. *Contemporary Clinical Trials*, **31**, 355-377. <https://doi.org/10.1016/j.cct.2010.04.004>
- [55] Nobre, A.C. (2020) Cognitive Neuroscience. In: Geddes, J.R., *et al.*, Eds., *New Oxford Textbook of Psychiatry*, Oxford University Press, 154-169. <https://doi.org/10.1093/med/9780198713005.003.0017>
- [56] Keeling, N. (2016) Payer Perspectives on Preemptive Pharmacogenetic Testing. <https://egrove.olemiss.edu/etd/735/>
- [57] Whyte, E.M., Dew, M.A., Gildengers, A., Lenze, E.J., Bharucha, A., Mulsant, B.H., *et al.* (2004) Time Course of Response to Antidepressants in Late-Life Major Depression. *Drugs & Aging*, **21**, 531-554. <https://doi.org/10.2165/00002512-200421080-00004>
- [58] Fornaro, M., Anastasia, A., Novello, S., Fusco, A., Pariano, R., De Berardis, D., *et al.* (2019) The Emergence of Loss of Efficacy during Antidepressant Drug Treatment for Major Depressive Disorder: An Integrative Review of Evidence, Mechanisms, and Clinical Implications. *Pharmacological Research*, **139**, 494-502. <https://doi.org/10.1016/j.phrs.2018.10.025>
- [59] Haddad, P.M., Talbot, P.S., Anderson, I.M. and McAllister-Williams, R.H. (2015) Managing Inadequate Antidepressant Response in Depressive Illness. *British Medical Bulletin*, **115**, 183-201. <https://doi.org/10.1093/bmb/ldv034>
- [60] Sartorius, N., Baghai, T.C., Baldwin, D.S., Barrett, B., Brand, U., Fleischhacker, W., *et al.* (2007) Antidepressant Medications and Other Treatments of Depressive Disorders: A CINP Task Force Report Based on a Review of Evidence. *The International Journal of Neuropsychopharmacology*, **10**, S1-S207. <https://doi.org/10.1017/s1461145707008255>
- [61] Lapolla, T., Saltiel, P. and Silvershein, D. (2015) Major Depressive Disorder: Mechanism-Based Prescribing for Personalized Medicine. *Neuropsychiatric Disease and Treatment*, **11**, 875-888. <https://doi.org/10.2147/ndt.s73261>
- [62] Koch, E., Pardiñas, A.F., O’Connell, K.S., Selvaggi, P., Camacho Collados, J., Babic, A., *et al.* (2024) How Real-World Data Can Facilitate the Development of Precision Medicine Treatment in Psychiatry. *Biological Psychiatry*, **96**, 543-551. <https://doi.org/10.1016/j.biopsych.2024.01.001>
- [63] Baldwin, H., Loebel-Davidsohn, L., Oliver, D., Salazar de Pablo, G., Stahl, D., Riper, H., *et al.* (2022) Real-World Implementation of Precision Psychiatry: A Systematic Review of Barriers and Facilitators. *Brain Sciences*, **12**, Article 934. <https://doi.org/10.3390/brainsci12070934>
- [64] Roy-Byrne, P.P., Sherbourne, C.D., Craske, M.G., Stein, M.B., Katon, W., Sullivan, G., *et al.* (2003) Moving Treatment Research from Clinical Trials to the Real World. *Psychiatric Services*, **54**, 327-332. <https://doi.org/10.1176/appi.ps.54.3.327>
- [65] Simon, G. (1995) Cost-Effectiveness Comparisons Using “Real World” Randomized Trials: The Case of New Antidepressant Drugs. *Journal of Clinical Epidemiology*, **48**,

- 363-373. [https://doi.org/10.1016/0895-4356\(94\)00141-c](https://doi.org/10.1016/0895-4356(94)00141-c)
- [66] Meehan, A.J., Lewis, S.J., Fazel, S., Fusar-Poli, P., Steyerberg, E.W., Stahl, D., *et al.* (2022) Clinical Prediction Models in Psychiatry: A Systematic Review of Two Decades of Progress and Challenges. *Molecular Psychiatry*, **27**, 2700-2708. <https://doi.org/10.1038/s41380-022-01528-4>
- [67] Bzdok, D. and Meyer-Lindenberg, A. (2018) Machine Learning for Precision Psychiatry: Opportunities and Challenges. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, **3**, 223-230. <https://doi.org/10.1016/j.bpsc.2017.11.007>
- [68] Fusar-Poli, P., Hijazi, Z., Stahl, D. and Steyerberg, E.W. (2018) The Science of Prognosis in Psychiatry. *JAMA Psychiatry*, **75**, 1289-1297. <https://doi.org/10.1001/jamapsychiatry.2018.2530>