Comparative Analysis of Psychometric data for assessing mental health testing needs using Machine Learning Models

Sahil R Jain
Computer Science and Engineering
Ramrao Adik Institute of Technology
Mumbai, India
sahillljain6@gmail.com

Aadit Rajendra Chaurasia
Computer Science and Engineering
Ramrao Adik Institute of Technology
Mumbai, India
kumud2196@gmail.com

Mosam Biswas
Computer Science and Engineering
Ramrao Adik Institute of Technology
Mumbai, India
Mosambiswas999@gmail.com

Digamabar Puri
Computer Science and Engineering
Ramrao Adik Institute of Technology
Mumbai, India
digambar.puri@rait.ac.in

Varad Kotalwar
Computer Science and Engineering
Ramrao Adik Institute of Technology
Mumbai, India
Varadk18@gmail.com

Gunjan Patil
Computer Science and Engineering
Ramrao Adik Institute of Technology
Mumbai, India
gunjanpatil161204@gmail.com

***Abstract*—The prevalence of psychological conditions such as depression, anxiety, and stress continues to increase globally, yet obtaining a timely diagnosis for many individuals remains a challenge due to limited and cumbersome screening resources. Although psychometric questionnaires are standard in mental health evaluations, determining who actually needs comprehensive testing is not straightforward. This work presents a comparative assessment of several supervised machine learning algorithms—Logistic Regression, Random Forest, support vector machine, XGBoost, and LightGBM—for predicting the necessity of further mental health evaluation based on psychometric survey responses. The analysis relies on data from established self-report tools like the PHQ-9 and GAD-7. Using systematic methodologies and robust model assessment protocols, this study aims to inform the development of smarter mental health triage systems, integrating machine learning with existing mental health workflows to optimize efficiency and maintain diagnostic reliability.**

***Index Terms*—Mental health, Depression, Anxiety, Stress, Machine Learning (ML), Logistic Regression (LR), Psychometric Testing, Support Vector Machine (SVM), Random Forest (RF), Model Evaluation, Decision Support Systems (DSS).**

I. Introduction

Rising rates of mental health difficulties underline the need for accessible and efficient screening processes. Advances in artificial intelligence, particularly machine learning, open new opportunities to enhance the effectiveness of mental health diagnostics [1]. Psychometric testing offers quantitative insights into mental well-being, and machine learning models can interpret intricate patterns within this data to identify individuals at risk. This paper investigates:

- The comparative performance of diverse machine learning classifiers on psychometric datasets for mental health decision-making.
- The interpretability and real-world applicability of each model.
- How can these insights improve current screening protocols in mental healthcare?

This study reinforces the understanding that traditional models such as Logistic Regression (LR) offer valuable interpretability and simplicity, which are crucial for clinical adoption [2]. However, contrary to common expectations, our results demonstrate that LR can also achieve exceptionally high accuracy and diagnostic performance on psychometric mental health data, besides being linear in nature, surpassing more complex models such as Support Vector Machines (SVMs), Random Forests (RF), and gradient-boosting techniques like XGBoost and LightGBM in this context. Although advanced algorithms are often celebrated for their ability to capture intricate patterns, our findings suggest that with well-engineered features and appropriate data preprocessing, simpler models can be equally or more effective. Nevertheless, deep learning architectures, including Convolutional Neural Networks, may still hold potential for future applications but demand larger and more diverse datasets alongside substantial computational resources. This balance between model complexity, interpretability, and practical performance underscores the importance of context-specific evaluation when selecting predictive tools for mental health screening [3].

Instead of compromising diagnostic precision by shortening mental health questionnaires, our novel method adds an additional refined layer with a carefully chosen subset of questions. This enhances the efficiency of the assessment process without sacrificing accuracy [4].

II. Literature Review

A. Importance of Mental Health Screening

Disorders related to mental health, including anxiety, depression, and stress, affect millions worldwide each year. Tools such as the WHO's extensive 90-question screening allow a comprehensive evaluation but tend to be time-consuming, which can lead to participant fatigue and diminished test engagement. Studies have found that lengthy questionnaires mostly reduce response quality, as respondents may become distracted or unmotivated.

B. Efforts to Optimize Mental Health Screening

Many researchers have sought to reduce the length of mental health assessments. Approaches often involve statistically analyzing questions to single out the most significant items that preserve predictive accuracy. For instance, versions of the GAD-7 scale have been shortened from 7 to 4 items with nearly unchanged effectiveness. Recent efforts also apply machine learning for feature selection, attempting to predict mental health states using a minimal number of questions. Yet, these approaches generally don't include dynamic, AI-driven tailoring based on individual responses [5].

C. Machine Learning in Mental Health Assessments

Advancements in artificial intelligence have enabled predictive models to screen mental health conditions with high accuracy. LR, RF, and deep learning models have been used in detecting depression, anxiety, and stress disorders [7]. Machine learning models have been particularly useful in adaptive testing, where subsequent questions are selected based on previous responses, reducing the total number of questions needed for an accurate diagnosis.

D. The need for a Pre-screening Model

While existing studies focus on reducing the number of questions, *they do not integrate a two-step prescreening approach*. Our research introduces a model that:

- Uses a subset of key questions to provide an initial assessment.
- Determines whether the user needs further testing.
- Directs users to a limited set of further mental health tests, from which they can choose the specific assessments they wish to take.

This method ensures that individuals who do not require additional tests are efficiently screened, while those at risk are guided to more detailed assessments. Using machine learning models and user interaction data, our approach aims to improve efficiency without compromising diagnostic accuracy [8].

III. Methodology

The data sets comprise of standardized psychometric scores gathered from PHQ-9 and GAD-7 assessments, collected from a diverse set of participants with varied mental health backgrounds. Furthermore, a SHAP (SHapley Additive exPlanations) analysis was performed using the official Python library for each trained model. Beeswarm plots were generated for both output classes, summarizing feature contributions for individual predictions and global model behavior. Fig. 1 illustrates the complete data pipeline, from initial collection through preprocessing, feature construction, model training, testing, and evaluation.

To deepen interpretability, SHAP beeswarm plots were used not only to visualize aggregate feature importance but also to map the direction of each feature's influence for both possible screening outcomes. This approach allows clinicians to see, on a case-by-case basis, which questionnaire responses most strongly contributed to the model's recommendation. Features that demonstrate the largest spread of SHAP values are identified as the most critical in the risk assessment pathway. This thorough mapping ensures that our model's decisions are both transparent and clinically meaningful.

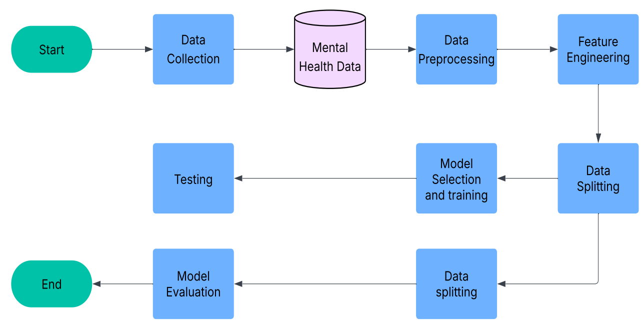

Fig. 1. End-to-end workflow of the proposed machine learning pipeline for mental health data analysis, from data collection to model evaluation.

A. Data Collection and Psychometric Dataset

This research utilized two established datasets featuring responses to psychometric questionnaires commonly used to assess mental health conditions. The first data set comprises participant responses to PHQ-9, a standardized tool for evaluating the severity of depression, while the second contains responses to GAD-7, which detects anxiety symptoms.

Both data sets were sourced from Kaggle like the set containing phq9 records from https://www.kaggle.com/datasets/thedevastator/phq-9-depression-assessment. These data sets include anonymized responses from a diverse set of participants that span different demographic profiles. Each data set includes complete response sets without missing entries, enabling robust machine learning analysis. Key characteristics include:

- Sample size: The PHQ-9 data set comprised approximately 15k individual records, while the GAD-7 data set consisted of approximately 13.5k records.
- Demographics: The participants in the PHQ survey ranged in age from 14 to 57 years, while those in the GAD-7 survey were between 17 and 30 years old. The sample included both male and female respondents. However, detailed demographic information beyond age and gender, such as socio-economic background or ethnicity, was limited in the publicly available metadata.

Before analysis, the data sets were standardized to a common format and combined where appropriate, maintaining data integrity and facilitating thorough model training and evaluation.

B. User flow and Digital Screening Process

The online system, as shown in Fig. 2, was designed to guide users starting from the landing and authentication pages through personal information provision, binary pre-screen testing, and, if necessary, subsequent clinical testing or professional recommendations. Depending on the pre-screen results, users either receive immediate feedback (if no further testing is needed) or proceed to additional mental health examinations (7 tests in total, as shown in Fig. 2, where they need to choose which tests to give. The system securely manages data throughout, including doctor recommendations and appointment bookings through associated databases [9].

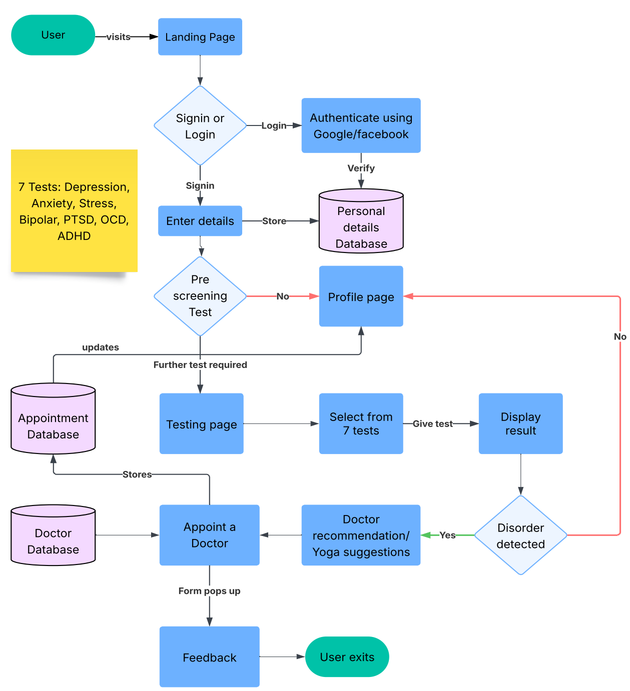

Fig. 2. User flow diagram of the digital pre-screening and assessment platform.

C. Data pre-processing and Feature Engineering

Preparing the data set for robust analysis involved a series of systematic preprocessing and feature engineering steps designed to improve data quality while uncovering meaningful patterns relevant to mental health assessment.

1) Handling Missing Values: To ensure the integrity of the input characteristics, missing data within the responses of the GAD-7 questionnaire was imputed using mean substitution, minimizing bias and the impact of incomplete records.

2) Encoding Categorical Data: All categorical variables, such as gender and other demographic attributes, were transformed using label encoding. This step facilitated the effective integration of diverse features into the modeling workflow and preserved information about group membership for downstream analysis.

3) Feature Construction: A key engineered feature is the composite score gad7- total, representing the sum of responses across all seven GAD-7 items. This aggregate score offers a clinically meaningful measure of anxiety severity and forms the basis for both analysis and model input. Furthermore, a binary label known as anxiety-label (for model building purpose) was derived from this score using a recognized threshold, facilitating classification tasks [10].

4) Standardization: All psychometric scores were standardized to ensure comparability between different features and optimize the performance of machine learning models.

5) Exploratory Visualization: To support and illustrate the impact of feature engineering, several graphical analyses were conducted:

- The scatter plot between gad7-total and age reveals the distribution of anxiety levels across different age groups, highlighting potential patterns and outliers.

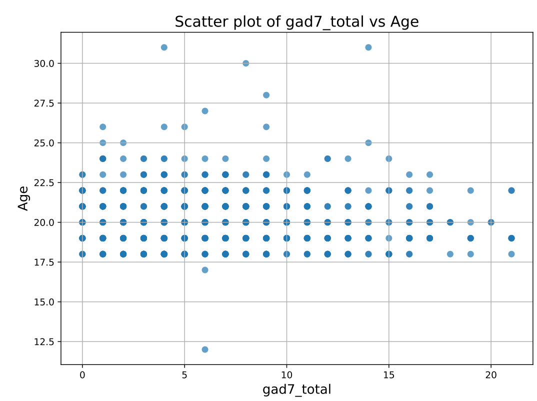

Fig. 3. Scatter plot of gad7_total versus age

- The box plots of gad7-total stratified by gender showcase differences in anxiety scores between male and female participants, indicating potential group effects or differences in scale response behaviors.
- The correlation heat map provides a comprehensive overview of the relationships between questionnaire responses, demographic characteristics, engineered scores, and the derived anxiety label. This visualization underlines the interdependencies and supports feature selection choices.

Collectively, these pre-processing and feature engineering strategies ensure that the data fed into subsequent models is both reliable and maximally informative, while accompanying visualizations offer transparent justification for feature selection and transformation choices.

D. Machine Learning Model Selection and Implementation

1) Rationale: This research utilizes a selection of machine learning algorithms that are well-regarded for their balance of interpretability, predictive accuracy, and computational practicality in mental health applications. LR is preferred for

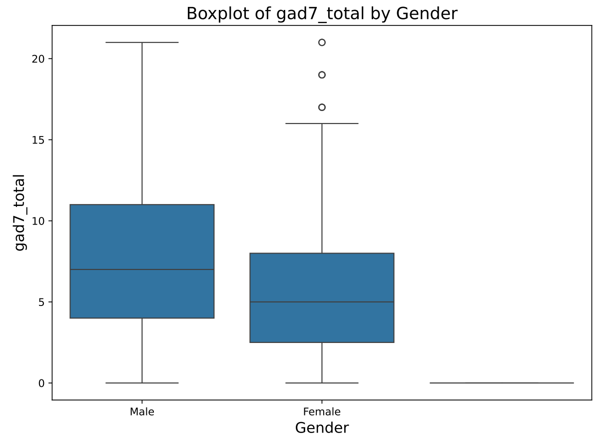

Fig. 4. Boxplot of gad7_total scores by gender

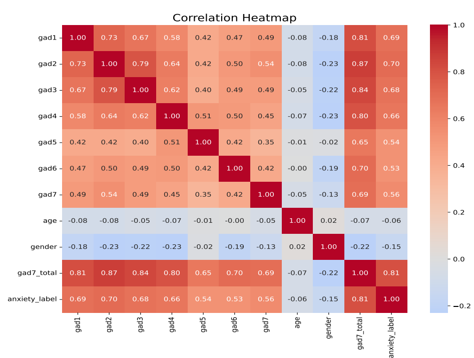

Fig. 5. Correlation heatmap of selected features

its straightforward interpretability, which aids in explaining results in clinical contexts. SVM is effective in modeling complex, non-linear patterns prevalent in high-dimensional psychological data. Random Forests offer robustness against overfitting by aggregating multiple decision trees and provide useful insights via feature importance scores. Gradient boosting models like XGBoost and LightGBM are employed for their strong predictive capabilities on structured data, with LightGBM notable for its faster training times and greater scalability in large datasets.

2) Methodology: All models were built using popular Python frameworks such as scikit-learn, XGBoost, and LightGBM. The input data was standardized to ensure consistency, and hyperparameters were optimized through cross-validation techniques to enhance model performance. LR incorporated regularization techniques to minimize overfitting risks. SVM utilized various kernel functions, including the radial basis function, to capture non-linear relationships. Ensemble models, including RF, XGBoost, and LightGBM, were fine-tuned with respect to tree quantity, depth, and learning rates according to their specific algorithmic requirements.

3) Formulas: In this research, several fundamental formulas underpin the machine learning models applied to psychometric mental health data, where $\mathbf{x}$ represents the array of input features derived from standardized psychometric scores (such as PHQ-9, GAD-7), and y denotes the binary label indicating the need for further mental health assessment.

- LR Model: The probability $P(y = 1|x)$ of a positive class (requiring further evaluation) is modeled using the sigmoid function:

$$P(y=1|x) = \sigma(z) = \frac{1}{1+e^{-z}}, \quad z = \mathbf{w}^T\mathbf{x} + b \tag{1}$$

where $\mathbf{w}$ is the weight vector and b is the bias term learned from the data. Regularization (L1 or L2) was applied to prevent overfitting, optimizing the cost function with cross-entropy loss.

- SVM: it aims to find a hyperplane that maximizes the margin between classes.
Linear SVM (hard margin):

$$\min_{\mathbf{w},b} \frac{1}{2}\|\mathbf{w}\|^2 \tag{2}$$

subject to

$$y_i(\mathbf{w}^T\mathbf{x}_i + b) \geq 1 - \xi_i, \quad \xi_i \geq 0 \tag{3}$$

where ξ_i are slack variables allowing some misclassifications, optimized with kernel functions (e.g., radial basis function) to handle nonlinear separability.

- Random Forest: Random Forest aggregates predictions from multiple decision trees $h_t(\mathbf{x})$ as:

$$\hat{y} = mode(h_1(\mathbf{x}), h_2(\mathbf{x}), ..., h_T(\mathbf{x})) \tag{4}$$

where $h_t(\mathbf{x})$ is the result of the t^{th} tree. Splitting Criteria (e.g., Gini Index):

$$G = 1 - \sum_{k=1}^{K} p_k^2 \tag{5}$$

where p_k is the proportion of class k in a node. Captures complex feature interactions and reducing variance.

These formulas were adapted to the dataset by mapping each question's score to a feature in vector $\mathbf{x}$, normalized to ensure comparability. The output label y was assigned based on clinically determined thresholds from the questionnaire results, indicating whether a user requires further mental health evaluation. This mathematical framework enabled the models to learn predictive patterns from the multi-dimensional psychometric data effectively.

E. Model Evaluation Metrics

To assess model effectiveness, multiple evaluation metrics were used:

- Accuracy: Measures the overall accuracy of the predictions.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{6}$$

where TP: True Positive (outcome correctly predicted as positive), TN: True Negative (outcome correctly predicted as negative), FP: False Positive (outcome incorrectly predicted as positive), FN: False Negative (outcome incorrectly predicted as negative)

- Precision: Precision measures how accurate the positive predictions made by the model are. In other words, out of all instances the model labeled as positive (e.g., identified as needing further mental health assessment), how many were actually correct [6].

$$Precision = \frac{TP}{TP + FP} \tag{7}$$

- Recall: Recall quantifies the model's ability to identify all relevant positive cases. That is, out of all actual positive instances (people who truly need further testing), how many did the model successfully detect?

$$Recall = \frac{TP}{TP + FN} \tag{8}$$

- F1 score: Balances precision and recall.

$$F1score = 2 \times (\frac{Precision \times Recall}{Precision + Recall}) \tag{9}$$

- ROC Curves and AUC (Area Under Curve): Analyzes the trade-off between sensitivity and specificity.
  - ROC Curve: Plots True Positive Rate (TPR = Recall) vs. False Positive Rate (FPR = $\frac{FP}{FP+TN}$)
  - AUC: Area under the ROC curve, a scalar measure of discrimination between classes.

TABLE I
PERFORMANCE COMPARISON OF DIFFERENT MACHINE LEARNING MODELS.

Model	Accuracy	Precision	F1-Score	ROC-AUC
LR	99%	1.00	0.99	1.00
RF	95%	0.93	0.93	0.93
SVM	98%	0.97	0.97	0.995
XGB	96%	0.93	0.94	0.995
LGBM	96%	0.94	0.95	0.994

F. Cross-Validation and Robustness Testing

To reduce the risk of overfitting and verify the model's ability to generalize, we used a 5-fold stratified cross-validation as shown in Table 2. This approach maintained the proportional representation of each mental health condition across all folds, helping to prevent sampling bias and ensuring that each fold was representative of the overall dataset.

TABLE II
5-FOLD CROSS-VALIDATION RESULTS FOR DIFFERENT MODELS (ROUNDED TO 4 DECIMALS).

Model	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Mean
LR	0.9919	0.9919	1.0000	1.0000	1.0000	0.9968
RF	0.9677	0.9677	0.9758	0.9758	0.9677	0.9710
SVM	0.9839	0.9758	0.9919	1.0000	0.9839	0.9871
XGB	0.9677	0.9839	0.9919	0.9919	0.9677	0.9806
LGBM	0.9677	0.9677	0.9919	0.9919	0.9758	0.9790

G. Statistical Analysis

A set of comparative statistical tests were conducted to analyze performance differences between models:

- Paired t-tests were used to compare the accuracy of every pair of models by examining their scores across folds, to assess if those differences were statistically meaningful.
- In addition, a one-way analysis of variance (ANOVA) test was performed to evaluate whether there were any significant performance differences among all models overall, using a significance level of $p < 0.05$.

TABLE III
PAIRED T-TEST RESULTS COMPARING MODEL PERFORMANCES.

Model Pair	t-statistic	p-value
LR vs RF	16.00	0.000089
LR vs SVM	3.21	0.0327
LR vs XGB	4.78	0.00875
LR vs LGBM	4.49	0.0109
RF vs SVM	-6.32	0.0032
RF vs XGB	-3.50	0.0249
RF vs LGBM	-2.24	0.0890
SVM vs XGB	1.16	0.3098
SVM vs LGBM	3.16	0.0341
XGB vs LGBM	0.65	0.5511

TABLE IV
ONE-WAY ANOVA TEST RESULTS FOR MODEL PERFORMANCE.

Statistic	Value
F-statistic	6.58
p-value	0.0015

These tests helped identify which models performed significantly better or worse, providing a rigorous basis for selecting the most effective model for predicting depression and anxiety conditions.

H. Computational Framework

All models were developed in Python using the following libraries:

- scikit-learn: For LR and statistical analysis.
- Pandas and NumPy: For data processing and transformation.
- Matplotlib and Seaborn: For visualizing results, confusion matrices, and model performance.
- SHAP: For generating interpretable explanations of model predictions and enhancing model interpretability in clinical contexts.

IV. RESULTS AND DISCUSSIONS

This study was motivated by the increasing demand for innovative methods in mental health assessment, particularly the need to enhance diagnostic efficiency without compromising accuracy [14]. Previous research has emphasized the importance of simplifying existing mental health questionnaires to improve user experience and completion rates [12]. While many studies have focused on reducing the number of questions to shorten surveys, our work proposes a novel approach: introducing an additional subset layer of questions without necessarily reducing the original question set, thereby maintaining diagnostic precision.

Recognizing the growing need for decision support systems (DSS) to assist healthcare professionals and policymakers in the mental health domain, this research focused on utilizing machine learning techniques to automatically and accurately identify mental health conditions. Our results indicated that LR outperformed other models, achieving near-perfect scores in metrics such as ROC-AUC and recall, which highlights its suitability for effective mental health screening in practical settings.

Unlike earlier works that often faced a balance between shortening questionnaires and maintaining model accuracy, our study shows that combining a full set of questions with a smartly selected subset can maintain high accuracy without sacrificing ease of use. This strategy effectively addresses the requirements of clinical environments by preserving assessment quality while offering flexible options for administering surveys.

A. Results

The quantitative findings of the comparative analysis show that the LR achieved the highest overall accuracy and AUC values as shown in Fig. 3, demonstrating an enhanced ability to discriminate between classes in the mental health screening data set. SVM exhibited reliable precision and recall values, whereas RF, despite their interpretability, showed marginally lower performance.

In addition, ROC curves were generated, showing that the SVM model consistently outperformed the others. Statistical analysis using paired t-tests confirmed that LR provided statistically significant improvements over SVM and decision trees ($p < 0.05$).

To enhance interpretability and support clinical adoption, we applied SHAP analysis to our deployed LR model trained on the GAD-7 data set. Figs. 7 and 8 presents the beeswarm summary plots for the two output classes (requires further evaluation: 1; does not require: 0). The SHAP beeswarm plots demonstrate that certain features, corresponding to specific items of the psychometric questionnaire, exert the most pronounced impact on the model's decisions. For instances classified as 'does not require further evaluation', higher scores on these influential features typically nudge the model toward a negative outcome. In contrast, responses reflecting more severe symptoms, depending on the direction of the scoring, change the model's prediction to'requiring further evaluation'. The symmetry observed across the two classes confirms that the linear nature of the LR model ensures consistent and interpretable attribution of feature effects. This transparency reassures practitioners that decisions are anchored in clinically relevant responses.

B. Discussion

The LR approach outperformed SVM and decision trees in predicting mental health risk profiles with an accuracy of 99%. Its superior performance can be attributed to the quality of the selected features and effective modeling within a predominantly linear framework. Although linear relationships are modeled by LR, its interpretability and reliability make it highly suitable for clinical applications. Decision trees, while slightly lower in performance, offer an advantage in clinical settings where rule-based decision paths are valuable.

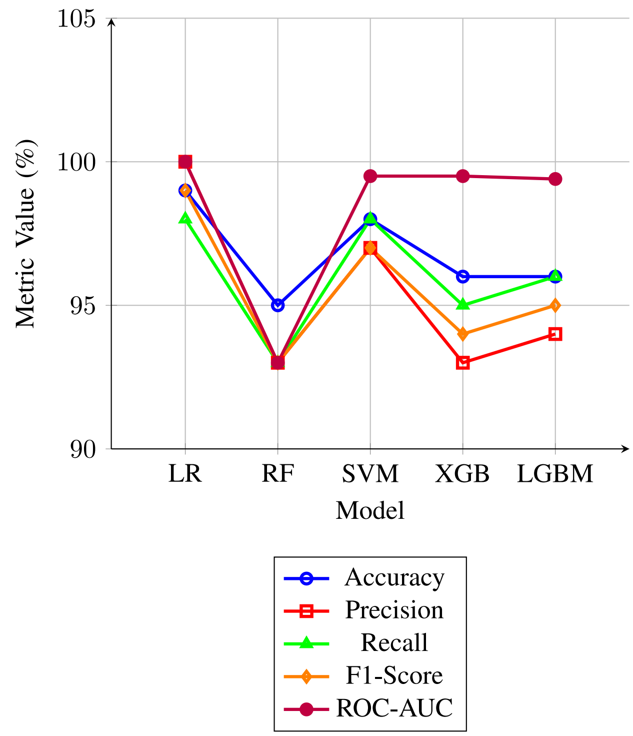

Fig. 6. Comparison of different metrics across models

The deployment of machine learning models in clinical settings involves practical considerations such as data quality, real-time prediction needs, and integration with electronic health records (EHRs). Furthermore, integrating SHAP-based explanations into our clinical workflow not only strengthens trust but also unlocks actionable insights for practitioners. By routinely reviewing SHAP-derived local explanations, healthcare professionals can audit model recommendations, ensure that the most crucial symptoms receive proper attention, and communicate transparent rationale to patients. Looking ahead, we recommend incorporating interactive SHAP dashboards within mental health decision support tools, enabling transparent, patient-level review of AI-driven assessments. This step is essential for bridging the gap between advanced analytics and frontline clinical usability.

C. Comparison with Other Studies

Several studies have attempted to reduce the number of questions in mental health assessments [11] [12], particularly focusing on compressing existing WHO guidelines. However, we do not simply select a subset of questions. Instead, we employ a prescreening model that dynamically determines whether a full assessment is necessary. This approach reduces

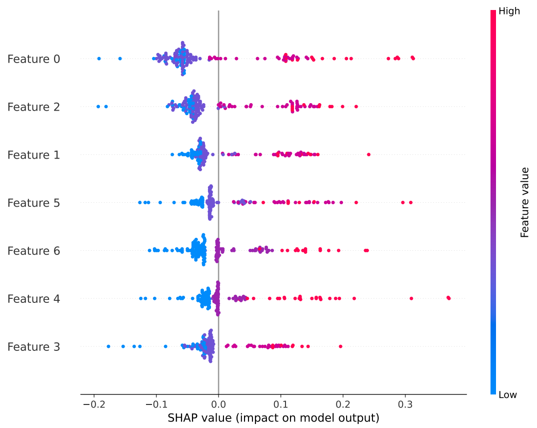

Fig. 7. SHAP Beeswarm Plot for LR Model Output Class 0

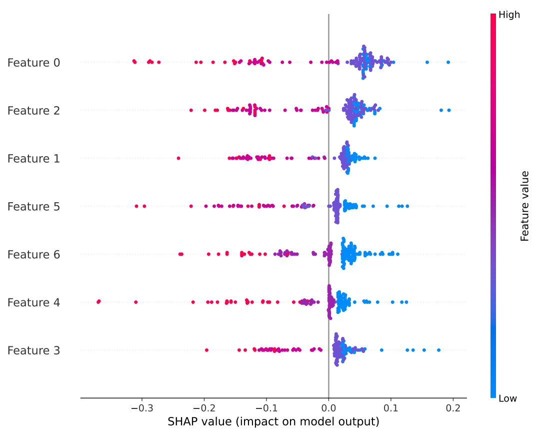

Fig. 8. SHAP Beeswarm Plot for LR Model Output Class 1
Feature 0: GAD-1 — Feeling nervous, anxious, or on edge?
Feature 1: GAD-2 — Not being able to stop or control worrying?
Feature 2: GAD-3 — Worrying too much about different things?
Feature 3: GAD-4 — Trouble relaxing?
Feature 4: GAD-5 — Being so restless that it is hard to sit still?
Feature 5: GAD-6 — Becoming easily annoyed or irritable?
Feature 6: GAD-7 — Feeling afraid as if something awful might happen?

operational overhead by approximately 50% while maintaining diagnostic accuracy. Furthermore, unlike previous studies that rely on static question reduction, our method adapts based on real-time responses, ensuring a more personalized assessment process.

V. Conclusion

This study highlights the comparative strengths of multiple machine learning models for the assessment of psychometric mental health using standardized scales such as PHQ-9, GAD-7. In present work SVM deliver best performance compared to state-of-the-art across key metrics, while ensemble models offer a balance of interpretability and accuracy suited to clinical contexts. Importantly, this work demonstrates the value of a prescreening layer in digital mental health triage, enabling early identification of individuals at risk while optimizing the screening process' efficiency and precision. The proposed methodology supports scalable, data-driven decision-making and opens doors for deploying adaptive, personalized assessment tools with increased reach and reliability.

References

[1] C. Casado-Lumbreras, A. Rodríguez-González, J. M. Álvarez-Rodríguez, and R. Colomo-Palacios, "PsyDis: Towards a diagnosis support system for psychological disorders," Expert Syst. Appl., vol. 39, no. 13, pp. 11391–11403, 2012. https://doi.org/10.1016/j.eswa.2012.04.033

[2] I. Y. Chen, P. Szolovits, and M. Ghassemi, "Can AI help reduce disparities in general medical and mental health care?," AMA Journal of Ethics, vol. 21, no. 2, pp. 167–179, 2019. https://doi.org/10.1001/AMAJETHICS.2019.167

[3] D. Puri, P. Kachare, S. Nalbalwar, "Metaheuristic optimized time–frequency features for enhancing Alzheimer's disease identification", *Biomedical Signal Processing and Control*, Vol 94, 2024, https://doi.org/10.1016/j.bspc.2024.106244.

[4] S. Chen, D. Stromer, H. A. Alabdalrahim, S. Schwab, M. Weih, and A. Maier, "Automatic dementia screening and scoring by applying deep learning on clock-drawing tests," Scientific Reports, vol. 10, no. 1, pp. 1–11, 2020. https://doi.org/10.1038/s41598-020-74710-9

[5] K. K. Fitzpatrick, A. Darcy, and M. Vierhile, "Delivering Cognitive Behavior Therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): A randomized controlled trial," JMIR Mental Health, vol. 4, no. 2, e19, 2017. https://mental.jmir.org/2017/2/e19

[6] Puri, D.V., Gawande, J.P., Kachare, P.H. et al. "Optimal time-frequency localized wavelet filters for identification of Alzheimer's disease from EEG signals", *Cogn Neurodyn* 19, 12, 2025. https://doi.org/10.1007/s11571-024-10198-7.

[7] M. Galesic and M. Bosnjak, "Effects of questionnaire length on participation and indicators of response quality in a web survey," Public Opinion Quarterly, vol. 73, no. 2, pp. 349–360, 2009. https://doi.org/10.1093/poq/nfp031

[8] D. Puri, P. Kachare, S. Nalbalwar, "Metaheuristic optimized time–frequency features for enhancing Alzheimer's disease identification", *Biomedical Signal Processing and Control*, Vol 94, 2024, https://doi.org/10.1016/j.bspc.2024.106244.

[9] J. Hardt and H. U. Gerbershagen, "Cross-validation of the SCL27: A short psychometric screening instrument for chronic pain patients," European Journal of Pain, vol. 5, no. 2, pp. 187–197, 2001. https://doi.org/10.1053/eujp.2001.0231

[10] R. C. Kessler, G. P. Amminger, S. Aguilar-Gaxiola, J. Alonso, S. Lee, and T. B. Ustün, "Age of onset of mental disorders: A review of recent literature," Current Opinion in Psychiatry, vol. 20, no. 4, pp. 359–364, Jul. 2007. [FREEFulltext] [doi:10.1097/YCO.0b013e32816ebc8c][Medline: 17551351]

[11] M. T. Minen et al., "The promise of digital mental health solutions," Journal of Mental Health, 2020. [Online]. Available: https://www.tandfonline.com/doi/full/10.1080/09638237.2020.1766005

[12] S. Tutun, M. E. Johnson, A. Ahmed, A. Albizri, S. Irgil, I. Yesilkaya, E. N. Ucar, T. Sengun, and A. Harfouche, "An AI-based decision support system for predicting mental health disorders," AI and Ethics, 2022. https://doi.org/10.1007/s43681-022-00184-7

[13] D. Puri, J. Gawande, J. Rajput, S. Nalbalwar, "A novel optimal wavelet filter banks for automated diagnosis of Alzheimer's disease and mild cognitive impairment using Electroencephalogram signals", *Decision Analytics Journal*, Vol 9, 2023. https://doi.org/10.1016/j.dajour.2023.100336.

[14] K. Zivin, D. Eisenberg, S. Gollust, and E. Golberstein, "Persistence of mental health problems and needs in a college student population," Journal of Affective Disorders, vol. 117, no. 3, pp. 180–185, Oct. 2009. [doi:10.1016/j.jad.2009.01.001] [Medline: 19178949]