Comparative Machine Learning Approaches for Mental Health Condition Classification

Abhijit Mahendra Bhalekar, Krishna Sharma S, Saurabh Kumar Prajapati, Vaibhav Agrawal
Indian Institute of Information Technology, Lucknow
Email: {MCS24003, MCS24006, MCS24004, MCS24007}@iiitl.ac.in

***Abstract*—Accurate classification of mental health conditions is critical for timely intervention and improved patient outcomes. This study evaluates the effectiveness of three machine learning models—Binary Classification, LightGBM, and Support Vector Machines (SVM)—in predicting mental health conditions. The models were enhanced with preprocessing techniques like PCA for dimensionality reduction and SMOTE for class balancing. A comprehensive evaluation based on accuracy, precision, recall, and F1-score reveals that the SVM model, integrated with PCA and SMOTE, provides superior performance in handling high-dimensional and imbalanced data. The findings underscore the potential of machine learning in supporting mental health diagnosis, highlighting the need for advanced preprocessing and optimization techniques for reliable and scalable solutions.**

***Index Terms*—Mental Health Classification, Machine Learning, Support Vector Machines (SVM), LightGBM, PCA**

I. Introduction

Mental health is crucial for overall well-being, forming the way an individual thinks, feels, and functions. It really affects everyday life in factors such as the ability to interact with people, working, and dealing with stress. Mental health conditions like rest depression and anxiety, bipolar disorder and schizophrenia are very common mental illnesses and are the most common causes of disability around the world.

Mental health issues among disabled people are the most undiagnosed issues, but even if diagnosed, most cases are poorly managed and dominated due to organizational stigma, resource inadequacies, and classic approaches in diagnosis. These are strictly based upon clinical reviews, interviews, and self-reported exposure surveys which at times are subjective, lengthy, and inconsistent hence end up delaying the diagnosis and treatment hours.

The increase of technological development in recent years especially artificial intelligence and ML has opened new doors to diagnose and manage more mental health conditions. ML stochastic analysis of large and complex data sets can find patterns and make reliable predictions. So if active lifestyle, behaviors and other socio-demographic variables can greatly validate and better identify and classify the disease so that it helps the providers. This can lead to faster, more accurate, and more consistent diagnoses, enabling early intervention and better patient outcomes.

In this research, the functional evaluation of three most popular machine learning algorithms used for the prediction of the state is carried out: Logistic Regression, Light Gradient Boosting Machine, Support Vector Machines (SVM). This is the most basic technique available, which categorizes a sample in two ways, those who are 'sick' and those who are 'non-sick' people. This means that a model based on this kind of intuition can always be employed as the first criterion. There are also even larger gaps of training speed and efficiency of the ensemble on large datasets from the performance of single models. One of such species of boosting which performs well on large dataset is LightGBM. Support Vector Machine can be referred to as reliable techniques in very high dimensional spaces and therefore, they address classification challenges that have many different features to deal with.

One of the continuous issues which has been encountered in the identification of effective models for the classification of mental health issues was class imbalance. Typically, in different datasets the number of non-ill individuals far exceeds the number of people suffering from mental disorders. This kind of dominance may result in erroneous predictions since the model will always be build orientated to the most populated class. In order to achieve this, we opt for a method called SMOTE (Synthetic Minority Over-sampling Technique) which is able to create examples that are not real for the minority class and thus balance the entire dataset. Furthermore, PCA – Principal Component Analysis is utilized here for the purpose of data dimensionality reduction while preserving significant factor

The main aim of this study is to assess which machine learning model effectively predicts the severity of the mental health condition with the aid of socio-demographic and behavioral variables. Our dataset includes age, gender, occupation, stress, sleeping patterns, and physical activities which aid in classifying mental health severity into *None, Mild, Moderate,* and *Severe* levels. We consider accuracy, precision, recall, and F1-score, among other metrics, in the assessment of each model in terms of its value and its narcissistic limitations.

By delving into these models, we hope that we will be able to help in the creation of AI tools for diagnosis relating to mental health which will allow healthcare providers to offer care with accuracy and in a reasonable timeframe. The ability to detect mental health issues at an early stage together with a suitable approach to treatment can vastly improve patient outcome while alleviating the strain on healthcare resources with a view to improvement of the quality of life in general. Finally, this research is aimed at realising how machine learning can be utilized in revolutionizing the mental health care landscape and subsequently, encouraging that diagnostic solutions become more efficient and available.

II. Related Work

Machine learning (ML) in predicting mental disorders has generated growing interest, as it can provide an effective diagnosis as well as help in the initiation of effective treatments faster. D. Ram et al. [1] conduct a detailed review of deep learning models, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), used in mental health prediction. These specific models are particularly well suited to finding intricate structures in microscopic images obtained from texts such as social network information and user-generated content. However, their work also pointed out the limitations associated with these models. Among the concerns that can be addressed are the need for ample amount of adequately annotated data, optimal computing power, and possible issues on the explainability of the model. Because of these problems, their use in real clinical scenarios, where the both transparency and efficiency are key concerns, is restricted.

Class imbalance is a common problem in mental health datasets, where much of the data concerning healthy people far surpasses data about patients with mental health disorders. C. Cortes et al. [2] investigated the ability of traditional ML algorithms, Support Vector Machines (SVM), and Random Forests, to be used for mental health classification tasks. SVMs are best suited for high-dimensional data since they learn linear or non-linear optimal separating hyperplanes [3]. However, biased predictions are found with imbalanced datasets since the model overemphasizes the majority class. Techniques like SMOTE, which generate synthetic samples of minority classes have thus been recommended, leading to improved model performance and bias mitigation [2].

Social media-based approaches have also shown promising results in identifying mental health trends and conditions. T. Chen et al. [4] applied deep learning models to large-scale social media datasets, using user-generated content to detect signs of mental health disorders, including depression and anxiety. Extrapolating from models like RNNs, the study showed that text-based analyses might become valid techniques for unearthing behavioral and emotional patterns. Nonetheless, ethical considerations, data privacy concerns, and the need for extensive labeled data remain challenges that must be addressed for better adoption [4].

Traditional ML models have proven to be strong in mental health classification, particularly due to their interpretability and ease of implementation. Mohan et al. [5] compared SVM and Random Forests for predicting mental health outcomes using structured health survey data. Their findings show that SVM, when optimized through feature selection and appropriate kernel tuning, is highly effective in high-dimensional spaces. Random Forests, introduced by [6], offer robust predictive performance but can lack the transparency needed for clinical applications. Balancing predictive accuracy and interpretability remains a key consideration in healthcare-oriented ML models.

Gradient boosting algorithms, or more specifically, LightGBM, also known as Light Gradient Boosting Machine, has been one of the most efficient algorithms for building predictive models. Its performance can be attributed to its capability to build models faster by employing the leaf-wise growth of the trees. This feature makes LightGBM particularly appropriate for mental health datasets that are complex and have many dimensions. Studies like that of Mabrouk et al. [7] have shown the effectiveness of LightGBM in classification. LightGBM makes it possible to process large datasets within a short period in order to discern patterns that can have correlations with a given condition, such as mental health, thereby increasing predictive capabilities in practice.

Dimensionality reduction techniques have assumed the critical role of optimizing ML models towards mental health classification. It is common to find high-dimensional data that is plagued with excessively many uninformative or noisy features that adversely affect model generalization. Rastogi et al. [8] and Ram et al. [1] studied the use of Principal Component Analysis (PCA) to reduce data dimensionality while preserving essential features. Whenever there is a potential threat of overfitting, PCA also aids in simplifying the model making it more rather interpretable. Combining it with techniques such as SMOTE which resolves class imbalance makes it successful.

Zhange et al.'s work taps into using socio-demographic traits and electronic health records, in an attempt to improve upon predictions of mental health [9]. They demonstrated the potential of combining diverse data sources to enhance predictive accuracy. Their paper also highlighted the importance of capturing a wide range of features, reflecting the multifaceted nature of mental health disorders. Including socio-demographic factors along with health records of individuals provide an integrated view for ML models towards assessing their mental health status.

Recent reviews by the works such as that of Sharma et al. [10] and Singh et al. [11] have further emphasized the critical role of feature selection, model interpretability, and data integration in ML-based mental health diagnosis. Advanced models like deep learning architectures also carry high predictive accuracy but comparable performance can also be attained using simpler models needing fewer computational resources. Traditional models are thus more attractive for deployment within resource-constrained settings where interpretability and efficiency become key.

Building upon these existing works, our study compares the performance of Binary Classification, LightGBM, and SVM models for mental health classification using structured socio-demographic and behavioral datasets. We use PCA to reduce dimensionality of the data used and SMOTE to balance class distributions, improving model accuracy and reliability. This research aims to contribute to the development of AI-driven tools for early detection and intervention in mental health care, ultimately enhancing diagnostic accuracy, patient outcomes, and the accessibility of mental health services.

III. Proposed Methodology

The main goal of this study is to evaluate and compare the effectiveness of three machine learning models—Logistic Regression, LightGBM (Light Gradient Boosting Machine), and Support Vector Machines (SVM)—to predict the severity of mental health based on socio-demographic and behavioral data. The methodology used in this research consists of several key steps, including data preprocessing, feature engineering, dimensionality reduction, handling class imbalance, and model training and evaluation.

A. Data Preprocessing and Feature Engineering

The dataset used in this study comprises socio-demographic attributes (such as age, gender, and occupation) and behavioral features (including stress levels, sleep patterns, physical activity, and work hours). Preprocessing was done to ensure data quality and suitability for machine learning model training. The preprocessing steps included:

- **Handling Missing Values:** Missing data were addressed using appropriate imputation techniques. Continuous features with missing values were imputed using the mean value, while categorical features were filled using the mode.
- **Normalization and Scaling:** Continuous features were normalized to a common scale using min-max scaling. This ensures uniform data distribution, enhancing model training efficiency.
- **Encoding Categorical Variables:** Categorical features, such as gender and occupation, were converted to numerical data using one-hot encoding. This transformation ensured that the machine learning models could interpret categorical data effectively without introducing biases due to any arbitrary numeric assignments.

B. Dimensionality Reduction Using PCA

High-dimensional data pose a problem to a model's performance because of its noise and irrelevant features. To mitigate this issue, Principal Component Analysis (PCA) was employed for dimensionality reduction. In PCA, the dataset is mapped into the lower-dimensional space through unconnected features of maximum variance-the principal components-of the dataset, hence reducing the computational complexity of models due to the decrease in size together with the risk of overfitting, making it quite useful in high-dimensional datasets typical of mental health classification.

C. Addressing Class Imbalance Using SMOTE

The presence of class imbalance is common in datasets of mental health where data constitutes more instances of healthy people than people with mental health disorders. Therefore, biased predictions might occur when using an imbalanced dataset due to the majority class. To address this, the Synthetic Minority Over-sampling Technique (SMOTE) was applied. SMOTE generates synthetic samples for the minority class by interpolating between existing instances, effectively balancing the class distribution. This mainly enhances the capacity of the model to distinguish between underrepresented classes and generally raises performance.

D. Model Selection and Training

We picked three models of machine learning for this experiment: Logistic Regression, LightGBM, and SVM. Each was trained and rated against each other to serve the purpose of providing a classification of mental health.

1) Binary Classification: A baseline solution was developed as a binary classification model with logistic regression. In the context of the problem, this is the kind of a model that assumes to take on one of two values: ”not healthy” or ”afflicted by mental health disorder.” Logistic regression was chosen because of its simplicity and interpretability, which also made this a good benchmarking choice in comparison to more advanced models.

2) LightGBM Model: LightGBM (Light Gradient Boosting Machine) is a highly efficient gradient boosting framework designed for fast and accurate model training. It builds multiple decision trees iteratively, optimizing a loss function to improve prediction accuracy. Grid search and cross-validation were applied for hyperparameters such as the number of leaves, learning rate, and max depth to tweak the process. Large-dataset handling with efficient usage of light-wise tree growth makes it particularly good for complex feature interaction mental health classification tasks.

3) SVM Model with PCA and SMOTE: The Support Vector Machine (SVM) is a robust model for high-dimensional classification tasks. Several SVM-based models were developed; the two used in this case used a linear and radial basis function (RBF) kernels with different decision boundaries. SVM combined with PCA for dimensionality reduction, and SMOTE for dealing with class imbalance, allows for almost better classification ability of highly complex data. The best hyperparameters for SVM and Gaussian kernel values along with values for the regularization parameter C was determined using a grid search over permutations of these settings.

E. Model Evaluation Metrics

To evaluate the performance of the models, several metrics were used, including:

- **Accuracy:** The proportion of correctly classified instances out of the total instances.
- **Precision:** The ratio of correctly predicted positive observations to the total predicted positives, measuring the model's ability to avoid false positives.
- **Recall:** The ratio of correctly predicted positive observations to all observations in the actual positive class, assessing the model's ability to detect true positives.
- **F1-Score:** The harmonic mean of precision and recall, providing a balanced measure of model performance.

These metrics were computed for each model to compare their effectiveness in predicting mental health severity.

F. Experimental Setup and Cross-Validation

A cross-validation-based approach would ensure that performance evaluation was strong and reliable. A stratified k-fold cross-validation was used to split classes directly related to the folds for stratification of the training and testing dataset. This led to less bias and gave a holistic review of the model's performance on unidentified data. Cross-validation was implemented to carry out hyperparameter tuning in order to find the best parameters for models.

G. Implementation Details

Both of the following models have been built through the Scikit-learn and Pandas libraries. Basic data prep work, as well as, feature manipulation and dimensionality reduction has been done through efficient and standard python libraries for easy and reproducible model training.

The method presented in this paper provides a repeatable procedure for evaluating various models of machine learning for the tasks of mental health classification with strong focus on data wrangling, dimensionality reduction and class imbalance issues. This study provides a strategy of employing both simple and complex models in a bid to find the best models which fit the data for the purpose of mental health prediction which is both accurate and interpretable.

IV. Performance Evaluation

This section includes extensive evaluation, analysis, and interpretation of the results obtained from the evaluation of the machine learning models proposed for the tasks of mental health classification as well as the target data set from which these models where derived.

A. Dataset Description

This dataset was obtained from the Mental Health Dataset (Bhadra Mohit, 2022). The control population is rather broad and includes not only socio-demographic, such as age, gender, occupation, level of education, and marital status, but also behavioral features within the sphere of mental health disease. This dataset has a volume of 1000 data points and features such as:

- **Age:** Integer age of the response participant, rounded to the nearest number.
- **Gender:** This is a categorical feature that captures the respondent's gender.
- **Occupation:** The job of the respondent is classified as a categorical variable.
- **Stress Levels:** Self-reported stress levels on a scale, reflecting mental and emotional stress.
- **Sleep Patterns:** Average hours of sleep per day, representing sleep quality and habits.
- **Physical Activity:** Frequency or intensity of physical activity, impacting overall health and well-being.
- **Work Hours:** Number of hours worked per week, which may influence stress and mental health.

The target variable represents the severity of the mental health condition which is categorized into four classes: *None*, *Mild*, *Moderate*, and *Severe*. The dataset was preprocessed to handle missing values, normalize continuous features, and encode categorical variables, as detailed in the proposed methodology section.

B. Experimental Results and Analysis

The performance of the three machine learning models—Binary Classification, LightGBM, and SVM—was evaluated using accuracy, precision, recall, and F1-score. Table I summarizes the results for each model.

TABLE I
PERFORMANCE METRICS FOR MENTAL HEALTH CLASSIFICATION MODELS

Model	Accuracy	Precision	Recall	F1-Score
Logistic Regression	50.3%	0.49	0.50	0.46
LightGBM	51%	0.52	0.52	0.48
SVM with PCA and SMOTE	55.5%	0.56	0.56	0.55

1) Binary Classification (Logistic Regression): The Binary Classification model, implemented using logistic regression, served as a baseline for performance comparison. With an accuracy of 50.3%, the model was able to distinguish between individuals with and without mental health conditions based on a simplified binary target variable. Although it offered a basic classification method, its low predictive power also suggested a requirement for advanced models.

2) LightGBM Model: The accuracy of the LightGBM model stands at 49.5% which is a bit poorer than the results of Binary Classification. Because of its lower accuracy output, LightGBM managed to obtain fair precision and recall values. The leaf-wise tree construction approach of LightGBM, which is essential for large data set, helped in detecting complex interaction of features. This could of course be result of lower complexity of the data sets that were used in this study.

3) SVM Model with PCA and SMOTE: The Support Vector Machine (SVM) model, integrated with Principal Component Analysis (PCA) and SMOTE technique was more effective than both Binary Classification and LightGBM models. SVM had the highest accuracy of 55.5% along with an impressive precision, recall, and F1-score value of 0.56. With application of PCA, the number of features in the dataset was reduced and therefore the chances of SVM finding optimal decision boundaries increased. Also, SMOTE consistently achieved a better distribution of the classes throughout the models leading to improved performance in detection of likely minority instances and reduced bias as well.

C. Conclusion

The performance analysis shows an advantage of the combined efforts of SVM model with dimensionality reduction and class balancing methods to achieve better performance compared to a simpler Binary Classification and LightGBM models. This improved performance can be attributed to the enhancement of the data preprocessing steps such as PCA and

SMOTE in such cases of high dimensional data with class imbalance.

Even though in this study SVM was the most performing one, it should be noted that there was a modest overall prediction accuracy for all models. This indicates that provision of more diverse features, scaling of datasets and the application of unsupervised deep learning models are likely to result in even better predictions.

LightGBM's performance during the undertakings was lower than was anticipated but such was probably a result of low dataset size and complexity of the feature interactions. Further studies may diffuse hyperparameter optimization, feature engineering and employing ensemble approaches in order to optimize on the predictive capacity of LightGBM.

D. Limitations and Future Directions

Several limitations were identified in this study. There was limited dataset and this could have limited the generalization power of the models constructed. Increasing the dataset by increasing samples and variability in features can be helpful. Also, applying more sophisticated techniques, including deep learning models and ensemble methods, may improve effectiveness for more sophisticated tasks of mental health prediction.

Future directions should also include the combination of multiple databases like EHR, wearable devices and surveys in order to get a more holistic understanding of people's mental health. As a result of expanding the range and scope of data set, machine learning models will enhance their performance for mental health diagnosis and intervention.

This research exemplifies results in effective prediction of mental health using machine learning models, including SVM with PCA and SMOTE. However, it is important to remember that further development in the areas of data quality and preprocessing and modeling will enable the full use of artificial intelligence based solutions for mental health problems.

V. Conclusion

This research compared three models of machine learning, the Binary Classification model, the LightGBM and the SVM, in predicting mental health conditions. The findings revealed that it was SVM that reigned supreme after being fused with PCA and SMOTE techniques. Despite this, the accuracy of such models was only acceptable and for further enhancement more complex approaches such as larger amounts of data or data manipulation are needed. It would be advantageous to focus in the next efforts establishing more diverse types of data and implementing deeper learning architectures to improve the prediction rate.

Acknowledgment

The authors would like to express their deepest gratitude to Dr. Sushil Kumar Tiwari, our subject teacher, for his invaluable guidance and unwavering support for the duration of this study. His knowledge and profound insights have been a key factor in shaping the direction of this research.

References

[1] D. Ram, A. Patel, and P. Singh, "Mental health prediction using deep learning: A systematic review," *Health Information Science and Systems*, vol. 9, no. 1, p. 50, 2021.

[2] X. Dai, Y. Wang, and H. Li, "Improving mental health diagnosis using machine learning algorithms: A systematic review," *Healthcare Analytics*, vol. 5, no. 3, p. 100064, 2020.

[3] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[4] T. Chen, X. Yu, W. Zhang, and Z. Zhang, "Deep learning and its application for mental health in social media: A survey," *IEEE Access*, vol. 8, pp. 48 956–48 973, 2020.

[5] R. Mohan, R. Gupta, and A. Shankar, "A novel approach to mental health classification using random forests and svm," *Journal of Machine Learning for Healthcare*, vol. 6, pp. 83–98, 2020.

[6] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[7] B. Mabrouk, N. Jazzar, L. Sallemi, and A. B. Hamida, "A comparative study of pca and lda for dimensionality reduction in a 4-way classification framework," *J App Mat Sci & Engg Res*, vol. 8, no. 1, pp. 1–6, 2024.

[8] A. K. Rastogi, S. Taterh, and B. S. Kumar, "Dimensionality reduction algorithms in machine learning: A theoretical and experimental comparison," *Engineering Proceedings*, vol. 59, no. 1, p. 82, 2023.

[9] H. Zhang, Y. Xu, J. Geller, and E. Jones, "Machine learning approaches for mental health prediction using socio-demographic and electronic health record data," *Journal of Medical Internet Research*, vol. 22, no. 6, p. e20630, 2020.

[10] N. Sharma, P. Yadav, and R. Agarwal, "Role of machine learning in mental health diagnosis and prediction: A review," *International Journal of Medical Informatics*, vol. 141, p. 104173, 2020.

[11] S. Singh, N. Kumar, and N. Pathak, "Machine learning approaches for prediction of mental health disorders: A systematic review," *Journal of Artificial Intelligence in Medicine*, vol. 115, p. 101016, 2022.

[12] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, "Lightgbm: A highly efficient gradient boosting decision tree," in *Advances in Neural Information Processing Systems*, 2017, pp. 3146–3154.

[13] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.

[14] T. Fawcett, "An introduction to roc analysis," *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861–874, 2006.

[15] T. Hastie, R. Tibshirani, and J. Friedman, "The elements of statistical learning: Data mining, inference, and prediction," *The Elements of Statistical Learning*, 2009.

[16] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg *et al.*, "Scikit-learn: Machine learning in python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.