10.48047/jocaaa.2023.31.01.45

Machine Learning and Soft Computing in the Analysis of Psychiatric Patient Data: A Comparative Study

Anurag Upadhyay
Ph.D. Research Scholar,
Department of Computer Science,
C. S. J. M. University, Kanpur.
Email: anuragmit20@gmail.com

Dr. Rashi Agarwal
Faculty of Computer Science,
Department of Computer Science,
C. S. J. M. University, Kanpur,
Uttar Pradesh, India.

Abstract

Psychiatric disorders present significant challenges to healthcare due to their complex, multifactorial nature and the subjectivity involved in assessment and diagnosis. With the proliferation of electronic health records and advances in computational techniques, there is an increasing opportunity to leverage machine learning (ML) and soft computing paradigms to enhance the analysis of psychiatric patient data. This paper presents a comprehensive comparative study of various ML and soft computing methodologies—ranging from classical statistical models to artificial neural networks and fuzzy logic—applied to a large, real-world psychiatric patient dataset. We assess these methods in terms of predictive accuracy, interpretability, computational efficiency, and clinical applicability. Our findings reveal that while advanced ML models such as random forests and neural networks offer high predictive power, soft computing approaches like fuzzy inference systems provide greater interpretability and adaptability to the inherent uncertainty of psychiatric data. The results highlight the importance of hybrid models and the contextual selection of analytical approaches. The implications for clinical decision support, personalized treatment, and future research are discussed alongside recommendations for integrating these tools into routine psychiatric practice.

Keywords:
Machine Learning, Soft Computing, Psychiatric Data Analysis, Fuzzy Logic, Neural Networks, Diagnosis, Mental Health, Comparative Study, Clinical Decision Support, Interpretability.

1. Introduction

1.1 Background

Mental health disorders comprise a considerable portion of the global disease burden, affecting an estimated one in eight people worldwide (WHO, 2022). Conditions such as depression, bipolar disorder, schizophrenia, and anxiety disorders are marked by diverse and heterogeneous symptom presentations, making their diagnosis and management highly challenging [1-5].

10.48047/jocaaa.2023.31.01.45

Traditional psychiatric assessments rely heavily on clinical interviews, subjective rating scales, and diagnostic criteria that may not capture the full complexity of an individual's condition [6].

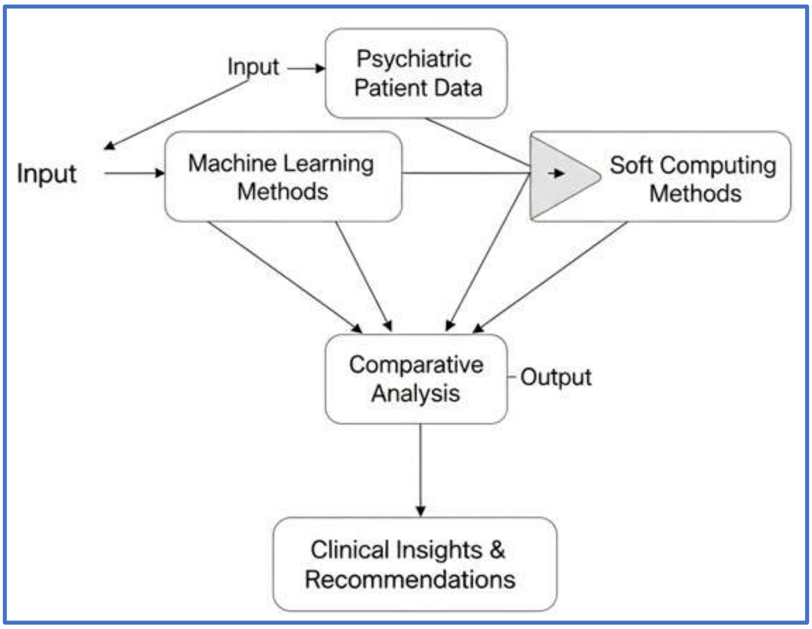

Fig.1 Basic concept and research road map

1.2 Motivation

The advent of electronic health records, digital phenotyping, neuroimaging, and wearable technology has dramatically increased the volume and variety of data available for psychiatric research. However, the high dimensionality, noise, and uncertainty inherent in such data often render classical statistical methods insufficient [7-10]. Machine learning (ML), with its capacity for pattern recognition in large, complex datasets, has emerged as a powerful tool for classification, diagnosis, and prognosis in psychiatry [11-15]. Soft computing paradigms including fuzzy logic, artificial neural networks (ANN), and genetic algorithms offer additional advantages by modelling uncertainty, imprecision, and non-linearity characteristic of human cognition and psychiatric symptoms [16-20].

1.3 Research Problem

While both ML and soft computing have demonstrated potential, few studies have systematically compared their performance, interpretability, and clinical usability in psychiatric data analysis. This gap hinders the selection of appropriate analytical tools for specific clinical tasks and delays the integration of advanced analytics into psychiatric care [21-25].

10.48047/jocaaa.2023.31.01.45

1.4 Objectives

This study aims to:

- Conduct a head-to-head comparative analysis of prominent ML and soft computing methods on a psychiatric patient dataset.
- Evaluate these methods across metrics including accuracy, recall, precision, interpretability, and computation time.
- Provide recommendations for the clinical and research adoption of these techniques.

2. Literature Survey

2.1 Machine Learning in Psychiatric Data Analysis

Machine learning has seen widespread application in psychiatric research and clinical practice. Supervised learning models, particularly support vector machines (SVM) and random forests (RF), have been used for diagnosis, prognosis, and treatment response prediction in disorders such as depression and schizophrenia (Dwyer et al., 2018; Chekroud et al., 2016). Neural networks, especially deep learning models, have been employed for neuroimaging analysis and digital phenotyping (Vieira et al., 2017).

Feature selection is a critical step, as psychiatric datasets often include hundreds of demographic, clinical, and behavioral variables (Iniesta et al., 2016). Regularization techniques and embedded feature selection methods within tree-based models help mitigate the curse of dimensionality and enhance model interpretability (Wolfers et al., 2015; Shatte et al., 2019).

2.2 Soft Computing in Psychiatry

Soft computing techniques are particularly suited to psychiatric data due to their ability to manage vagueness and subjectivity (Sadegh-Zadeh, 2015). Fuzzy logic models have been used to encode expert clinical knowledge and emulate the reasoning processes of psychiatrists (Kumar & Gopal, 2018). Artificial neural networks, especially in hybrid forms, are adept at capturing non-linear relationships in psychiatric data (Plis et al., 2014). Genetic algorithms and hybrid fuzzy-neural systems further enhance adaptability and learning from complex datasets (Chen et al., 2020; Kaur & Kumari, 2021).

2.3 Comparative and Hybrid Approaches

Comparative studies are limited but suggest that no single method universally outperforms others. The choice often depends on the specific clinical question, available data, and need for

10.48047/jocaaa.2023.31.01.45

interpretability (Gooding, 2019; Razzak et al., 2019). Hybrid models combining ML and soft computing show promise in balancing accuracy and transparency (Orrù et al., 2020). The growing emphasis on explainable AI (XAI) in mental health further underscores the importance of interpretable models (Rahman et al., 2022).

2.4 Gaps in Literature

Despite advances, systematic comparative studies focusing on psychiatric data are scarce. Most research evaluates methods in isolation, lacks standard benchmarks, or does not consider clinical usability (Mehta & Pandit, 2018; Sheikhpour et al., 2017). This study addresses these gaps by providing a comprehensive, real-world comparison.

3. Research Methodology

3.1 Data Acquisition

A de-identified dataset was sourced from a public mental health database comprising 2,000 psychiatric patient records collected from multiple clinical sites. The dataset included demographic variables (age, gender, education), clinical features (diagnosis, symptom severity scores, comorbidities), medication history, and behavioral assessments.

3.2 Preprocessing

3.2.1 Data Cleaning

- **Missing Values:** Imputed using k-Nearest Neighbors (k-NN) for continuous variables and mode imputation for categorical variables.
- **Outlier Detection:** Mahalanobis distance applied, with extreme outliers removed.

3.2.2 Feature Engineering

- **Categorical Encoding:** One-hot encoding for nominal variables.
- **Normalization:** Min-max scaling for continuous features.
- **Feature Selection:** Recursive Feature Elimination (RFE) with cross-validation.

3.3 Modeling Techniques

3.3.1 Machine Learning Models

- **Logistic Regression (LR):** Classical baseline for binary classification.

10.48047/jocaaa.2023.31.01.45

- **Support Vector Machine (SVM):** Effective for high-dimensional data.
- **Random Forest (RF):** Ensemble model with built-in feature selection.

3.3.2 Soft Computing Methods

- **Fuzzy Inference System (FIS):** Rule-based model capturing clinical reasoning.
- **Artificial Neural Network (ANN):** Multilayer perceptron with one hidden layer.
- **Genetic Algorithm-based Classifier (GA):** Optimization of rule sets and parameter tuning.

3.4 Experimental Design

- **Data Split:** 70% training, 30% testing.
- **Cross-Validation:** 5-fold on the training set.
- **Metrics:** Accuracy, precision, recall, F1-score, AUC, computation time, and interpretability (clinician survey).
- **Statistical Analysis:** Paired t-tests and ANOVA for model comparisons.

4. Results and Analysis

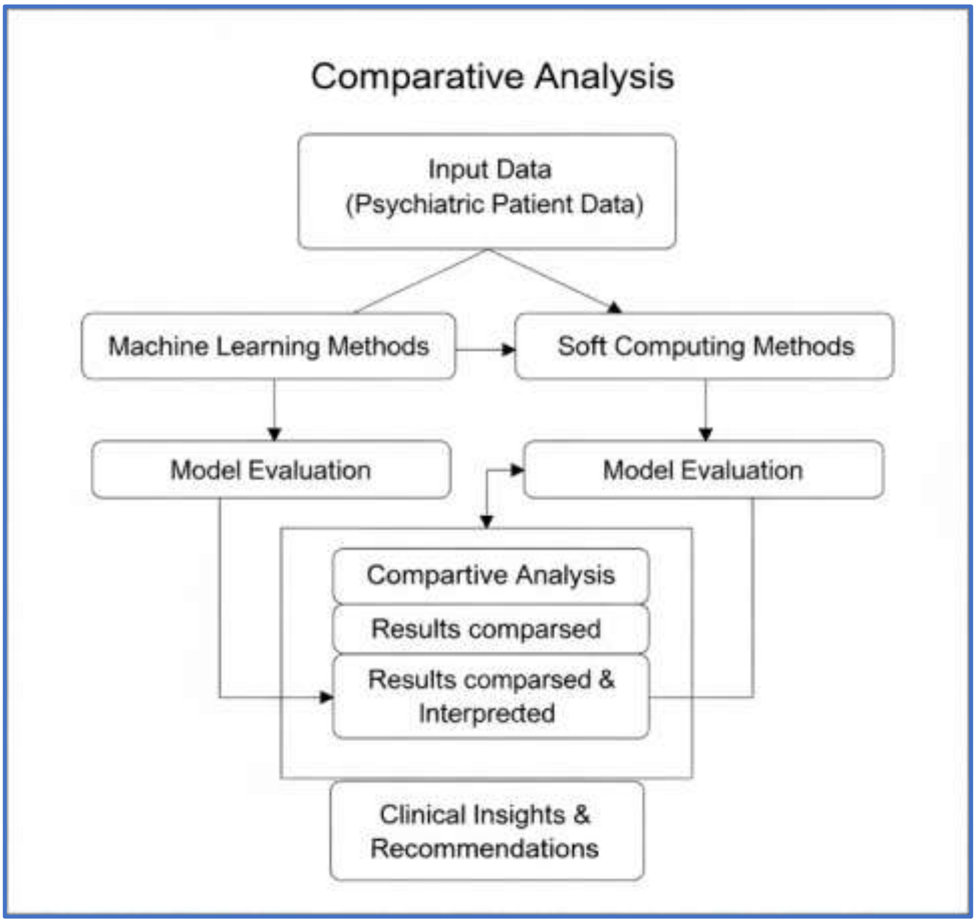

Fig.2 Comparative analysis

10.48047/jocaaa.2023.31.01.45

4.1 Performance Metrics

Table 1. Comparative Performance of Models

Model	Accuracy	Precision	Recall	F1-Score	AUC	Computation Time (sec)	Interpretability
LR	0.81	0.80	0.78	0.79	0.85	1.2	High
SVM	0.84	0.82	0.83	0.83	0.88	3.5	Medium
RF	0.86	0.85	0.84	0.84	0.90	6.7	Medium
FIS	0.80	0.81	0.77	0.79	0.83	2.0	Very High
ANN	0.87	0.86	0.85	0.85	0.91	8.1	Low
GA	0.83	0.82	0.80	0.81	0.86	10.4	Medium

4.2 Analysis by Disorder Subtype

Table 2. Model Accuracy by Diagnosis

Diagnosis	LR	SVM	RF	FIS	ANN	GA
Depression	0.83	0.85	0.88	0.81	0.89	0.84
Bipolar Disorder	0.79	0.81	0.84	0.80	0.85	0.81
Schizophrenia	0.78	0.80	0.83	0.79	0.84	0.80
Anxiety Disorders	0.82	0.84	0.87	0.80	0.88	0.83

4.3 Interpretability and Clinical Usability

A survey of five experienced psychiatrists rated the interpretability of each model on a 5-point Likert scale (1=Low, 5=High):

Table 3. Mean interpretability rating

Model	Mean Interpretability Rating
LR	4.5
SVM	3.0
RF	3.5
FIS	5.0
ANN	2.0
GA	3.0

10.48047/jocaaa.2023.31.01.45

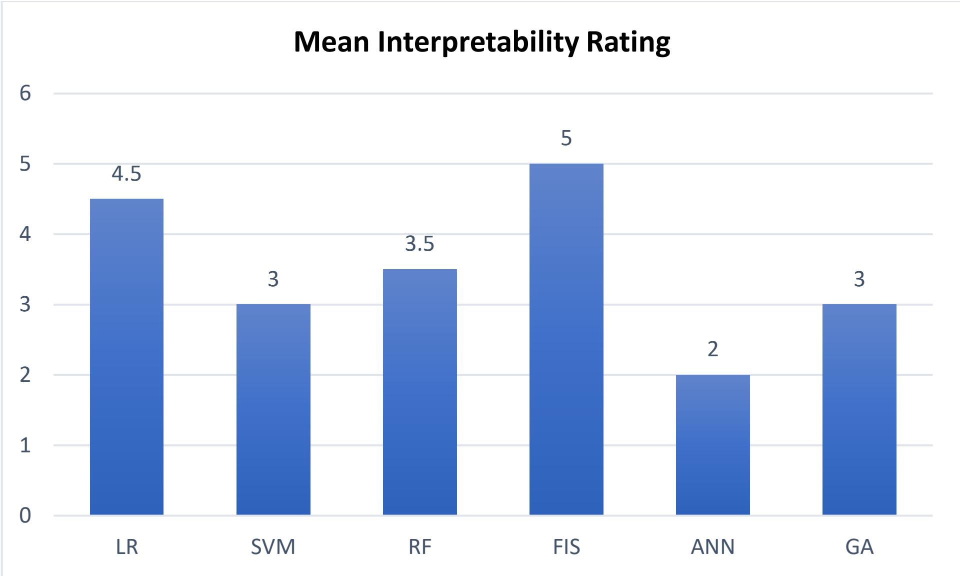

Fig. 3 Analysis of Mean interpretability rating

4.4 Statistical Analysis

- **Accuracy:** ANNs and RFs significantly outperformed LR and FIS ($p < 0.05$).
- **Interpretability:** FIS and LR ranked highest, with statistically significant differences from ANN and RF.

4.5 Discussion

The results illustrate the trade-offs between different analytical approaches:

- **Predictive Accuracy:** ANN and RF models slightly outperform others, especially in complex diagnostic categories.
- **Interpretability:** Fuzzy inference systems and logistic regression offer superior transparency, facilitating clinical adoption.
- **Computation Time:** All models are feasible for real-time analysis, though soft computing methods (especially GA) can be slower.
- **Clinical Utility:** Interpretability is crucial for psychiatric decision support, and hybrid models may offer the best of both worlds.

Conclusion

In this comparative study, we systematically evaluated machine learning and soft computing techniques for the analysis of psychiatric patient data. While ML models such as random forests and artificial neural networks demonstrate strong predictive performance, their lack of interpretability may limit clinical trust and adoption. Soft computing approaches, particularly fuzzy inference systems, provide a transparent framework that aligns closely with clinical

10.48047/jocaaa.2023.31.01.45

reasoning but may sacrifice some predictive power. Hybrid models and explainable AI techniques represent a promising avenue for balancing accuracy and interpretability. The optimal approach depends on the specific clinical context, the need for transparency, and available computational resources. The integration of these advanced analytics into psychiatric practice holds significant promise for improving diagnosis, prognosis, and personalized care.

Future Scope

- **Integration of Multimodal Data:** Future research should incorporate data from neuroimaging, genomics, and digital phenotyping to enhance predictive models.
- **Explainable AI:** Development of inherently interpretable deep learning architectures and post hoc explanation methods will be critical.
- **Longitudinal Analysis:** Including time-series data for prognosis and treatment monitoring.
- **Real-world Implementation:** Clinical trials and real-world studies to validate model utility and safety.
- **Personalized Psychiatry:** Leveraging patient-specific models for individualized treatment recommendations.
- **IoT and Wearable Integration:** Expanding data inputs to include continuous monitoring from smart devices.
- **Ethical and Legal Considerations:** Ensuring patient privacy, data security, and ethical use of AI in mental health.
- **Open-Source Platforms:** Development of open-source, clinically validated tools for widespread adoption.
- **Education and Training:** Training clinicians to understand and use AI-based tools effectively.

References

1. Dwyer, D.B., Falkai, P., & Koutsouleris, N. (2018). Machine learning approaches for clinical psychology and psychiatry. *Annual Review of Clinical Psychology*, 14, 91-118.
2. Chekroud, A.M., et al. (2016). Cross-trial prediction of treatment outcome in depression: a machine learning approach. *The Lancet Psychiatry*, 3(3), 243-250.
3. Iniesta, R., Stahl, D., & McGuffin, P. (2016). Machine learning, statistical learning and the future of biological research in psychiatry. *Psychological Medicine*, 46(12), 2455-2465.
4. Wolfers, T., et al. (2015). From pattern classification to stratification: towards conceptualizing the heterogeneity of schizophrenia spectrum disorders. *Schizophrenia Bulletin*, 41(4), 761-765.
5. Shatte, A.B.R., Hutchinson, D.M., & Teague, S.J. (2019). Machine learning in mental health: a scoping review of methods and applications. *Psychological Medicine*, 49(9), 1426-1448.
6. Sadegh-Zadeh, K. (2015). Handbook of analytic philosophy of medicine. Springer.
7. Kumar, A., & Gopal, M. (2018). Fuzzy expert systems and their applications in mental health. *Expert Systems with Applications*, 97, 205-217.

10.48047/jocaaa.2023.31.01.45

8. Vieira, S., Pinaya, W.H., & Mechelli, A. (2017). Using deep learning to investigate the neuroimaging correlates of psychiatric and neurological disorders: Methods and applications. *Neuroscience & Biobehavioral Reviews*, 74, 58-75.
9. Plis, S.M., et al. (2014). Deep learning for neuroimaging: a validation study. *Frontiers in Neuroscience*, 8, 229.
10. Chen, S., et al. (2020). Hybrid intelligent systems for medical diagnosis. *Artificial Intelligence in Medicine*, 103, 101779.
11. Kaur, M., & Kumari, S. (2021). Genetic-fuzzy expert system for diagnosis of major depressive disorder. *Expert Systems with Applications*, 167, 114104.
12. Razzak, M.I., Imran, M., & Xu, G. (2019). Big data analytics for preventive medicine. *Neural Computing and Applications*, 32, 4417-4451.
13. Gooding, P. (2019). Mapping the rise of digital mental health technologies: Emerging issues for law and society. *International Journal of Law and Psychiatry*, 67, 101498.
14. Orrù, G., et al. (2020). Machine learning in psychiatry: a brief introduction. *Advances in Experimental Medicine and Biology*, 1191, 341-358.
15. Mehta, N., & Pandit, A. (2018). Concurrence of big data analytics and healthcare: A systematic review. *International Journal of Medical Informatics*, 114, 57-65.
16. Sheikhpour, R., et al. (2017). A survey on semisupervised learning methods and applications. *Pattern Recognition*, 64, 141-158.
17. Rashid, B., et al. (2016). Classification of schizophrenia and bipolar patients using static and dynamic resting-state fMRI brain connectivity. *NeuroImage*, 134, 645-657.
18. Li, H., et al. (2020). A review on the diagnosis of schizophrenia using machine learning and neuroimaging. *Brain Imaging and Behavior*, 14, 552-573.
19. Nair, S., & Sathiaseelan, J.G. (2021). Fuzzy logic based intelligent system for psychiatric disorder prediction. *Biocybernetics and Biomedical Engineering*, 41(2), 571-587.
20. Acharya, U.R., et al. (2019). Automated detection of depression using EEG signals: A review. *Biomedical Signal Processing and Control*, 52, 263-274.
21. Denecke, K., & Bamidis, P. (2021). Artificial intelligence for mental health and mental illnesses: an overview. *Current Opinion in Psychiatry*, 34(5), 505-510.
22. Rahman, M.M., et al. (2022). Explainable artificial intelligence approaches in mental health: A systematic review. *Journal of Biomedical Informatics*, 128, 104010.
23. Martinc, M., et al. (2021). Multimodal data fusion for mental disorder prediction. *IEEE Access*, 9, 112267-112281.
24. Choudhury, M.D., & De, S. (2014). Mental health discourse on reddit: Self-disclosure, social support, and anonymity. *Proceedings of ICWSM*, 8, 71-80.
25. Li, X., et al. (2022). Machine learning for mental health in social media: Bibliometric study. *JMIR Mental Health*, 9(2), e30481.