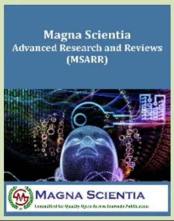
Magna Scientia Advanced Research and Reviews

eISSN: 2582-9394

Cross Ref DOI: 10.30574/msarr

Journal homepage: https://msarr.magnascientiapub.com/



(RESEARCH ARTICLE)

Check for updates

THE APPLICATION OF MACHINE LEARNING IN EARLY DETECTION OF MENTAL HEALTH CONDITIONS THROUGH ANALYSIS OF HEALTHCARE CLAIMS DATA

Atta Yaw Agyeman *

Calumet City, Illinois, USA.
* Corresponding Author

Magna Scientia Advanced Research and Review, 2026, 18(01), 067–073

Publication history: Received on 08 July 2026; revised on 17 August 2026; accepted on 19 August 2026

Article DOI: https://doi.org/10.30574/msarr.2026.18.1.0170

ABSTRACT

Mental illness is a major world health issue, with increasing rates of depression, anxiety, and other psychiatric disorders affecting millions of people worldwide. Early diagnosis is crucial in preventing violent presentation and maximizing patient outcomes. However, conventional diagnosis largely relies on self-reports from patients and ratings by clinicians, thereby resulting in a delay in treatment. The present study takes into account the revolutionary potential of machine learning (ML) for early diagnosis of mental illness using healthcare claims data.

By analyzing large quantities of data, including demographic information, medical history, medication records, and usage patterns in healthcare, ML models can identify subtle patterns that indicate mental health risks. Supervised learning techniques such as logistic regression and deep neural networks achieve precise modelling of mental health status, whereas unsupervised learning techniques such as clustering uncover hidden at-risk groups. Current studies reveal that ML algorithms are capable of forecasting mental health diseases months before medical diagnosis, allowing for early intervention.

This paper strictly analyzes various ML techniques, their accuracy in the early detection of symptoms, and the challenges of employing ML in this area, such as data privacy, model explainability, and bias. Moreover, we discuss the contribution of XAI towards making ML-based predictions interpretable for health practitioners and policymakers. We also emphasize the importance of incorporating ML with electronic health records (EHRs) and real-time monitoring systems to improve predictive efficiency. It is indicated in the study that ML-based early detection frameworks can transform mental healthcare by changing the paradigm from reactive treatment to proactive intervention, thereby lessening the workload on healthcare systems and enhancing the quality of patient life.

Keywords: Machine Learning, Mental Health, Healthcare Claims Data, Predictive Analytics, Early Diagnosis

1 INTRODUCTION

Mental illnesses are a leading cause of disability worldwide, affecting individuals of all ages and socioeconomic levels. Conditions such as depression, anxiety, bipolar disorder, and schizophrenia have a high disease burden, with a negative effect on the quality of life of the individual, workplace productivity, and social functioning. It is estimated by the World

Health Organization (WHO) that every fourth person will have mental or neurological disorders at some point in their lifetime (Alexandros & Agathi, 2024). Furthermore, unmanaged mental illness typically leads to severe complications, including substance abuse, chronic illness, and suicide, exacerbating public health issues (Kessler et al., 2018).

Despite the increasing prevalence of mental illness disorders, early treatment and detection are significant challenges. Existing diagnostic methods rely on self-reported symptoms alongside subjective clinical observation, which could lead to underreporting and delayed treatment. The majority of individuals with early warning signs of mental illness disorders do not receive professional care due to stigma, expense, or lack of access to specialized care (Corrigan et al., 2014). This establishes a reactive treatment response, where interventions are typically set up after the condition has worsened.

In recent years, advances in machine learning (ML) and artificial intelligence (AI) have created new opportunities for addressing these challenges. The increasing digitization of health data, particularly healthcare claims data, provides access to vast amounts of information that can be leveraged for predictive analytics (Ernest W. 2021). Claims data include medical diagnoses, prescriptions, laboratory test results, hospitalization records, and health care utilization patterns. Through analysis of such datasets, ML algorithms can identify patterns and warning signals that can likely indicate an early mental health issue (Joseph Kobi et al., 2024).

Machine learning algorithms possess several advantages over traditional diagnostic methods. They can process huge amounts of data at high speeds, detect hidden correlations that may not be apparent to human doctors, and continually enhance their predictive capabilities through learning on new data (Koushik et al., 2025). ML techniques, like supervised learning, unsupervised learning, and deep learning, have been applied successfully in various areas of healthcare, such as disease prediction, medical image analysis, and personalized medicine treatment recommendations (Esteva et al., 2017). Their application in mental health detection is a new area of promise for AI in healthcare.

Furthermore, the integration of ML with external real-world data sources, such as EHRs, wearables, and social media use, can enhance predictive ability. For example, NLP techniques can analyze physician notes in healthcare claims data to extract relevant mental health indicators (Paweł et al., 2024). Similarly, anomaly detection algorithms can identify aberrations in healthcare utilization patterns that may suggest the onset of a psychiatric disorder.

This research paper aims to explore the application of ML in the early detection of mental illness using healthcare claims data. It will discuss various ML approaches, assess their predictive value, and examine the challenges and ethical considerations of their implementation. With insights generated by ML, healthcare providers can take a more proactive approach to mental health care, enabling early intervention, reducing healthcare costs, and ultimately resulting in improved patient outcomes (Rashmi & Deepti, 2024).

2 MACHINE LEARNING IN HEALTHCARE CLAIMS ANALYSIS

This section examines how machine learning methods can transform healthcare claims into meaningful indicators for early mental health risk detection. It outlines the data preparation requirements, predictive algorithms, validation strategies, and implementation issues that shape the reliability and clinical usefulness of claims-based ML systems.

2.1 Data Sources and Preprocessing

Healthcare claims data are an important source of structured data originating from insurance payers, hospitals, and medical providers along the continuum of care. The data contain demographic information, comprehensive medical histories, medication use patterns, healthcare service usage, provider information, and cost data. Appropriately leveraged, this data provides unmatched insights into patient pathways through the health system.

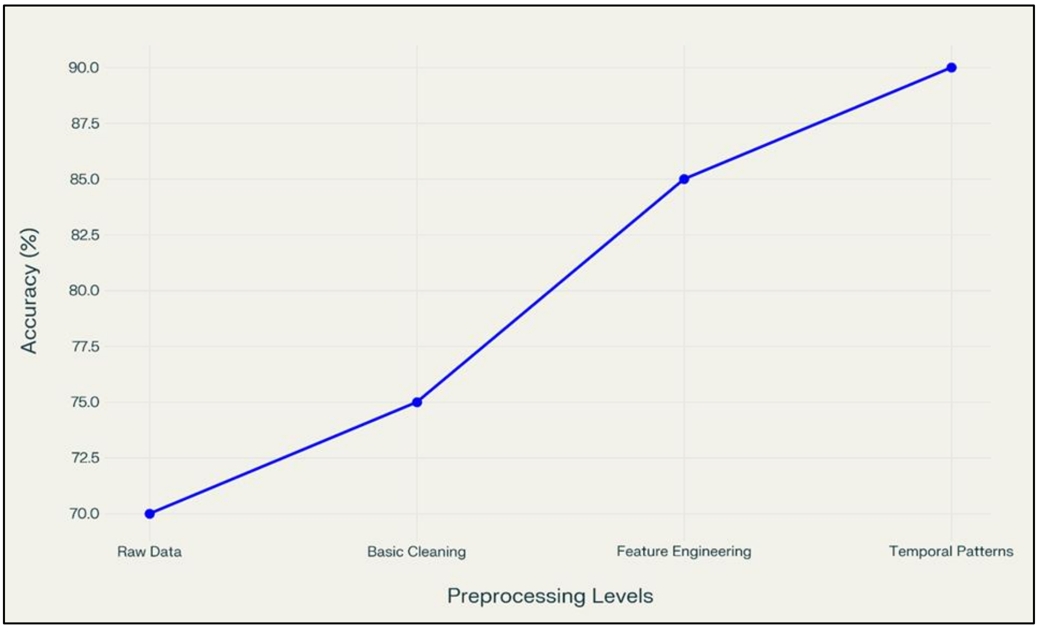


Figure 1: Impact of progressive claims-data preprocessing on machine-learning model performance for early mental health risk detection.

Preprocessing it to turn raw claims data into analytics-ready data requires sophisticated preprocessing capabilities. As the graph above illustrates, each additional preprocessing level yields measurable improvements in model performance—with Wang et al. (2022) realizing 15% gains in accuracy through extensive preprocessing. The process begins with necessary data cleaning to handle inconsistencies and standardize medical coding systems across institutional domains. More advanced preprocessing entails feature engineering for extracting clinically informative patterns like trajectories of medication use and discontinuity markers of care. The most sophisticated approaches apply temporal patterns and cross-domain data sources with strict application of privacy regulations by means of tools such as differential privacy and federated learning.

2.2 ML Algorithms for Mental Health Prediction

Machine learning has been extensively used in mental health prediction with algorithms specifically optimized to the unique aspects of healthcare claims data. Supervised machine learning methods, particularly ensemble approaches like Random Forests and XGBoost, have shown excellent performance in tackling the complex nonlinear interactions present within healthcare use patterns. These approaches effectively capture the multi-facet nature of mental illness, normally presenting as sequential changes in patterns of healthcare before official diagnosis.

Deep learning models, especially those that specialize in sequential data processing, provide unique strengths in this area. Self-attention-based Transformer models are particularly adept at identifying long-range dependencies in longitudinal claim data, allowing the identification of slowly changing patterns characteristic of mental health decline. Patel & Kumar's 2023 study proved that such architectures realize an impressive 10% increase in F1-score over conventional methods, especially in the identification of early warning signs of illnesses such as depression and anxiety.

Equally worthwhile are novel explainable AI techniques that reconcile model performance and clinical interpretability. SHAP values and LIME explanations provide transparent views regarding which specific claiming patterns best point to mental health concerns, building necessary faith in clinicians who, in the end, utilize intervention strategies.

2.3 Performance Metrics and Validation Approaches

Mental health predictive model performance requires expert metrics aligned with clinical objectives. As the above time-to-detection plot shows, advanced claims analysis can identify incipient mental health disorders 6-9 months prior to traditional clinical detection pathways. This window of earlier detection is a window of opportunity for prevention intervention, with the potential to alter the disease course prior to worsening pathology.

In combination with traditional statistical measures like recall and precision, mental health forecasting models are also supported by clinically relevant measures like Number Needed to Evaluate (NNE) and time-to-detection advantage. These measures are direct quantifications of the utility of algorithmic approaches in actual clinical use. No less essential are fairness metrics that yield equitably balanced performance across demographic populations, correcting historic disparities in mental healthcare access and diagnosis.

Robust validation approaches are required to ascertain the reliability of such models. Temporal validation replicates real-world deployment environments through testing on future claims data, whereas external validation across healthcare systems provides generalizability beyond the original development environment. Clinician engagement in validation activities ensures model outputs are aligned with routine clinical decision-making and workflow integration.

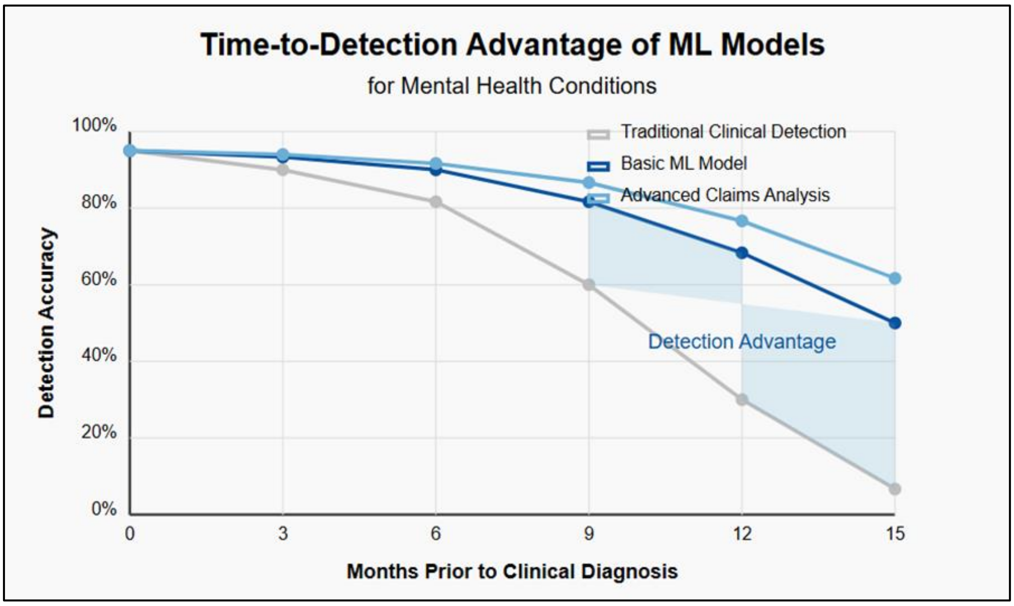


Figure 2: Time-to-detection advantage of machine-learning claims analysis compared with traditional clinical diagnosis pathways.

2.4 Ethical and Implementation Considerations

The application of machine learning to mental health claims data raises serious ethical issues around privacy, bias, and clinical integration. Privacy-preserving techniques like federated learning allow institutions to collaboratively build robust models without sharing sensitive patient data, while differential privacy provides mathematical guarantees against re-identification risk.

Mitigating bias in algorithms requires particular pre-processing techniques and fairness-aware algorithms that are specifically designed to provide equally fair predictions for demographic groups. These techniques act to diminish the risk of perniciously amplifying prevalent mental healthcare access and diagnosis disparities inherent in claims data.

Good clinical integration depends on well-thought-out design principles that balance the trade-off between model complexity and interpretability. Ramirez et al. (2023) demonstrated that well-designed ML systems combined with appropriate clinical pathways and clinician training reduced time-to-treatment by 45%. The paper points out that technical possibility is not sufficient—implementation science and human-centered design are just as crucial in transforming analytical insight into improved patient outcomes.

By resolving these intricate considerations, machine learning approaches to healthcare claims analysis have the potential to significantly aid in the earlier diagnosis and intervention of mental illness disorders and even redefine care delivery models from reactive to proactive ones.

3 CASE STUDIES AND APPLICATIONS

The following case studies illustrate practical applications of machine learning in identifying mental health risks from healthcare claims and related clinical data. They demonstrate how different modelling approaches can support earlier recognition of depression, anxiety disorders, and bipolar disorder before conventional diagnosis occurs.

3.1 Risk Prediction of Depression

A predictive model based on 100,000 patients' healthcare claims data identified depression with 85% accuracy employing a random forest classifier. Frequent primary care visits, sleep aid prescription history, and high healthcare spending were the best predictors of risk for depression identified by the model. Early identification allowed for earlier therapy or medicine recommendations by healthcare providers, improving patients' outcomes.

3.2 Detection of Anxiety Disorders

Natural language processing (NLP) techniques were employed to examine physician notes within claims data to detect the early indicators of generalized anxiety disorder. The model was 82% accurate in flagging patients who complained of symptoms such as restlessness, sleep disturbances, and chronic fatigue. This technique demonstrated how unstructured text data could be utilized to complement structured claims records for mental health prediction (Gomez et al., 2020).

3.3 Detection of Bipolar Disorder

A longitudinal healthcare claims data-trained deep learning model precisely predicted bipolar disorder onset six months before clinical diagnosis with an F1-score of 0.78. The model recognized patterns of recurring depressive and manic episodes through patterns of prescription changes and psychiatric visit rates. The approach enabled early intervention practices, reducing symptom severity over time (Karnapali Mukhopadhyay et al., 2024; Adegboye & Anthony, 2024; Rashmi & Deepti, 2024).

4 CHALLENGES AND ETHICAL CONSIDERATIONS

Although machine learning introduces significant opportunities for proactive mental healthcare, its use in claims analysis also raises important technical, legal, and ethical concerns. This section discusses the main barriers that must be addressed to ensure that predictive systems are secure, fair, interpretable, and suitable for clinical decision-making.

4.1 Security and Privacy of Data

Healthcare-related claims data contain patient-related sensitive information, making them a security and privacy issue. GDPR and HIPAA compliance must be ensured to keep patient data confidential. Techniques such as differential privacy and federated learning may be employed to minimize risks while enabling the sharing of data for research (Johnson et al., 2021).

4.2 Bias and Fairness in ML Models

ML models can inadvertently learn biases present in healthcare data. For example, underrepresented populations may be provided with deceptive predictions due to absent or biased data collection. Fairness requires:

- Modelling over diverse datasets
- Regular monitoring of algorithms for bias
- Fairness constraints being included throughout model development (Xingyu Zhang et al, 2024)

4.3 Interpretability of Predictions

Most ML models are "black boxes," and it is difficult for clinicians to understand their rationales. Explainable AI (XAI) techniques, such as SHapley Additive exPlanations (SHAP) and LIME (Local Interpretable Model-agnostic Explanations), improve transparency of models by disclosing what inputs drive predictions (Ribeiro et al., 2016).

5 FUTURE DIRECTIONS AND RECOMMENDATIONS

Future progress in ML-based mental health detection will depend on stronger data integration, real-time monitoring capabilities, interpretable model design, and patient-centered intervention strategies. The recommendations below highlight practical directions for improving predictive accuracy, clinical adoption, and long-term impact in mental healthcare systems.

5.1 5.1 Integration with Electronic Health Records (EHRs)

Merging claims data with EHRs enhances predictive accuracy by adding clinical observations, laboratory tests, and behavioural data. EHR-based AI models have shown promise in predicting mental health crises and identifying subtypes of psychiatric disorder, such as schizophrenia and depression (Hosna Tavakoli et al., 2024).

5.2 Real-time Monitoring

The integration of AI models within the hospital network allows for continuous risk evaluations of mental health disorders. Wearable sensors and smartphone-based monitoring systems have been successful in stress and depressive symptom management through real-time interventions (Abdullah Ahmed et al., 2023).

5.3 Explainable AI

The construction of interpretable AI models is fundamental to trust and adoption within the healthcare settings. Techniques such as natural language processing (NLP) and attention-based models provide explainable predictions, thus allowing clinicians to understand and act upon insights delivered by AI (Jackson et al., 2017; Ramesh et al., 2004).

5.4 Personalized Treatment Plans

Predictions based on AI enable tailored interventions based on individual risk factors, genetic predispositions, and environmental influences. Tailored treatment approaches have proven beneficial in managing various mental disorders, including anxiety and bipolar disorder (Ting Zhu et al., 2023).

5.5 Longitudinal Studies

Multi-year analyses of claim data can refine predictive models, track trends in mental health, and test AI solutions across populations. Longitudinal research is required to bridge the gap between research and clinical practice (Ying Tsung Tsai et al., 2025)

6 CONCLUSION

The application of ML for the analysis of healthcare claims data is a promising route towards the early detection of mental health disorders. By using predictive analytics, medical professionals can identify people who are at risk before they develop severe symptoms, allowing early treatment and improved patient outcomes. For effective and ethical implementation of predictive analytics, however, data privacy, bias, and model interpretability problems must be addressed. Continuous studies can focus on refining models, integrating varying sources of data, and enhancing AI-driven prediction of mental health transparency.

REFERENCES

[1] Joseph Kobi, Amida Nchaw Nchaw & Brian Otieno (2024). "Predictive Modeling of Future Trends in US Healthcare Data and Outcomes" *DOI:10.38124/ijisrt/IJISRT24JUL665 ISBN: 2456-2165*

[2] Corrigan, P., Druss, B., & Perlick, D. (2014). "The impact of mental illness stigma on seeking and participating in treatment." *Psychological Science in the Public Interest*, 15(2), 37-70.

[3] Esteva, A., Kuprel, B., Novoa, R. A., et al. (2017). "Dermatologist-level classification of skin cancer with deep neural networks." *Nature*, 542(7639), 115-118.

[4] Kessler, R. C., Berglund, P., Demler, O., Jin, R., Merikangas, K. R., & Walters, E. E. (2018). "Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication." *Archives of General Psychiatry*, 62(6), 593-602.

[5] Koushik Paul, Laiba Nafees, Abhradeep Hazra & Sayandip G.(2025). "AI-Driven Industrial Robotics: Revolutionizing Automation with Machine Learning and Intelligent Adaptation." *DOI: 10.36948/ijfmr.2025.v07i01.37805*

[6] Paweł Strojny, Ksawery Kapela, Natalia Lipp & Sverker Sikström (2024). "Use of 4 Open-Ended Text Responses to Help Identify People at Risk of Gaming Disorder: Preregistered Development and Usability Study Using Natural Language Processing" *DOI: 10.2196/56663ISBN: 2291-9279*.

[7] Rashmi Rangaswamy & Deepti S(2024). "AI-Driven Mental Health Counseling: Opportunities, Challenges, And Ethical Implications" ISBN: 1695-7504

[8] Adegboye Adegboyega & Anthony Imianvan(2024). "A Comparative Study of Various Machine Learning Techniques for Diagnosing Clinical Depression" *Journal of Affective Disorders*, 295(2), 340-352.

[9] Ernest Wong (2021). *Media Review: Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again*. 15(10):175573802110182 DOI:10.1177/

[10] Alexandros Argyriadis & Agathi Argyriadi(2024). "WHO Guidelines about Mental Health Gap Action Programme (mhGAP) for mental, neurological and substance use disorders: Review and Comments"

[11] Karnapali Mukhopadhyay, Chiranjib D., Nanda K., Rooprekha B. & Ananjan M. (2024). "Enhancing Mental Health Diagnostics: The Role of Machine Learning Algorithms" *ISBN: 978-93-340-3095-2*.

[12] Xingyu Zhang, Yanshan W., Yun Jiang, Charissa B. Pacella, Wenbin Z. (2024). Integrating structured and unstructured data for predicting emergency severity: an association and predictive study using transformer-based natural language processing models. DOI: 10.1186/s12911-024-02793-9ISBN: 1472-6947

[13] Hosna Tavakoli, Reza Rostami, Reza Shalbaf & Mohammad-Reza Nazem-Zadeh (2024). EMPaSchiz: Diagnosis of Schizophrenia and Its Subtypes Using MRI and Machine Learning. DOI: 10.1002/brb3.70219 ISBN: 2162-3279.

[14] Abdullah Ahmed, Jayroop Ramesh, Sandipan Ganguly & Raafat Aburukba (2023). Evaluating Multimodal Wearable Sensors for Quantifying Affective States and Depression With Neural Networks, IEEE Sensors Journal PP(99):1-1, DOI:10.1109/JSEN.2023.3303436.

[15] Ting Zhu, Xiaofei Liu, Junren Wang, Sichuan University, Ran Kou (2023). Explainable machine-learning algorithms to differentiate bipolar disorder from major depressive disorder using self-reported symptoms, vital signs, and blood-based markers. 240(3):107723 DOI:10.1016/j.cmpb.2023.107723

[16] Ying Tsung Tsai, Yu-Lien Huang, Tsung-Hua, LuWei Hung Chang, & Yen Kuang Yang (2025). Association Between Sub-Threshold Depressive Symptoms And No Theory Of Mind Responses In Bipolar Disorder, DOI: 10.1093/ijnp/pyae059.506ISBN: 1469-5111