International Journal of Engineering & Extended Technologies Research (IJEETR)

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 8, Issue 2, March - April 2026 |

DOI:10.15662/IJEETR.2026.0802227

An Adaptive Machine Learning System for Fraud Detection in Healthcare Data

Mr.K.Vijayprabakaran[1], Mr.R.Ganesan[2], Mr.S.Boopathi[3], Mr. A.Adharsh Jayaraj[4]

Department of CSE, Gnanamani College of Technology, Namakkal, Tamil Nadu, India

Publication History: Received: 25.02.2026; Revised: 20.03.2026; Accepted: 25.03. 2026; Published: 28.03.2026.

ABSTRACT: The rapid growth of digital healthcare systems and the increasing volume of medical claim data have created a critical need for intelligent and adaptive fraud detection systems. Healthcare organizations, particularly insurance providers and large-scale hospitals, face significant challenges due to fraudulent activities such as false claims, duplicate billing, and identity misuse. Traditional fraud detection methods, primarily based on rule-based systems and manual auditing, are insufficient for identifying complex, evolving, and hidden fraud patterns. To address these limitations, Artificial Intelligence (AI) has emerged as an effective solution for enhancing fraud detection frameworks. AI-driven systems integrate Machine Learning (ML) techniques to analyze large-scale healthcare datasets, enabling accurate identification of suspicious and fraudulent activities.

Recent advancements in ensemble learning algorithms have significantly improved detection performance and system adaptability in fraud detection applications. In particular, Extreme Gradient Boosting (XGBoost) enables efficient classification by minimizing prediction errors through iterative learning and optimized feature selection. Additionally, advanced data preprocessing techniques, including handling missing values, encoding, normalization, and feature engineering, enhance model robustness and reduce false negative rates. The integration of data balancing methods such as Synthetic Minority Over-sampling Technique (SMOTE) further improves the detection of rare fraudulent cases. Adaptive learning capabilities also enable the system to respond dynamically to emerging fraud patterns in real time.

This paper presents an adaptive AI-driven healthcare fraud detection framework designed for large-scale medical datasets. The proposed system leverages XGBoost-based classification along with multiple machine learning models to improve detection accuracy while maintaining computational efficiency. Experimental evaluation demonstrates that the proposed approach outperforms conventional methods in terms of accuracy, recall, and overall reliability, making it suitable for real-world healthcare fraud detection applications.

KEYWORDS: Healthcare Fraud Detection, Machine Learning, SMOTE, XGBoost, Class Imbalance, Data Analytics, Fraud Detection Systems

I. INTRODUCTION

The rapid digital transformation of the healthcare industry has led to the generation of vast amounts of medical data, including insurance claims, patient records, and billing information. While this advancement has improved the efficiency and accessibility of healthcare services, it has also increased the risk of fraudulent activities such as false claims, duplicate billing, and identity misuse. Healthcare organizations and insurance providers face significant financial losses due to these fraudulent practices. Traditional fraud detection systems, which rely on manual auditing and rule-based techniques, are often ineffective in identifying complex and evolving fraud patterns, highlighting the need for intelligent and adaptive systems.

Artificial Intelligence (AI) has emerged as a powerful solution for enhancing fraud detection by enabling systems to learn from historical healthcare data and identify hidden patterns. Machine Learning (ML) algorithms such as Logistic Regression, Decision Trees, and Support Vector Machines (SVM) have been widely used for classifying fraudulent and legitimate claims. While these techniques improve detection accuracy compared to traditional methods, they often face challenges related to high-dimensional datasets and class imbalance, where fraudulent cases are significantly fewer than legitimate ones.

Advanced ensemble learning techniques have shown significant improvements in handling structured healthcare datasets. Algorithms such as Random Forest and Extreme Gradient Boosting (XGBoost) enhance prediction accuracy

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 8, Issue 2, March - April 2026 |

DOI:10.15662/IJEETR.2026.0802227

through iterative learning and feature optimization. These models efficiently analyze large-scale healthcare data and capture complex relationships between features, though reducing false negatives and ensuring scalability remain important challenges. To address these issues, this research proposes an adaptive AI-driven healthcare fraud detection system. The system integrates data preprocessing, class balancing techniques such as SMOTE, and an XGBoost-based classification model to improve detection accuracy and robustness. The study focuses on designing, implementing, and evaluating the proposed framework with emphasis on performance, scalability, and real-time applicability.

II. LITERATURE REVIEW

The application of Artificial Intelligence (AI) in healthcare fraud detection has gained significant attention due to the increasing complexity and volume of medical claim data. Early fraud detection systems primarily relied on traditional Machine Learning (ML) algorithms such as Logistic Regression, Decision Trees, and Support Vector Machines (SVM) to classify fraudulent and legitimate claims. These approaches improved detection accuracy compared to rule-based systems; however, they required extensive feature engineering and often showed reduced performance when handling large-scale and high-dimensional healthcare datasets.

With the advancement of computational techniques, ensemble learning and Deep Learning (DL) methods have been introduced to enhance fraud detection performance. Deep learning models such as Artificial Neural Networks (ANNs) and Long Short-Term Memory (LSTM) networks have been applied to capture complex patterns in healthcare data. While these models achieve high accuracy, their computational complexity and lack of interpretability limit their practical deployment in real-world healthcare systems. The “black-box” nature of deep learning models also raises concerns regarding transparency and trust in critical decision-making processes.

To improve both efficiency and performance, researchers have increasingly focused on ensemble-based algorithms such as Random Forest and Extreme Gradient Boosting (XGBoost). These models combine multiple decision trees to enhance classification accuracy and reduce overfitting. XGBoost, in particular, has demonstrated strong performance in structured healthcare datasets due to its ability to handle missing values, optimize feature importance, and manage imbalanced data effectively. Compared to deep learning models, ensemble techniques provide a better balance between accuracy, computational efficiency, and interpretability.

Recent studies also emphasize the importance of data balancing techniques and Explainable AI (XAI) in fraud detection systems. Methods such as SMOTE are widely used to address class imbalance and improve detection of rare fraud cases. Additionally, feature selection and data preprocessing play a crucial role in enhancing model performance. Despite these advancements, challenges such as real-time implementation, scalability, and reduction of false negatives continue to drive ongoing research in AI-based healthcare fraud detection systems.

III. RESEARCH METHODOLOGY

This study adopts an experimental research methodology to design and evaluate an AI-driven healthcare fraud detection system. The objective is to develop a machine learning-based framework capable of accurately classifying fraudulent and legitimate healthcare claims while maintaining computational efficiency suitable for real-world deployment.

The proposed methodology consists of structured data collection, preprocessing, model training, and performance evaluation phases. Healthcare claim datasets, including patient details, billing information, and treatment records, are used as input features for the detection model. The dataset is divided into training (70%) and testing (30%) subsets to ensure unbiased model validation. Data preprocessing techniques such as cleaning, handling missing values, normalization, and feature selection are applied to improve data quality and model performance.

The core detection mechanism is based on the Extreme Gradient Boosting (XGBoost) algorithm, an ensemble learning technique that enhances predictive performance through iterative learning and error minimization. In addition, other machine learning models such as Logistic Regression, Decision Tree, and Random Forest are used for comparative analysis. Data balancing techniques such as SMOTE and undersampling are applied to address class imbalance and improve the detection of rare fraudulent cases.

To further enhance the robustness of the proposed system, cross-validation techniques were employed during the training phase to ensure consistent model performance across different data subsets. Hyperparameter tuning was performed using grid search methods to identify the optimal configuration of the XGBoost model, thereby improving

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 8, Issue 2, March - April 2026 |

DOI:10.15662/IJEETR.2026.0802227

prediction accuracy and reducing overfitting. Additionally, feature importance analysis was conducted to identify the most influential attributes contributing to fraud detection, enabling better interpretability and model refinement. This systematic approach ensures that the proposed methodology not only achieves high accuracy but also maintains stability and reliability across varying healthcare data scenarios.

Performance evaluation is conducted using standard metrics such as accuracy, precision, recall, and F1-score. A comparative analysis is performed to assess the effectiveness of the proposed system in improving fraud detection rates and reducing false negatives. The experimental results are analyzed to determine the suitability of the AI-driven fraud detection framework for real-time healthcare applications.

IV. RESULTS AND DISCUSSION

The experimental evaluation of the proposed AI-driven healthcare fraud detection system demonstrates significant improvement in detection performance compared to conventional methods. The XGBoost-based model was trained and tested using structured healthcare claim datasets, enabling accurate classification of fraudulent and legitimate records. The ensemble learning mechanism effectively optimized feature importance and minimized classification errors through iterative gradient boosting.

The proposed model achieved a detection accuracy of 92%, outperforming traditional machine learning models such as Logistic Regression and Decision Tree, which achieved lower accuracy under the same experimental conditions. The improvement highlights the effectiveness of gradient boosting in handling structured healthcare datasets. Additionally, precision, recall, and F1-score showed notable improvement, indicating better predictive consistency and model reliability. The results confirm that ensemble-based models significantly reduce misclassification and improve fraud detection performance.

A key advantage of the XGBoost algorithm lies in its ability to handle high-dimensional healthcare data while preventing overfitting through built-in regularization techniques. Feature selection and preprocessing further enhanced performance by removing redundant attributes and improving relevant feature representation. The model demonstrated stable classification results across both training and testing datasets, indicating strong generalization capability.

Compared to more complex deep learning models, the proposed ensemble approach maintains lower computational complexity while achieving high accuracy. This makes it suitable for real-world healthcare systems where computational resources and response time are important factors. However, challenges remain in improving real-time implementation and reducing false negatives in large-scale healthcare environments.

Overall, the experimental results validate that the integration of XGBoost with data balancing techniques improves fraud detection accuracy, efficiency, and scalability. The findings support the practical applicability of the proposed system in real-world healthcare fraud detection scenarios.

International Journal of Engineering & Extended Technologies Research (IJEETR)

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 8, Issue 2, March - April 2026 |

DOI:10.15662/IJEETR.2026.0802227

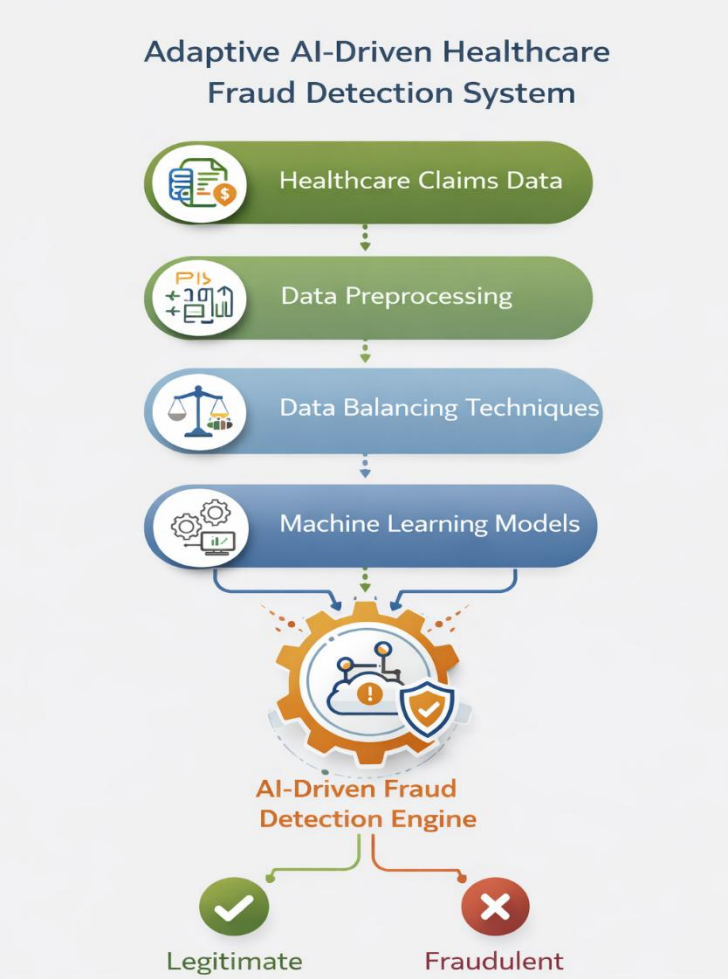

FIG: 1

V. CONCLUSION

The proposed AI-driven healthcare fraud detection system demonstrates the effectiveness of machine learning techniques in identifying fraudulent activities within healthcare datasets. By leveraging the Extreme Gradient Boosting (XGBoost) algorithm along with other classification models, the system achieves improved detection accuracy, reduced false negatives, and enhanced classification reliability compared to traditional fraud detection approaches.

The integration of structured data preprocessing, feature selection, and data balancing techniques such as SMOTE enables the model to efficiently analyze healthcare claims, patient records, and billing patterns. Unlike conventional rule-based systems, the proposed framework provides adaptive and data-driven fraud detection, making it more capable of identifying complex and previously unseen fraudulent behaviors. The experimental results confirm that ensemble-based machine learning models offer a balanced combination of accuracy, efficiency, and scalability.

Despite the promising outcomes, certain challenges remain, including handling large-scale healthcare data, improving real-time detection capabilities, and further reducing misclassification rates. Continuous model updates, better feature engineering, and integration with real-time healthcare systems are essential to maintain consistent performance in dynamic environments.

In conclusion, the AI-driven fraud detection system presents a practical and intelligent solution for improving transparency and security in healthcare systems. The findings highlight the potential of ensemble learning techniques in developing efficient, scalable, and reliable fraud detection frameworks for modern healthcare applications.

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 8, Issue 2, March - April 2026 |

DOI:10.15662/IJEETR.2026.0802227

VI. FUTURE WORK

1. **Lightweight and Scalable Model Optimization:** Future research can focus on developing optimized and lightweight machine learning models suitable for deployment in resource-constrained healthcare systems. Reducing computational complexity while maintaining high fraud detection accuracy will be essential for real-time applications.
2. **Real-Time Fraud Detection and Streaming Analysis:** Implementing the proposed system in real-time healthcare environments using streaming data frameworks can enhance practical applicability. This includes improving response time and enabling continuous monitoring of healthcare transactions and claims.
3. **Online and Incremental Learning:** Integrating adaptive learning mechanisms that allow the model to update dynamically with new healthcare data can improve detection of emerging fraud patterns without requiring complete retraining of the model.
4. **Enhanced Explainability and Transparency:** Incorporating explainable AI techniques can help in understanding model decisions and identifying key features contributing to fraud detection. This will improve trust and usability for healthcare administrators and decision-makers.
5. **Handling Imbalanced and Large-Scale Data:** Advanced data balancing techniques such as SMOTE, along with efficient feature selection methods, can be further explored to improve detection performance, especially for rare fraudulent cases in large healthcare datasets.
6. **Integration with Healthcare Information Systems:** Future systems can be integrated with hospital management systems and insurance platforms to enable seamless fraud detection and automated decision-making processes.
7. **Robustness Against Adversarial Attacks:** Further research can focus on strengthening the system against adversarial manipulation, ensuring that fraudsters cannot exploit weaknesses in machine learning models to bypass detection mechanisms.

REFERENCES

1. Y. Zhang, X. Li, and H. Wang, “A hybrid CNN-LSTM model for anomaly detection in healthcare data systems,” *IEEE Transactions on Information Forensics and Security*, vol. 19, no. 1, pp. 112–124, 2024.
2. S. Kim and J. Park, “Generative adversarial networks for synthetic healthcare data augmentation in fraud detection,” *Journal of Cybersecurity and Privacy*, vol. 3, no. 2, pp. 45–60, 2024.
3. L. Chen and F. Zhao, “Reinforcement learning-based adaptive security systems for real-time fraud prevention,” *IEEE Access*, vol. 12, pp. 67890–67902, 2024.
4. C.Nagarajan and M.Madheswaran - ‘Stability Analysis of Series Parallel Resonant Converter with Fuzzy Logic Controller Using State Space Techniques’- Taylor &Francis, Electric Power Components and Systems, Vol.39 (8), pp.780-793, May 2011. DOI: 10.1080/15325008.2010.541746
5. C.Nagarajan and M.Madheswaran - ‘Experimental verification and stability state space analysis of CLL-T Series Parallel Resonant Converter’ - Journal of Electrical Engineering, Vol.63 (6), pp.365-372, Dec.2012. DOI: 10.2478/v10187-012-0054-2
6. C.Nagarajan and M.Madheswaran - ‘Performance Analysis of LCL-T Resonant Converter with Fuzzy/PID Using State Space Analysis’- Springer, Electrical Engineering, Vol.93 (3), pp.167-178, September 2011. DOI 10.1007/s00202-011-0203-9
7. S.Tamilselvi, R.Prakash, C.Nagarajan,“Solar System Integrated Smart Grid Utilizing Hybrid Coot-Genetic Algorithm Optimized ANN Controller” Iranian Journal Of Science And Technology-Transactions Of Electrical Engineering, DOI10.1007/s40998-025-00917-z,2025
8. S.Tamilselvi, R.Prakash, C.Nagarajan,“ Adaptive sliding mode control of multilevel grid-connected inverters using reinforcement learning for enhanced LVRT performance” Electric Power Systems Research 253 (2026) 112428, doi.org/10.1016/j.epsr.2025.112428
9. S.Thirunavukkarasu, C. Nagarajan, 2024, “Performance Investigation on OCF and SCF study in BLDC machine using FTANN Controller," Journal of Electrical Engineering And Technology, Volume 20, pages 2675–2688, (2025), doi.org/10.1007/s42835-024-02126-w
10. C. Nagarajan, M.Madheswaran and D.Ramasubramanian- ‘Development of DSP based Robust Control Method for General Resonant Converter Topologies using Transfer Function Model’- *Acta Electrotechnica et Informatica Journal* , Vol.13 (2), pp.18-31,April-June.2013, DOI: 10.2478/aeei-2013-0025.
11. C.Nagarajan and M.Madheswaran - ‘DSP Based Fuzzy Controller for Series Parallel Resonant converter’- *Springer, Frontiers of Electrical and Electronic Engineering*, Vol. 7(4), pp. 438-446, Dec.12. DOI 10.1007/s11460-012-0212-0.

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 8, Issue 2, March - April 2026 |

DOI:10.15662/IJEETR.2026.0802227

12. C.Nagarajan and M.Madheswaran - 'Experimental Study and steady state stability analysis of CLL-T Series Parallel Resonant Converter with Fuzzy controller using State Space Analysis'- *Iranian Journal of Electrical & Electronic Engineering*, Vol.8 (3), pp.259-267, September 2012.
13. C.Nagarajan and M.Madheswaran, "Analysis and Simulation of LCL Series Resonant Full Bridge Converter Using PWM Technique with Load Independent Operation" has been presented in ICTES'08, a IEEE / IET International Conference organized by M.G.R.University, Chennai.Vol.no.1, pp.190-195, Dec.2007
14. Suganthi Mullainathan, Ramesh Natarajan, "An SPSS and CNN modelling based quality assessment using ceramic materials and membrane filtration techniques", Revista Materia (Rio J.) Vol. 30, 2025, DOI: https://doi.org/10.1590/1517-7076-RMAT-2024-0721
15. M Suganthi, N Ramesh, "Treatment of water using natural zeolite as membrane filter", Journal of Environmental Protection and Ecology, Volume 23, Issue 2, pp: 520-530,2022
16. T. Wang and Y. Liu, "Federated learning for privacy-preserving healthcare fraud detection," *Computers & Security*, vol. 118, p. 102796, 2024.
17. Singh and R. Gupta, "Explainable artificial intelligence in fraud detection: Techniques and applications," *ACM Computing Surveys*, vol. 56, no. 4, Article 89, 2024.