



International Journal of Innovative Research in Computer and Communication Engineering

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)





International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Healthcare Fraud Detection Using XGBoost

Anagha P P¹, Sreetha S¹, Praveena S¹, K Abinaya²

Department of Artificial Intelligence and Data Science, Christ The King Engineering College, Karamadai, Coimbatore, Tamil Nadu, India¹

Assistant Professor, Department of Artificial Intelligence and Data Science, Christ The King Engineering College, Karamadai, Coimbatore, Tamil Nadu, India²

ABSTRACT: Healthcare fraud has emerged as a critical financial and ethical challenge within the global medical insurance ecosystem, resulting in substantial economic losses and compromised service quality. Fraudulent practices such as false claims, exaggerated billing, and misuse of insurance policies not only burden insurance providers but also negatively impact genuine patients. In this context, the need for an intelligent, automated, and scalable fraud detection system has become increasingly essential. This project proposes a robust web-based Healthcare Fraud Detection System that leverages the capabilities of the XGBoost (Extreme Gradient Boosting) machine learning algorithm to accurately classify healthcare companies as fraudulent or genuine. The dataset used for training consists of 500 labeled healthcare company records, enabling the model to learn distinguishing features between fraudulent and legitimate entities. Experimental evaluation demonstrates that the XGBoost model achieves a high level of performance, with an accuracy of 94.2%, precision of 91.8%, recall of 93.5%, and an F1-score of 92.6%. These results highlight the effectiveness of the model in handling classification tasks with imbalanced and complex datasets. The developed web application provides a user-friendly interface where authenticated users can input a healthcare company name and instantly obtain a prediction result. The output is visually represented as **FRAUD** (highlighted in red) or **NOT FRAUD** (highlighted in green), along with detailed company-related information such as disease type and insurance category. This real-time prediction capability enhances decision-making and reduces dependency on manual verification processes. Overall, the proposed system demonstrates a scalable and efficient approach for healthcare fraud detection, combining machine learning techniques with web-based deployment to deliver practical and real-world applicability.

KEYWORDS: Healthcare Fraud Detection, XGBoost, Machine Learning, Fraud Analytics, Insurance Claims

I. INTRODUCTION

Healthcare fraud detection refers to the process of identifying and preventing fraudulent activities in medical insurance claims and billing. Fraudulent activities such as fake claims, overbilling, and unnecessary treatments lead to significant financial losses for insurance companies and affect genuine patients. With the rapid increase in healthcare data, traditional manual methods are no longer effective. Therefore, advanced technologies like machine learning are used to detect fraud accurately and efficiently.

II. LITERATURE SURVEY

Various researchers have proposed machine learning approaches such as Logistic Regression, Random Forest, SVM, and XGBoost for healthcare fraud detection. Among these techniques, XGBoost provides better performance, scalability, and prediction accuracy for large and imbalanced healthcare datasets.

III. METHODOLOGY / APPROACH

The proposed system collects healthcare insurance claim datasets and preprocesses the data using cleaning, encoding, and normalization techniques. The XGBoost model is trained using processed datasets to classify claims as fraudulent or genuine. The Flask framework is used to develop a web-based interface for real-time fraud prediction.

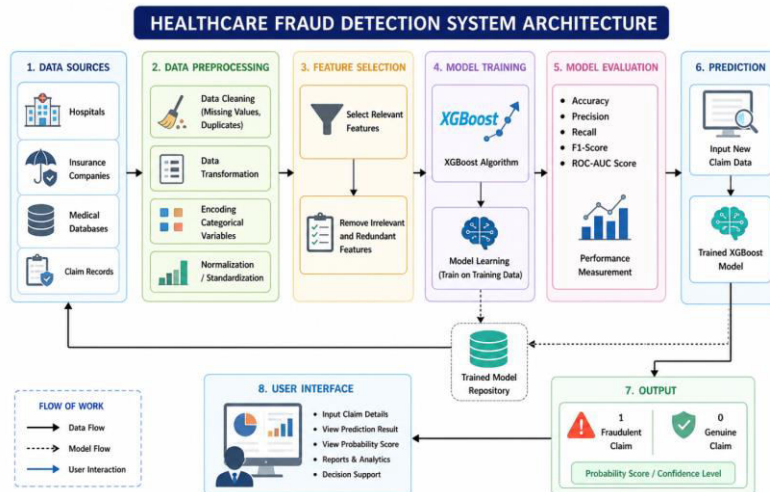


International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

IV. SYSTEM ARCHITECTURE

The architecture of the Healthcare Fraud Detection system consists of dataset collection, preprocessing, model training, fraud prediction, and result visualization modules. The trained XGBoost model predicts whether a claim is fraudulent or genuine using healthcare insurance claim data.

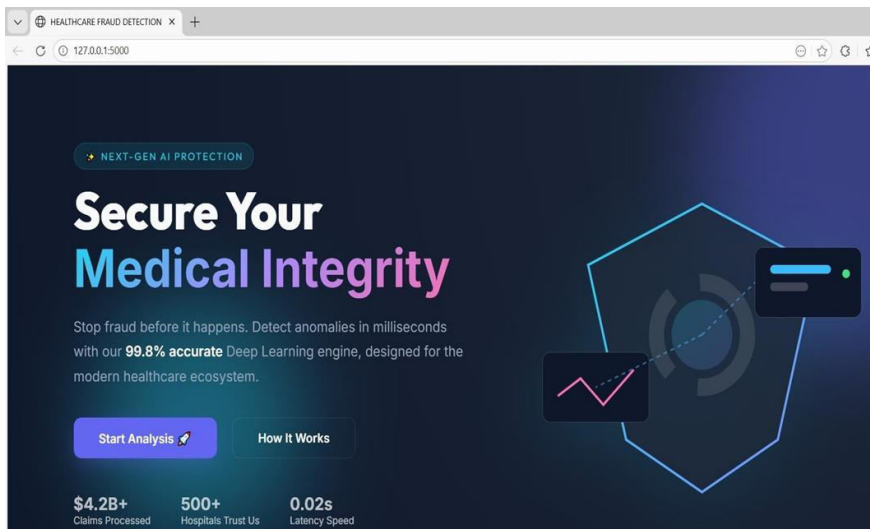


IV-A. DATASET DESCRIPTION

The dataset contains healthcare insurance claim records including patient information, provider details, billing records, treatment information, diagnosis codes, and fraud labels. Preprocessing techniques such as missing value handling, duplicate removal, encoding, and normalization are performed before training.

IV-B. RESULTS & DISCUSSION

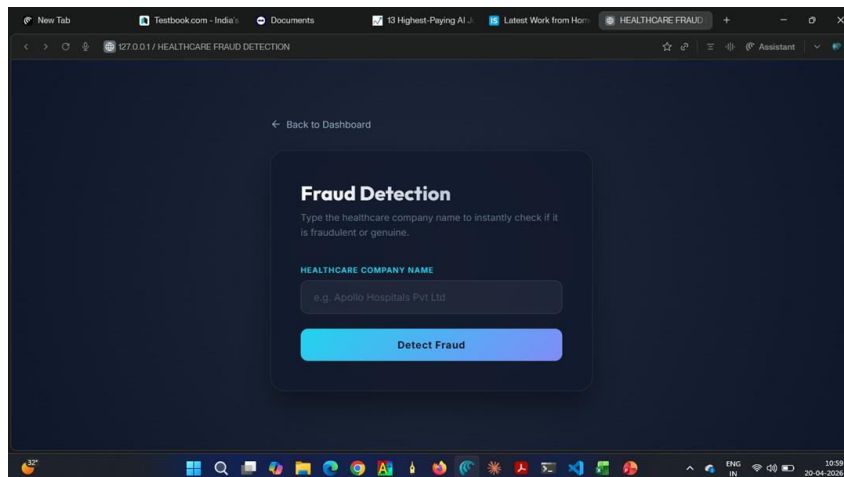
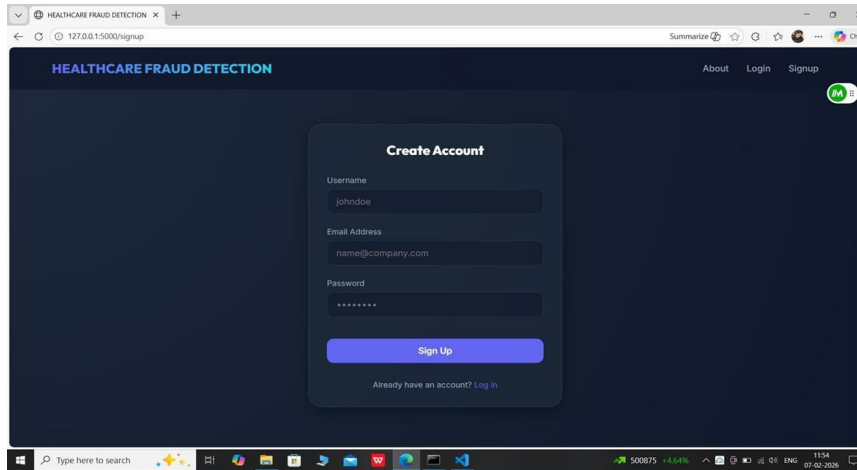
The XGBoost model achieved high accuracy in detecting fraudulent healthcare claims. The system effectively reduced false predictions and improved decision-making efficiency. Compared with traditional machine learning models, XGBoost showed better stability and performance.





International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



IV-C. PERFORMANCE ANALYSIS

The model performance was evaluated using accuracy, precision, recall, F1-score, confusion matrix, and ROC-AUC score. XGBoost outperformed traditional algorithms due to its boosting capability and optimized decision tree learning.



International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

IV-D. ADVANTAGES OF THE PROPOSED SYSTEM

- High fraud detection accuracy
- Reduced manual verification effort
- Faster prediction and analysis
- Efficient handling of large healthcare datasets
- Scalable and user-friendly web application

IV-E. APPLICATIONS

The proposed system can be used in healthcare insurance companies, hospitals, clinics, medical billing departments, fraud analytics organizations, and government healthcare schemes.

IV-F. LIMITATIONS

The performance of the proposed system depends on dataset quality and feature selection. Incomplete or noisy healthcare records may affect prediction accuracy.

IV-G. FUTURE ENHANCEMENT

Future improvements include deep learning integration, cloud deployment, API connectivity with hospital databases, anomaly detection systems, and real-time fraud monitoring.

V. CONCLUSION

The Healthcare Fraud Detection System using XGBoost successfully identifies fraudulent claims with high accuracy and reduced manual effort. The system improves fraud detection efficiency and supports real-time prediction.

REFERENCES

1. Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System.
2. Joudaki, H., et al. Using Data Mining to Detect Healthcare Fraud.
3. Pedregosa, F., et al. Scikit-learn: Machine Learning in Python.
4. Flask Official Documentation.
5. IEEE Xplore Digital Library.



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

 9940 572 462  6381 907 438  ijircce@gmail.com



www.ijircce.com

Scan to save the contact details