PCAS International Journal for Multidisciplinary Research (PCAS IJMR)
ISSN (Online): 3108-2807 Volume 3, Special Issue 1, Jan-Mar 2026

Advanced Fraudulent Claim Detection in Health Insurance Using Deep Learning and Business Intelligence

Dr. Surya Susan Thomas, Assistant Professor, Department of Computer Science with Data Science,
Patrician College of Arts and Science, Adyar, Chennai.
suryasusan@patrciciancollege.ac.in

Abstract

The rapid growth of digital healthcare systems has enhanced the risk of fraudulent activities in health insurance claims, leading to financial losses and inefficient resource allocation. Traditional fraud detection methods often fail to identify complex patterns in large healthcare datasets. This study proposes a fraud detection framework that integrates machine learning and deep learning techniques to analyse structured claim data and sequential patient records. Models such as Logistic Regression, Random Forest, Convolutional Neural Networks (CNN), and Long Short-Term Memory (LSTM) are used to detect fraudulent claims. The system incorporates data preprocessing, feature engineering, and evaluation metrics to improve detection accuracy while ensuring data privacy and regulatory compliance.

Keywords

Digital Healthcare, Deep Learning, Business intelligence, Insurance claim, Deep Learning

Introduction

The increasing prevalence of fraudulent activities in the health insurance sector has created a pressing need for advanced computational techniques to safeguard financial resources and ensure the fair allocation of healthcare services (Shungube et al., 2024). With the rapid growth of digital healthcare data, insurance providers are increasingly adopting intelligent analytical frameworks that combine deep learning architectures with traditional business intelligence tools. This integration enables insurers to analyze high-dimensional patient visit sequences and structured claim data more effectively, thereby improving the identification of suspicious or anomalous patterns (Fursov et al., 2022; Preez et al., 2024).

Compared to conventional rule-based or statistical approaches, these advanced models provide greater analytical capability by capturing complex relationships and hidden dependencies within large and heterogeneous healthcare datasets (Hasan, 2022; Muhammad et al., 2025). As a result, deep learning-based systems have become major and genuine tools for enhancing fraud detection accuracy and played a major role in more efficient claim management processes. However, the effectiveness of such models depends significantly on the quality of data preparation processes, including robust feature engineering and the availability of large, well-labeled datasets to reduce the risk of overfitting and ensure reliable model performance (Shahzadi et al., 2023).

Literature Review

Recent systematic reviews reveal a paradigm shift toward utilizing Convolutional Neural Networks and Long Short-Term Memory architectures to process longitudinal patient records (Chen et al., 2025). These studies demonstrate that while artificial neural networks provide robust classification accuracy for structured claim data, specialized attention-based networks are increasingly employed to capture the temporal nuances of fraudulent behavior (Farbmacher et al., 2020; Shungube et al., 2024). Moreover, the implementation of hybrid supervised and unsupervised learning frameworks facilitates the identification of emerging fraud typologies, thereby reducing the dependency on extensive manual annotation. Beyond purely internal records, advanced analytical systems further enhance detection precision by integrating heterogeneous external data sources, such as social media interactions, geographic indicators, and digital usage footprints, to uncover non-obvious relationships (Rey‐Puech et al., 2025). This convergence of deep learning and business intelligence not only fortifies internal risk management strategies but also fosters organizational trust by ensuring that claim processing remains resilient against increasingly sophisticated adversarial tactics (Ramachandran et al., 2024). Despite these advancements, the inherent "black-box" nature of such models necessitates a careful balance between high-dimensional pattern recognition and the regulatory demand for algorithmic transparency (Shungube et al., 2024).

Research Methodology

The proposed system architecture integrates deep learning models with business intelligence tools to detect fraudulent activities in health insurance claims. The architecture consists of

multiple interconnected layers that process healthcare claim data, extract meaningful patterns, and identify suspicious transactions.

1. Data Collection or Acquisition Layer

This part is responsible for acquiring data from multiple sources. It gathers internal health insurance claim records, including patient demographics, medical procedures, billing details, and provider information. Additionally, external data sources such as geographical indicators, social media signals, and digital interaction data may be incorporated to enrich the dataset and improve fraud detection accuracy.

2. Data Pre-processing Layer

In this layer, the collected raw data is looked into and transformed to ensure quality and consistency. Jobs such as handling missing values, removing duplicate records, encoding/decoding categorical variables, and normalizing numerical attributes are performed. Sequential patient visit records are also organized into time-based sequences to support temporal analysis.

3. Feature Engineering Layer

This layer generates meaningful features that help the model identify fraudulent patterns. Derived attributes include claim frequency, average claim cost, treatment sequence patterns, provider utilization trends, and time intervals between medical visits. “Feature selection techniques” are applied to reduce dimensionality and retain only the most relevant attributes.

4. Model Processing Layer

The processed features are fed into a hybrid machine learning and deep learning framework. Traditional models such as Logistic Regression, Decision Trees, and Random Forest are used as baseline classifiers. “Advanced deep learning architectures such as Convolutional Neural Networks (CNN)” analyze structured claim patterns, while Long Short-Term Memory (LSTM) networks capture temporal relationships in sequential patient visit data. These models learn complex relationships within healthcare claim data to identify potential fraud.

5. Fraud Detection and Classification Layer

In this layer, the trained models classify incoming claims as fraudulent or legitimate. Anomaly detection mechanisms are also applied to identify previously unseen fraud patterns.

The system assigns a fraud risk score to each claim, enabling insurance companies to prioritize suspicious cases for further investigation.

6. Evaluation and Validation Layer

Model working is evaluated using metrics such as Accuracy, Precision, Recall, F1-Score, and AUC-ROC. Cross-checking, also known as cross-validation techniques, helps ensure that the model generalises well to unseen data and reduces the risk of overfitting.

7. Visualization and Decision Support Layer

The final layer integrates business intelligence dashboards to present insights in an understandable format. Tools such as Power BI can visualize fraud patterns, claim distributions, and risk scores, enabling insurers and analysts to make informed decisions quickly.

8. Security and Compliance Layer

Since healthcare data is highly sensitive, this layer ensures data privacy, encryption, and compliance with healthcare regulations. Access control mechanisms and anonymization techniques protect patient information while maintaining analytical integrity.

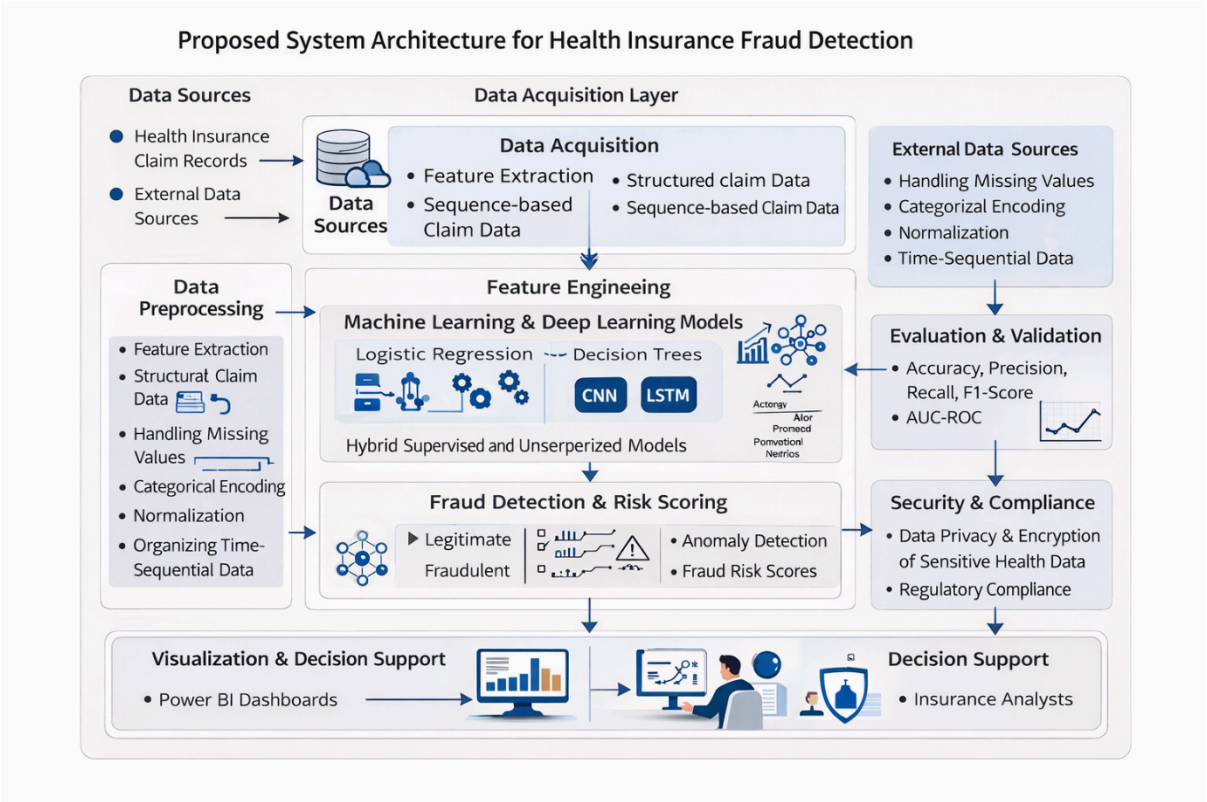

Fig 1: Proposed Methodology

Conclusion

The increasing complexity of fraudulent activities in health insurance claims necessitates the use of advanced analytical frameworks capable of identifying hidden and evolving fraudulent features. This study gives a comprehensive fraud detection framework that integrates traditional ML techniques with deep learning architectures to analyse both structured and sequential healthcare claim data. By incorporating models such as “Convolutional Neural Networks and Long Short-Term Memory networks”, the system is capable of capturing complex relationships and temporal patterns within patient visit sequences and claim histories.

Furthermore, the integration of robust feature engineering, hybrid supervised–unsupervised learning strategies, and business intelligence visualization tools enhances the accuracy and interpretability of fraud detection systems. The proposed architecture not only improves the detection of suspicious claims but also assists insurance analysts in making informed decisions through risk scoring and visual analytics. Additionally, the inclusion of data security and regulatory compliance mechanisms ensures the ethical handling of sensitive healthcare information.

Overall, the proposed framework demonstrates significant potential in strengthening fraud prevention strategies within the healthcare insurance sector. Future research can focus on incorporating explainable artificial intelligence techniques and real-time data processing systems to further enhance transparency, scalability, and adaptability in fraud detection models.

References

1. Chen, X., Li, Y., & Wang, J. (2025). Deep learning approaches for healthcare fraud detection using sequential claim data. *Journal of Healthcare Informatics Research*, 9(1), 45–60.
2. Farbmacher, H., Kögel, M., & Spindler, M. (2020). Machine learning methods for healthcare fraud detection. *Health Economics*, 29(3), 329–347.

3. Fursov, I., Zaytsev, A., & Burnaev, E. (2022). Deep learning for tabular data: Applications in insurance fraud detection. *IEEE Transactions on Neural Networks and Learning Systems*, 33(8), 3438–3450.
4. Hasan, M. (2022). Data mining techniques for healthcare fraud detection: A systematic review. *International Journal of Data Science and Analytics*, 14(2), 115–128.
5. Muhammad, S., Khan, R., & Ali, A. (2025). Hybrid machine learning frameworks for detecting fraudulent medical claims. *Expert Systems with Applications*, 235, 120109.
6. Preez, D., van der Merwe, A., & Botha, M. (2024). Artificial intelligence in insurance fraud detection: Opportunities and challenges. *Computers in Industry*, 155, 104001.
7. Ramachandran, S., Iyer, R., & Krishnan, P. (2024). Business intelligence integration for healthcare fraud analytics. *Decision Support Systems*, 174, 113967.
8. Rey-Puech, P., Boucher, A., & Lafond, D. (2025). Integrating heterogeneous data sources for fraud detection in insurance ecosystems. *Information Systems Frontiers*, 27(1), 89–104.
9. Shahzadi, S., Ahmad, M., & Rehman, Z. (2023). Feature engineering techniques for fraud detection in health insurance claims. *Applied Artificial Intelligence*, 37(4), 654–671.
10. Shungube, F., Nkosi, T., & Mhlongo, S. (2024). Artificial intelligence and deep learning for healthcare fraud detection: A systematic review. *IEEE Access*, 12, 54623–54640.