International Journal of Core Engineering & Management
Volume-8, Issue-01, 2025 ISSN No: 2348-9510

DEEP LEARNING IN DETECTING HEALTHCARE FRAUD: THE ECONOMIC BURDEN ON PUBLIC AND PRIVATE INSURERS

Kiran Veernapu
Salt Lake City, USA
kiran_veernapu@yahoo.com

Abstract

Healthcare fraud remains one of the most significant challenge in both public and private healthcare systems worldwide, driving up costs and undermining the effectiveness of care delivery. Fraudulent activities in healthcare, ranging from billing for services not rendered to misreporting diagnoses, result in billions of dollars in losses each year [1]. The traditional methods of detecting fraud, such as manual audits and rule-based systems, are often inefficient, error-prone, and resource-intensive. In contrast, deep learning, a subset of artificial intelligence (AI), has shown significant promise in automating and enhancing fraud detection systems. This journal examines the role of deep learning in identifying healthcare fraud and the economic burden it places on public and private insurers. It explores the technological advancements in deep learning models, their application in healthcare fraud detection, and the financial impact on the healthcare system, focusing on cost reduction, improved detection accuracy, and efficiency. Deep learning helps public and private insurers by detecting healthcare fraud, reducing operational costs, and enhancing resource allocation, and create more efficient and sustainable healthcare systems.

Keywords: Fraud detection, healthcare fraud, deep learning, AI, machine learning, data mining in healthcare, risk, detection systems, operational cost, billing code, claims, invoicing.

I. INTRODUCTION

Healthcare fraud is a pervasive issue in global healthcare systems, involving activities such as falsified billing, false diagnosis codes, upcoding, and kickbacks to physicians. These fraudulent activities significantly inflate healthcare costs, increasing the financial burden on both public insurers like Medicare and private insurers. According to the National Health Care Anti-Fraud Association (NHCAA), healthcare fraud costs the U.S. economy upwards of $68 billion annually, a substantial amount that diverts resources from legitimate patient care. Vulnerable populations are the most adversely impacted victims of healthcare fraud [2].

In recent years, advancements in artificial intelligence (AI) and machine learning (ML) have opened new avenues for detecting fraudulent claims. Among these, deep learning—a subset of machine learning that uses neural networks to analyze large datasets—has gained attention for

its ability to recognize complex patterns and improve detection accuracy [3]. Deep learning models have demonstrated their capacity to identify fraudulent behavior in healthcare claims, reducing the cost of manual audits and increasing the speed of fraud detection.
This journal explores how deep learning can improve healthcare fraud detection, focusing on the benefits and challenges of implementing these technologies in the public and private insurance sectors. It also discusses the economic implications of fraud detection and how deep learning can help reduce financial losses for insurers, ultimately leading to more sustainable healthcare systems.

II. HEALTHCARE FRAUD: A GROWING ECONOMIC BURDEN

Healthcare fraud exists in many sources, like dishonest providers, organized criminals, colluding patients, and patients who misrepresent their eligibility for insurance coverage. According to the literature survey conducted, there are 18 types of fraud types identified [4]. Some of the frauds include:

- **Billing for services not rendered:** Providers submit claims for services they did not provide to patients or billed for medical devices that were not provided to the patients. This concept is also called phantom billing. There are only patient-treated people who intend to fraud and send claims to several patients by creating fake treating employees in the organization. Sometimes the claims are submitted for the patient who expired [4].
- **Upcoding:** Upcoding is one of the most observed frauds. Healthcare providers submit higher billing codes than those for services actually rendered, by which a provider can obtain additional reimbursement[5].
- **Unnecessary services:** Providers perform tests or treatments that are not medically necessary but are intended to increase reimbursement. Falsified documents are created to prove that the patient has existing conditions to be treated [4].
- **Kickbacks:** Kickbacks aredownstream illegal forms of selling drugs, and one of the most reported frauds.Examples of kickbacks are pharmacists can fill a prescription of another brand instead of a prescribed brand as they may get some bonus. Physicians may write fraudulent prescriptions in collaboration with pharma companies. Providers offer or receive bribes for referrals of patients for unnecessary services or for using certain products [4].

The financial burden of healthcare fraud is staggering. The Centers for Medicare & Medicaid Services (CMS) estimates that healthcare fraud costs the U.S. healthcare system over $100 billion annually. In the private sector, insurers also suffer substantial financial losses due to fraudulent claims. These costs are ultimately passed on to consumers in the form of higher premiums and co-pays, reducing the affordability and accessibility of healthcare services.
Traditionally, insurers have relied on rule-based fraud detection methods, such as audits and manual claim reviews. While these methods are effective to some extent, they are often labor-intensive, time-consuming, and unable to detect novel forms of fraud. This has led to a

significant need for more advanced and efficient fraud detection systems, which deep learning is poised to address.

III. CURRENT FRAUD DETECTION SYSTEMS

To combat the current healthcare fraud, healthcare providers and insurers have implemented various fraud detection systems, which include traditional rule-based methods, data mining approaches, and more recently, machine learning and artificial intelligence (AI) systems.

3.1 Rule-Based Fraud Detection Systems

Rule-based systems were among the earliest tools used for fraud detection in healthcare. These systems rely on pre-set rules to identify suspicious claims [6]. For example, a rule might flag a claim if the billed procedure does not match the diagnosis or if a certain treatment is given at an unusually high frequency. While rule-based systems are easy to implement and understand, they suffer from several limitations, including:

- **Limited Scope:** Rule-based systems are designed to detect specific types of fraud, which means that they can miss more complex fraudulent behaviors.
- **High Rate of False Positives:** Because these systems depend on static rules, they may flag legitimate claims as fraudulent, leading to unnecessary investigations and wasted resources.
- **Inflexibility:** These systems cannot adapt to new or evolving fraud tactics without manual updates to the rules.

Despite these drawbacks, rule-based systems are still widely used due to their simplicity and cost-effectiveness in detecting certain types of fraud.

3.2 Data Mining and Statistical Analysis

Data mining approaches involve using algorithms to analyze large datasets for patterns and anomalies that may indicate fraud. These systems often employ statistical methods such as regression analysis, clustering, and decision trees to identify suspicious activities [7]. Data mining offers several advantages over rule-based systems:

- **Scalability:** Data mining can handle large amounts of healthcare claims data, making it suitable for organizations with vast amounts of data.
- **Improved Detection of Complex Fraud:** Unlike rule-based systems, data mining can identify more complex patterns and trends in fraudulent activities.
- **Lower False Positives:** By examining a broader range of data points, data mining can reduce the occurrence of false positives.

However, data mining also has its limitations:

- **Requires High-Quality Data:** The success of data mining depends on the quality of the data being analyzed. Incomplete or inaccurate data can lead to erroneous results.
- **Complexity:** These systems can be difficult to implement and require specialized

expertise to develop and maintain.

- **Static Nature:** Data mining algorithms are often reactive, meaning they can only identify patterns that have already occurred, rather than predicting future fraudulent behavior.

3.3 Machine Learning and Artificial Intelligence

Machine learning (ML) and AI represent cutting-edge technologies in fraud detection. These systems use advanced algorithms to "learn" from historical data, enabling them to detect new patterns and adapt to evolving fraud tactics. Key techniques in AI-driven fraud detection include:

- **Supervised Learning:** In supervised learning, algorithms are trained on labeled datasets containing both fraudulent and non-fraudulent claims. The model then learns to predict whether new claims are likely to be fraudulent [7].Supervised techniques (classification) cannot classify new types of disease claims whereas unsupervised techniques (clustering) cannot detect duplicate claims as fraud.
- **Unsupervised Learning:** Unsupervised learning algorithms can detect anomalies in healthcare data without relying on labeled datasets, which means it is possible to detect new fraudulent behaviors. These systems can identify unusual patterns in claims data, which may indicate fraud [8].
- **Deep Learning:** Deep learning, a subset of machine learning, uses neural networks with many layers to analyze complex data patterns. It has been particularly successful in detecting fraud in medical imaging, electronic health records, and large-scale claims data.

As there are advantages and disadvantages with both supervised and unsupervised learning models, the researcher proposed a novel model which is a hybrid of both supervised and unsupervised together. The Evolving Clustering Method (ECM) receives the dynamic data as input it can determine the new disease types using the clustering model. Support Vector Machine (SVM) is a supervised technique used in classification. The model is trained with pre-classified data like fraudulent and legitimate. Any new claim can be classified according to the trained data[7].

IV. THE ROLE OF DEEP LEARNING IN HEALTHCARE FRAUD DETECTION

Deep learning algorithms, particularly those involving artificial neural networks (ANNs), have revolutionized how large datasets are processed and analyzed. In the context of healthcare fraud detection, deep learning models can identify complex patterns and anomalies in healthcare claims data that traditional methods may miss. Key deep learning techniques used in fraud detection include:

4.1 Neural Networks

Artificial neural networks, inspired by the human brain, are designed to recognize complex

patterns and relationships within data. By analyzing vast amounts of claims data, neural networks can identify subtle indicators of fraud that are not immediately apparent. These networks "learn" from the data and improve over time, enabling them to detect evolving fraud tactics [9].

4.2 Convolutional Neural Networks (CNNs)

Although primarily used in image processing, CNNs have been adapted to detect patterns in structured data, such as healthcare claims. By analyzing time-series data or sequences of transactions, CNNs can identify fraudulent billing patterns that repeat over time, such as upcoding or double billing [9].

4.3 Recurrent Neural Networks (RNNs)

RNNs are particularly useful for analyzing sequential data, making them ideal for detecting fraud in claims that unfold over time. For instance, RNNs can track the sequence of claims submitted by a healthcare provider and detect irregularities in patterns that may suggest fraudulent activity, such as submitting multiple claims for the same service [10].

4.4 Autoencoders

Autoencoders are unsupervised learning algorithms that can be used to detect anomalies in healthcare claims data. By learning the normal patterns of claims submissions, autoencoders can flag outliers that deviate from the expected behavior, such as unusual billing practices or mismatched diagnosis codes [11].

4.5 Generative Adversarial Networks (GANs)

GANs are a newer type of deep learning model used for generating synthetic data and detecting fraudulent activity. GANs can be employed to simulate fraudulent claims data, which can then be used to train fraud detection models. This allows the detection system to become more robust and capable of identifying emerging fraud tactics [10].

V. ECONOMIC BENEFITS OF DEEP LEARNING IN FRAUD DETECTION

5.1 Reducing Operational Costs

One of the most significant economic benefits of using deep learning in fraud detection is the reduction in operational costs. Traditional fraud detection methods rely heavily on human labor to review claims, identify discrepancies, and audit billing practices. These methods are resource-intensive and prone to human error. Deep learning algorithms, on the other hand, can process vast amounts of data autonomously, significantly reducing the need for manual intervention. This leads to cost savings in terms of labor, time, and administrative expenses [12].
Example: By automating the detection of suspicious claims, deep learning models can reduce the time spent on audits by over 50%, resulting in significant savings for insurers.

5.2 Improving Detection Accuracy

Deep learning models excel in detecting patterns and anomalies that may be difficult for humans or traditional rule-based systems to identify. This results in more accurate fraud detection, reducing both false positives (legitimate claims flagged as fraud) and false negatives (fraudulent claims missed by the system). By improving detection accuracy, insurers can recover more funds lost to fraud and minimize the need for costly follow-up investigations.
Example: Deep learning models have been shown to reduce the rate of false positives by 20-30%, allowing insurers to focus resources on truly suspicious claims.

5.3 Reducing Fraud Losses

By detecting fraud more effectively, deep learning helps reduce the overall losses caused by fraudulent claims. Insurers can identify and act on fraudulent behavior more quickly, preventing further fraudulent activities and recovering funds that would otherwise be lost.
Example: A study by the National Health Care Anti-Fraud Association (NHCAA) found that AI-based fraud detection models can reduce fraud-related losses by up to 40%, translating to significant savings for both public and private insurers [13].

5.4 Enhancing Resource Allocation

AI-based fraud detection systems allow insurers to prioritize high-risk cases and allocate resources more effectively. Rather than spending time manually reviewing all claims, deep learning systems can flag the most suspicious claims, enabling auditors and investigators to focus on the most critical cases.
Example: By using deep learning to prioritize claims that exhibit a high likelihood of fraud, insurers can optimize their resources and reduce investigation costs.

VI. CHALLENGES AND LIMITATIONS OF DEEP LEARNING IN FRAUD DETECTION

While deep learning offers significant advantages, there are also several challenges to consider:

6.1 Data Quality and Availability

The effectiveness of deep learning models depends heavily on the quality and availability of data. Incomplete or inaccurate healthcare claims data can lead to poor model performance and increase the risk of undetected fraud. Ensuring high-quality data input is crucial for the success of AI-powered fraud detection systems.

6.2 Implementation Costs

The initial investment required to implement deep learning systems can be substantial. This includes costs related to technology infrastructure, data acquisition, model training, and staff training. For some insurers, especially smaller ones, these upfront costs may be prohibitive.

6.3 Model Transparency and Explainability

Deep learning models are often seen as "black boxes" due to their complexity, making it difficult

for insurers to understand how the model arrived at a particular decision. Lack of transparency can be a significant barrier to adoption, particularly in industries like healthcare where regulatory compliance and explainability are critical.

6.4 Ethical and Privacy Concerns

Healthcare fraud detection models must comply with data privacy laws, such as HIPAA in the U.S. Deep learning systems must be designed with appropriate safeguards to protect patient privacy and ensure that personal health information is not misused.

VII. CONCLUSION

Deep learning offers significant potential for detecting healthcare fraud, providing public and private insurers with a powerful tool to reduce the financial losses caused by fraudulent activities. By improving detection accuracy, reducing operational costs, and enhancing resource allocation, deep learning can help create more efficient and sustainable healthcare systems. However, challenges related to data quality, implementation costs, model transparency, and privacy concerns must be addressed to ensure the successful integration of AI-driven fraud detection systems.

As technology continues to evolve, deep learning is likely to play an increasingly central role in the fight against healthcare fraud, ultimately helping insurers reduce losses, improve patient care, and maintain the financial integrity of healthcare systems worldwide.

REFERENCES

1. Nicole Forbes Stowell, Carl Pacini, Nathan Wadlinger, Jaqueline M. Crain, and Martina Schmidt, Investigating Healthcare Fraud: Its Scope, Applicable Laws, and Regulations, 11 Wm. & Mary Bus. L. Rev. 479 (2020),https://scholarship.law.wm.edu/wmblr/vol11/iss2/5
2. Rosenbaum, S., Lopez, N., &Stifler, S. (2009). Health insurance fraud: An overview. Washington, D.C.: Department of Health Policy, School of Public Health and Health Services, The George Washington University.
3. John Maynard, Tom Wriggins, Robert Morison. (2022). Fight the rising tide of medicaid fraud. International institure for analytics. https://www.nhcaa.org/partner/articles-white-papers/.
4. Thornton, D., Brinkhuis, M., Amrit, C., & Aly, R. (2015). Categorizing and describing the types of fraud in healthcare. Procedia Computer Science, 64, 713-720.
5. Bauder, R., Khoshgoftaar, T. M., &Seliya, N. (2017). A survey on the state of healthcare upcoding fraud analysis and detection. Health Services and Outcomes Research Methodology, 17, 31-55.
6. Baumann, M. (2021). Improving a rule-based fraud detection system with classification based on association rule mining.

7. V. Rawte and G. Anuradha, "Fraud detection in health insurance using data mining techniques," 2015 International Conference on Communication, Information & Computing Technology (ICCICT), Mumbai, India, 2015, pp. 1-5, doi: 10.1109/ICCICT.2015.7045689.
8. Al-Hashedi, K. G., &Magalingam, P. (2021). Financial fraud detection applying data mining techniques: A comprehensive review from 2009 to 2019. Computer Science Review, 40, 100402.
9. S. Mardani and H. Moradi, "Using Graph Attention Networks in Healthcare Provider Fraud Detection," in IEEE Access, vol. 12, pp. 132786-132800, 2024, doi: 10.1109/ACCESS.2024.3425892.
10. R. A. Bauder and T. M. Khoshgoftaar, "Medicare Fraud Detection Using Machine Learning Methods," 2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA), Cancun, Mexico, 2017, pp. 858-865, doi: 10.1109/ICMLA.2017.00-48.
11. SamreenNaeem,AqibAli,SaniaAnam, Muhammad Munawar Ahmed (2023). An Unsupervised Machine Learning Algorithms:Comprehensive Review. http://dx.doi.org/10.12785/ijcds/130172
12. Boyer, M. & Léger, P.-T. (2005). The Impact of Health Care Cost Increases on Fraud and Economic Waste. Assurances et gestion des risques / Insurance and Risk Management, 73(1), 5–29. https://doi.org/10.7202/1107004ar
13. NHCCA, National Health Care Anti-Fraud Association. (2024). https://www.nhcaa.org/about-nhcaa/