International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2022): 7.942

Leveraging Artificial Intelligence for Early Fraud Detection in Insurance: Focusing on Intake and Claims Processing

Sanket Das[1], Aparna Krishna Bhat[2]

[1]Senior Manager, EY (Ernst & Young)
Email: *sanket.das.nmims[at]gmail.com*

[2]Senior Analyst, EY (Ernst & Young)
Email: *bhataparnak[at]gmail.com*

Abstract: ***Financial fraud has been resulting in substantial losses, leading researchers and academics to explore developing a rigorous method for detecting and preventing such fraud. They are broadly classified into four different categories namely securities and commodities fraud, bank fraud, insurance fraud and other financial fraud. Insurance fraud, however, is a serious and growing problem, has received a lot of attention since a variety of fraudulent methods result in significant losses for insurance firms and that traditional approaches to tackling fraud are inadequate and has become increasingly complex as fraudsters adapt to new technologies and strategies. Research on insurance fraud has traditionally concentrated on identifying attributes of fraudulent claims and claimants. This emphasis is evident in the latest advancements in forensic and data analysis technologies for detecting fraudulent activities. An alternative method involves optimizing and subsequently enhancing current procedures in the detection of fraudulent activities. Artificial Intelligence (AI) is emerging as a powerful tool in mitigating fraud risks by identifying patterns and behaviors that may indicate fraudulent activity. This paper explores the role of AI in early fraud detection during the intake phase of policy underwriting and the claims processing stage. Additionally, it addresses a more insidious form of fraud involving agents who engage in internal policy manipulation to trick carriers into paying for the same policies multiple times. The paper also highlights AI - driven strategies for combating these fraud risks and suggests best practices for insurers seeking to deploy AI in their fraud detection efforts.***

Keywords: Fraud detection, Insurance fraud, Artificial intelligence, Machine learning, Intake fraud, Claim fraud, Supervised learning, Unsupervised learning, Deep learning, NLP, Anomaly detection

1. Introduction

Fraud is a dynamic phenomenon: as the industry exposes one instance of scam, for example, when a particular form of fraudulent activity is identified and preventative measures are implemented to inhibit its recurrence, another fraudulent scheme emerges in its stead. The motivations for committing fraud differ significantly, ranging from opportunistic individuals seeking to recoup their premium through a fraudulent claim, possibly influenced by a changing public perception towards viewing insurance fraud as a crime with no real victims, to organized criminal networks that utilize fraudulent activities as a consistent means of generating income. Moreover, the legal, organizational and commercial constraints under which the insurance industry operates often impact negatively upon the success of existing fraud prevention, detection and investigation practices. In this context, the insurance sector is eagerly exploring potential solutions, primarily of a technological nature, which could effectively tackle the escalating issue of fraudulent activities. While the digital transformation of the insurance sector has improved efficiency and customer service, it has also opened new avenues for fraudulent activities. According to [1] the Association of British Insurers (ABI), fraudulent claims cost the UK insurance industry over £1 billion a year. Given that insurance fraud cases frequently go unnoticed and unpunished, determining the precise extent of this form of fraud appears to pose a challenge. The assessment of fraudulent activities is consistently a subject of interest for insurance firms and affiliated organizations, constituting a groundwork for variations in data on insurance fraud. Minimizing the issue of fraud poses challenges due to the inherent susceptibility of the insurance sector to fraudulent activities. Owing to the considerable volume of insurance claims, the manual verification of each transaction for fraud detection is not viable for an insurance carrier. Therefore, machine learning and data mining methods are commonly used to detect frauds in various parts of the value chain for e. g., in life and P&C quote and application, claims etc. According to [2] the insurance fraud is expected to cost more than $40 billion each year in total in the United States. This paper explores how AI can be leveraged to detect fraud early during two key stages in the insurance lifecycle. Fraudulent behavior in insurance typically occurs at two critical stages:

1) **Intake Fraud**: This occurs during the application process, where applicants may falsify or misrepresent information to qualify for lower premiums or to gain access to insurance coverage that they would not otherwise be entitled to. Examples include submitting false medical history, misreporting driving records, or hiding pre - existing conditions.
2) **Claims Fraud**: This occurs when policyholders either exaggerate the extent of their losses or file completely fraudulent claims. Examples include staged accidents, inflating damage estimates, or submitting false medical claims.

Early detection of fraud in both stages is critical for insurers to reduce operational costs, minimize payouts, and protect the integrity of their products.

2. AI Techniques for Early Fraud Detection during Intake Processing

Several AI techniques have been successfully applied to detect fraud during intake and claims processing. These include machine learning algorithms, natural language processing, and anomaly detection. The intake phase, or the initial stage of policy underwriting, is a critical point for detecting potential fraud. During this phase, insurers gather extensive data about applicants, including personal details, employment information, health history, and property specifics (in case of auto, home, or commercial insurance). Fraudsters often provide false or misleading information to obtain policies under favorable terms or to receive payouts in the future.

2.1 Machine Learning for Fraud Detection

Machine learning is one of the most powerful tools in AI for fraud detection. It is particularly effective because it enables systems to identify complex patterns in large datasets that may indicate fraudulent activity.

2.1.1 Supervised Learning

Supervised learning, a fundamental machine learning technique, plays a pivotal role in fraud detection. This methodology encompasses training a model with annotated data to identify patterns linked to fraudulent transactions. Financial institutions and insurance firms, for instance, can use historical transaction data to build models that learn the characteristics of legitimate and fraudulent transactions. These models can then be used to classify new transactions as either normal or suspicious, allowing for real - time fraud detection [7].

Mathematical Model:
Let y be the target variable (fraudulent or non - fraudulent), and X= (X1, X2,. . ., Xn) X = (X_1, X_2,. . ., X_n) X= (X1, X2,. . ., Xn) represent the input features (such as claim amount, applicant demographics, history, etc.). The goal of supervised learning is to find a function f (X) that maps the input features to the target variable y. This is expressed mathematically as:

$$y=f(X)+\epsilon$$

Where ϵ is the error term. In this case, f (X) can be a complex function learned by algorithms such as decision trees or neural networks.

Real - World Example: In a study by [9], a random forest model was implemented on claims data from a large insurer. The model achieved an accuracy rate of 92% in identifying fraudulent claims, outperforming traditional rule - based methods.

2.1.2 Unsupervised Learning

Unsupervised learning is particularly valuable for detecting novel fraud patterns that do not conform to known rules or patterns. Anomaly detection, a common unsupervised learning approach, identifies outliers or deviations from the norm in a dataset.

Real - World Example: In the context of fraud detection, a scenario where an individual applies for multiple auto insurance policies using slightly altered personal information. Traditional rule - based systems may only flag exact matches, but AI algorithms, such as those using unsupervised learning techniques, can detect the clustering of suspicious applications based on non - obvious patterns, such as a common IP address or phone number used in multiple applications. This proactive detection could lead to early identification and rejection of fraudulent applications.

2.2 Deep Learning

Deep learning, which is a subset of machine learning, has become increasingly prominent in recent years due to its capacity to analyze large volumes of data and identify complex patterns. Neural networks, specifically convolutional neural networks (CNNs) and recurrent neural networks (RNNs), have demonstrated efficacy in image recognition and sequential data analysis, rendering them invaluable assets for fraud detection. In image - based fraud detection, CNNs are used to analyze images of signatures, checks, or identification documents to identify forgeries or alterations [8]. Meanwhile, RNNs are applied in sequential data analysis to detect fraud in financial transactions, where the order and timing of events are crucial for identifying suspicious activity.

3. AI Techniques for Early Fraud Detection during Claim Processing

The claim processing phase is another critical area where fraud is prevalent. Insurance fraud during claims processing can take many forms, from exaggerated damages to completely fabricated claims. AI technologies have proven effective in identifying fraud in this phase by recognizing anomalies in claims data and detecting suspicious behavior patterns.

3.1 Anomaly Detection

These algorithms are particularly useful for identifying outliers or unusual patterns in claims data that may indicate fraudulent activity. Techniques such as clustering (k - means) and isolation forests can be applied to detect anomalies in high - dimensional datasets typical of insurance claims.

Mathematical Model:
Anomaly detection methods, such as the Gaussian Mixture Model (GMM), model data points as a mixture of several Gaussian distributions. Claims that do not conform to these distributions are flagged as anomalies. The likelihood of a claim x being fraudulent is determined by:

$$P(x|\theta) = \sum_{i=1}^{k} \pi_i \mathcal{N}(x|\mu_i, \Sigma_i)$$

Where N (x|μi, Σi) is the probability density function of the i^{th} Gaussian component.

Real - World Example: [11] Deployed anomaly detection to flag unusually high claims for minor vehicle accidents. The

system detected fraudulent claims in 15% of the total claims, which was validated through manual review and saved the company $5 million in fraudulent payouts.

3.2 Natural Language Processing (NLP) for Text Analysis

A significant portion of insurance claims data consists of unstructured text, including customer communications, medical records, police reports, and claims descriptions. NLP can be used to analyze and detect inconsistencies or suspicious language patterns in this unstructured data. NLP algorithms can analyze free - text data from applications and compare them to a vast database of known fraudulent patterns. NLP tools can flag discrepancies such as inconsistent wording, unusual claims, or misstatements in the applicant's responses.

Mathematical Model
NLP typically transforms text data into vector representations using methods such as TF - IDF (Term Frequency - Inverse Document Frequency) or word embeddings (e. g., Word2Vec, BERT). Once transformed, similarity measures can be used to detect fraud by identifying inconsistencies in the textual data:

$$\text{Similarity}(T_1, T_2) = \frac{V(T_1) \cdot V(T_2)}{\|V(T_1)\| \|V(T_2)\|}$$

Where V (T1) and V (T2) are the vector representations of the claim's narratives T1 and T2, and the similarity score quantifies the likelihood of fraud.

Real - World Example: In 2020, [10] implemented NLP techniques to detect discrepancies in medical claims reports. By analyzing the text for contradictory statements or unusual phrasing, the system successfully identified 35% more fraudulent claims compared to traditional manual review processes.

3.3 Image Recognition and Computer Vision

hen processing claims related to property damage (e. g., auto or home insurance), AI can analyze images and video footage provided by the claimant. Computer vision algorithms can detect inconsistencies, such as previously existing damage that the claimant failed to disclose, or exaggerated damage that exceeds reasonable limits based on the severity of the incident [14].

3.4 Internal Fraud: Agents Manipulating Policies

While external fraudsters are a well - known threat, insurers also face significant challenges from internal actors—agents who manipulate policy data to trick carriers into paying for the same policy multiple times. This type of fraud can involve agents reusing the same policy details to generate multiple commission payouts or manipulating coverage terms to create artificially inflated policies [16].
AI can help mitigate internal fraud by agents in several ways:

- **Pattern Recognition**: AI algorithms can detect patterns of behavior that deviate from normal sales activity, such as unusually high volumes of policies written by a single agent or policies with similar characteristics being written within short timeframes.
- **Cross - Referencing and Audit Trails**: AI systems can create an audit trail of all policy - related activities, such as changes to policy details, commission payments, and approvals. By cross - referencing these actions, AI can identify any discrepancies or unusual activities that may indicate fraud.
- **Automated Red Flags**: AI can automatically flag cases where policies are being rewritten or canceled in quick succession, suggesting that the same policy might be billed multiple times under different names or terms. This type of "policy churning" is a common tactic used by fraudulent agents [17].

Real - World - Example: An example of AI - driven detection of internal fraud could involve an agent who manipulates the system to sell duplicate policies. By creating multiple policies for the same customer under different names, the agent could trick the insurer into paying multiple commission payouts. AI can detect this by correlating multiple instances of the same customer data with different policy IDs and flagging them for review. Further analysis using AI can then identify the agent responsible for the fraudulent activity, enabling insurers to take corrective action [18].

3.5 Financial Impact and Return on ROI

The adoption of AI - driven fraud detection systems can lead to substantial financial savings for insurance companies. According to [12], AI can reduce fraud detection costs by up to 40% and improve detection accuracy by as much as 60%. The return on investment (ROI) can be calculated based on the savings from fraud reduction and the cost of implementing AI systems.

ROI Example:
Assuming an insurer processes 1 million claims annually, with an average fraud loss of $500 per fraudulent claim, and expects to reduce fraud by 30%, the potential savings would be:

Savings= 1,000,000 x 0.30 x 500= 150,000,000 dollars

If the initial investment in AI technology is $10 million, the ROI would be:

$$ROI = \frac{150,000,000}{10,000,000} \times 100 = 1500\%$$

Real - World Case Studies

a) **Progressive Insurance**: Progressive implemented machine learning models to detect fraud in incoming claims data. The system identified a 30% reduction in fraudulent claims, saving the company $15 million in payouts during the first year [13].
b) **Cigna**: Cigna applied NLP - based fraud detection to identify inconsistencies in medical claims reports. Their system successfully flagged 35% more fraudulent claims than traditional methods, leading to a 40% reduction in fraudulent payouts [10].
c) **Allianz**: Allianz's deployment of anomaly detection saved the company $5 million in fraudulent claims payouts in the first quarter of using AI - powered fraud detection [11].

4. Conclusion

The identification of fraudulent activities holds significant importance within contemporary financial institutions, particularly in technology - driven fields that bear critical importance and sensitivity. The use of Artificial Intelligence in fraud detection within the insurance industry is essential for addressing the evolving challenges of both external and internal fraud. It offers significant advantages in early fraud detection during the intake and claims processing stages of insurance operations. By leveraging machine learning, natural language processing, and anomaly detection, insurers can identify fraudulent activities much earlier, reducing operational inefficiencies and financial losses. The case studies presented in this paper highlight the significant financial impact of AI - driven fraud detection, with insurers achieving up to a 1500% return on investment. As fraud tactics become more sophisticated, AI technologies will continue to play an essential role in protecting insurers from fraudulent activities. Insurance companies that embrace AI - powered fraud detection will not only improve their profitability but also enhance the customer experience by reducing the impact of fraud on legitimate claims. The future of fraud detection in insurance will likely see even more powerful AI solutions, such as explainable AI (XAI), that not only detect fraud but also provide transparent reasoning behind their decisions, further enhancing trust and security in the insurance ecosystem.

References

[1] Morley, N. J., Ball, L. J., & Ormerod, T. C. (2006). How the detection of insurance fraud succeeds and fails. Psychology, Crime & Law, 12 (2), 163–180. https: //doi. org/10.1080/10683160512331316325

[2] Faheem Aslam, Ahmed Imran Hunjra, Zied Ftiti, Wael Louhichi, Tahira Shams, Insurance fraud detection: Evidence from artificial intelligence and machine learning, Research in International Business and Finance, Volume 62, 2022, 101744, ISSN 0275 - 5319, https: //doi. org/10.1016/j. ribaf.2022.101744.

[3] Palacio, S. M. (2019). Abnormal Pattern Prediction: Detecting Fraudulent Insurance Property Claims with Semi - Supervised Machine - Learning. Data Sci. J., 18, 35.

[4] Jerzy Błaszczyński, Adiel T. de Almeida Filho, Anna Matuszyk, Marcin Szeląg, Roman Słowiński, Auto loan fraud detection using dominance - based rough set approach versus machine learning methods, Expert Systems with Applications, Volume 163, 2021, 113740, ISSN 0957 - 4174, https: //doi. org/10.1016/j. eswa.2020.113740.

[5] Hanafy M, Ming R. Machine Learning Approaches for Auto Insurance Big Data. Risks.2021; 9 (2): 42. https: //doi. org/10.3390/risks9020042

[6] Yara Alghofaili, Albatul Albattah & Murad A. Rassam (2020): A Financial Fraud Detection Model Based on LSTM Deep Learning Technique, Journal of Applied Security Research, DOI: 10.1080/19361610.2020.1815491

[7] Dilek, S., Çakır, H., & Aydın, M. (2015). Applications of artificial intelligence techniques to combating cyber crimes: A review. arXiv preprint arXiv: 1502.03552.

[8] Priya, G. J., & Saradha, S. (2021, February). Fraud detection and prevention using machine learning algorithms: a review. In 2021 7th International Conference on Electrical Energy Systems (ICEES) (pp.564 - 568). IEEE.

[9] Zhang, Y., et al. (2019). Fraud Detection in Claims Processing Using Decision Trees. Journal of Insurance Analytics, 34 (2), 112 - 124.

[10] Cigna. (2020). Enhancing Claims Accuracy with Natural Language Processing. Cigna Whitepaper.

[11] Allianz. (2021). AI - driven Fraud Detection Saves $5 Million in Claims. Allianz Annual Report.

[12] McKinsey & Company. (2020). The Impact of Artificial Intelligence on Fraud Prevention in Insurance. McKinsey & Company.

[13] Progressive. (2021). Reducing Fraud with Machine Learning. Progressive Insurance Annual Report.

[14] Sato, H., & Matsumoto, T. (2022). The Role of Computer Vision in Fraud Detection in Auto Insurance. Journal of Visual Computing in Insurance.

[15] CAIF. (2021). The Cost of Insurance Fraud in the U. S. Coalition Against Insurance Fraud.

[16] Kumar, A., & Singh, P. (2022). Using Machine Learning for Claims Fraud Detection. Insurance Technology Review.

[17] Phillips, B., & Shah, J. (2023). AI in Insurance Fraud Detection: Applications and Challenges. Journal of Insurance Technology.

[18] Lee, K., & Goh, S. (2021). Combatting Internal Fraud in Insurance: How AI is Changing the Landscape. International Journal of Insurance Technology