



Medical Insurance Fraud Detection Using Machine Learning

P.P. Shenoy^{1*}, P. K. Vidhate², S. C. Gund³

^{1*}Assistant Professor, Department Of Information Technology, Changu Kana Thakur Arts, Commerce and Science College, New Panvel, Maharashtra-410206 Email:- shenoypriya89@gmail.com

²Department Of Information Technology, Changu Kana Thakur Arts, Commerce and Science College, New Panvel, Maharashtra-410206 Email:- pratikshavidhate0112@gmail.com

³Department Of Information Technology, Changu Kana Thakur Arts, Commerce and Science College, New Panvel, Maharashtra-410206 Email:- shwetagund18@gmail.com

***Corresponding Author:** P.P. Shenoy

*Assistant Professor, Department Of Information Technology, Changu Kana Thakur Arts, Commerce and Science College, New Panvel, Maharashtra-410206 Email:- shenoypriya89@gmail.com

Abstract

Medical insurance fraud poses significant challenges to the healthcare industry, impacting financial resources and patient care. This research explores the application of machine learning methodologies to detect fraudulent activities within healthcare insurance claims. Medical insurance fraud detection is crucial to help insurance companies save money. Machine learning is a powerful tool that can be used to detect fraudulent activities in the healthcare industry. Fraud can be spread broadly and extremely costly to the therapeutic protection framework. Protection can be made unscrupulous and be a case designed to hide or alter such information meant for social insurance benefits. Cheats might be numerous and submitted by the protection guarantor or the safeguarded. The unscrupulous social insurance providers are the reason for extortion in the well-being segment.

This research approach is to apply machine learning to find incidents of medical insurance fraud automatically. In conclusion, machine learning is a promising tool for detecting medical insurance fraud. It can help insurers detect fraudulent activities in real time, saving money on bogus claims.

CC License
CC-BY-NC-SA 4.0

Keywords: machine learning, fraud detection, healthcare, insurance.

1. INTRODUCTION

Medical insurance fraud detection is identifying and preventing fraudulent claims or activities in the health insurance sector. Fraudulent claims can result from intentional deception, misrepresentation, or manipulation of facts by healthcare providers, insurance subscribers, or other parties involved in the insurance process. Fraudulent activities can lead to increased healthcare costs, reduced quality of care, and wasted financial resources.

Computer science can be categorized into machine learning, a form of artificial intelligence that teaches computers how to make inferences or decisions based on collected data without specifically instructing them. Machine learning can be used to detect medical insurance fraud by analyzing large amounts of insurance claims data and identifying patterns, anomalies, or outliers that indicate possible fraud. Machine learning can also

help to automate the fraud detection process, reduce human errors, and improve the accuracy and efficiency of fraud prevention.

This research endeavors to explore, analyze, and advance methodologies in medical insurance fraud detection through the lens of machine learning. By delving into the diverse spectrum of machine learning algorithms, feature engineering approaches, and ethical considerations, this study seeks to unravel the intricacies of detecting fraudulent behavior in healthcare insurance claims. The aim is not only to enhance accuracy and efficiency but also to ensure ethical compliance, transparency, and fairness in the detection process.

2. OBJECTIVE

- To search for irregular patterns or outliers in medical billing that could suggest fraud, misuse, and abuse.
- To save resources and time by correctly differentiating between genuine and fraudulent claims.
- To automate the process of detection and use technology to detect fraud before it occurs in the organization, and in time.
- to use previous experiences and sophisticated algorithms to keep on updating on various fraud tactics and improve the accuracy of detecting fraud by studying historical data.
- To protect insurance companies from financial losses caused by fraudulent claims that lead to the imposition of fair premiums for all parties involved.
- To Healthcare insurance-Enforcement of regulatory standards and ethical practices in the healthcare industry.

3. LIMITATIONS OF THE STUDY

- The level of success of machine learning models significantly depends on the quality and completeness of the data. Erroneous prediction may occur due to inaccurate, incomplete, or biased datasets that can hamper the performance of the model.
- Models are prone to bias because of the imbalance between genuine and fraudulent claims in datasets, which makes it difficult to accurately detect fraudsters.
- Identifying appropriate features (medical procedures, billing codes, patient data) for fraud detection is also challenging. If there is ineffective feature selection or design, it could affect the model's ability to distinguish between genuine and fake claims.
- The models developed in one dataset or region may not generalize well to other insurance systems, regions, healthcare practices., This could, however, limit the model's generalizability to different settings.
- Many of the machine learning models, particularly complex ones like neural networks, are not interpretable. It is often difficult to understand why a model predicts something specific thereby eroding confidence and acceptance of its use in practice.
- However, fraudsters are known for changing their tactics. Without periodic retraining and updating, machine learning models may find it hard to keep up with new and changing fraudulent behaviors.
- Accessible and effective fraud detection may be limited by compliance with regulations and ethics such as HIPAA in the US for the privacy of patients and data protection laws.
- It requires a lot of resources for the institution to develop and maintain effective machine-learning models for fraud detection.

4. LITERATURE REVIEW

Detecting medical insurance fraud using machine learning techniques has become an active area of research due to its potential to identify anomalies and irregularities in healthcare claims.

The research "Fraud detection in health insurance claims using machine learning algorithms" by D. Vineela and P. Swathi (2020) explores various machine learning algorithms such as supervised and unsupervised algorithms for fraud detection analysis and prediction. Health insurance fraud claims data anomalies have detected fraud in healthcare.[1]

"A comprehensive study of healthcare fraud detection based on Machine Learning" by Shivani S. Waghade and Prof. Aarti M. Karandikar (2018) provides a comprehensive study of fraud detection methods in healthcare. This paper explores the improvements in data mining and machine learning tools for fraud detection in healthcare. Detection of anomaly and detection of fraud behavioral profile methods based on machine learning techniques. [2]

“Fraud detection in health insurance using Ensemble learning methods” by Rimante Kunickaite, Monika Zdanaviciute, and Tomas Krila (2020) employs ensemble learning methods in healthcare. It processes a decision-tree-based model and ensemble learning methods such as bagging, random forest, and boosting for healthcare fraud detection. It combines multiple classifiers to improve detection accuracy. [3]

The research “Fraud Detection in Medical Insurance Claim System using Machine Learning: A Review” by Paresh Gohil, Dr. Sheshang Degadwala, Dhairya Vyas (2022) presents a framework for the selection of features to be used in machine learning, which will enable the robust categorization of insurance claims. It will demonstrate how these technologies might be used in the development of a system that can prevent certain kinds of fraud in the field of healthcare. [4]

“Fraud Claims Detection in Insurance Using Machine Learning” by Hritik Kalra, Ranvir Singh, Dr.T. Senthil Kumar (2022) explores a hybrid approach based on the advantages of both supervised and unsupervised machine learning techniques. Evolving clustering method is used to cluster insurance claims according to diseases and a support vector machine is used for the classification of duplicate claims. [5]

The purpose of “Fraudulent Insurance Claims Detection Using Machine Learning” by Arif Ismail Alrais (2022) is to flag suspicious fraudulent claims for insurance companies or hospitals using machine learning and data analysis techniques. It is a master's research project paper. And also, this research highlighted the applications, and their usage and explained that machine learning algorithms are used for analyzing healthcare data analytics. [6]

The research paper "Medical Fraud and Abuse Detection System Based on Machine Learning" by Conghai Zhang, Xinyao Xiao, and Chao Wu (2020) is about how to use machine learning to find out if there is any cheating or misuse of money in the healthcare system. The authors say that sometimes doctors or patients may do things that are not honest, such as giving or taking too many drugs, or pretending to be sick, to get more money from the health insurance. [7]

5. METHODOLOGY

5.1 Data collection method:

In this paper, we have used an online survey-based methodology.

5.2 Machine Learning Methodology:

Machine learning can be used to automate and improve the fraud detection process by reducing manual efforts and increasing accuracy and efficiency. Different types of machine learning methods can be used for fraud detection, such as supervised, unsupervised, and semi-supervised learning. Supervised learning is when the algorithm is trained on labeled data, i.e., data that has the correct output or class for each input. Unsupervised learning is when the algorithm is applied to unlabelled data, i.e., data that does not have any output or class information.

For survey-based methodology, supervised learning is the most common and widely used approach, as it can leverage the existing knowledge and labels of fraudulent and genuine claims to train and test the algorithm. Logistic regression works by estimating the probability of an outcome based on the input variables, such as the features of a health insurance claim. It then uses a threshold to classify the outcome as either 0 (false) or 1 (true). To detect medical insurance fraud by training the algorithm on labeled historical data, where the outcome of each claim is known (fraud or genuine). The algorithm can then learn the relationship between the input variables and the outcome, and use it to predict the outcome of new claims.

However, logistic regression has some limitations, such as:

- It assumes that the linear relationship between the input variables and their outcome, may not always hold in real-world data.
- It may suffer from class imbalance, where the number of fraudulent claims is much lower than the number of genuine claims, which can affect the accuracy and performance of the algorithm. Some of the popular supervised learning algorithms for fraud detection are:

1. Logistic regression: A linear model that predicts the probability of a binary outcome (fraud or genuine) based on the input features.

2. Decision tree: A tree-like structure that splits the data into smaller subsets based on certain criteria or rules, and assigns a class label to each leaf node.

3. Random forest: An ensemble method that combines multiple decision trees and aggregates their predictions by voting or averaging.

4. Artificial neural network: A network of interconnected nodes that mimics the biological neurons and learns complex nonlinear relationships between the input and output.

These algorithms can be applied to various features or attributes of health insurance claims, such as the claim amount, diagnosis code, provider type, policy type, age, gender, etc. The algorithms can also handle different types of data, such as numerical, categorical, textual, or temporal data. The performance and reliability of the algorithms can be evaluated using various metrics, such as accuracy, precision, recall, F1-score, specificity, and area under the curve (AUC).

5.3 Sample design:

For our study, we collected data from people using a series of questionnaires. Following is the sample questionnaire.

1. Do you currently have medical insurance?
2. How long have you been using medical insurance services?
3. What common issue do people face in insurance fraud?
4. Do you think insurance fraud is a common issue?
5. How you ever suspected fraudulent activities related to your medical insurance claims?
6. Have you or someone you know ever been a victim of medical insurance fraud?
7. How easy do you find the process of filing insurance claims?
8. If you suspect a fraudulent medical insurance claim, would you know how to report it to the insurance company?
9. How important do you think it is to detect and prevent fraud in insurance?

6. FINDING

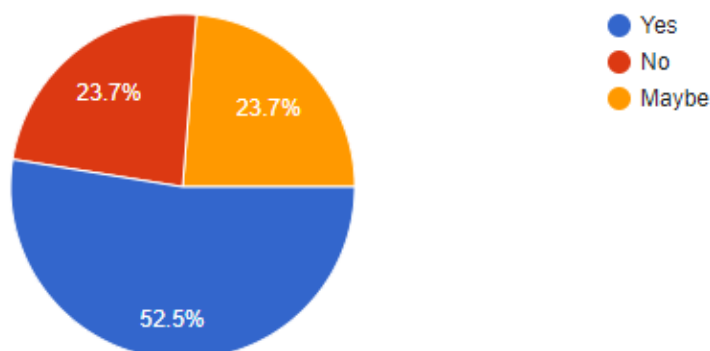
The sample size is 60. Data is collected by an online survey of 60 people from different age groups.

Of the total study, 40% were females and 60% were males.

46.7% of the people were students, 43.3% of the people from services, 3.3% of the people from business, 6.7% of the people from housewife. 63.3% of the people do not have medical insurance and 36.7% have medical insurance.

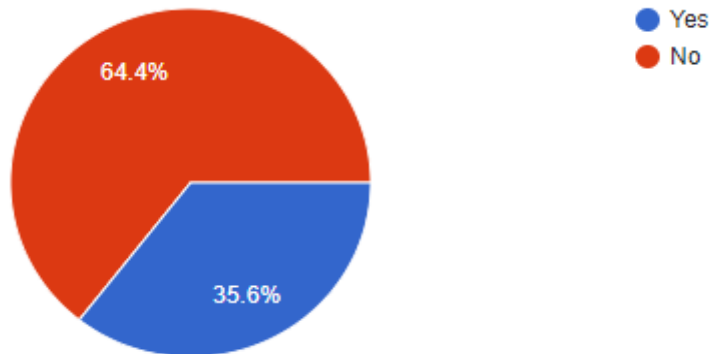
6.1 Insurance fraud is a common issue

- Premium diversion refers to stealing insurance premiums. It is the most popular form of insurance fraud. A regular insurance agent does not pass on premiums to the underwriter but rather holds them for personal gains.
- One of the popular fraudulent practices is the sale of insurance without a license, taking collection of premiums, and failing to settle claims.
- Victims of Fraud also suffer mental and physical traumas that may last a lifetime.
- 52.5% of people think that insurance fraud is the most common issue nowadays.



6.2 Suspected fraudulent activities related to your medical insurance claims

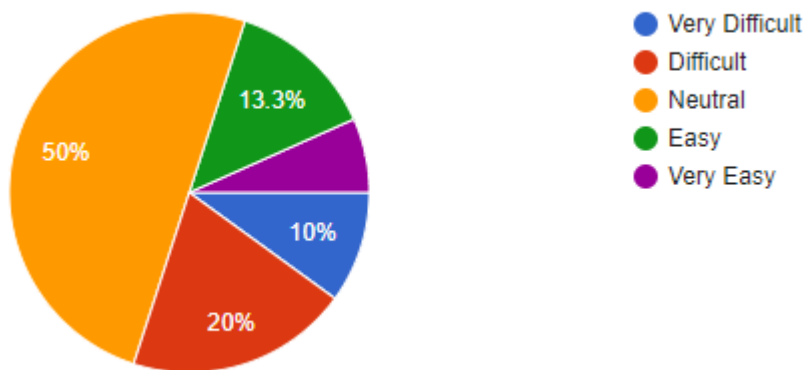
Dishonest insurance companies and insurance agents can defraud consumers by collecting premiums on non-existent policies that they do not intend or have the ability to pay. Such “companies” can sell policies at such prices below the usual market cost, aimed to draw shoppers eager to cut expenses. Oftentimes, consumers will be supplied with documents that seem genuine from an illegitimate insurance firm. These policies may even be personified by reputable insurance agents who have also been conned by unscrupulous companies.



6.3 The process of filing insurance claims

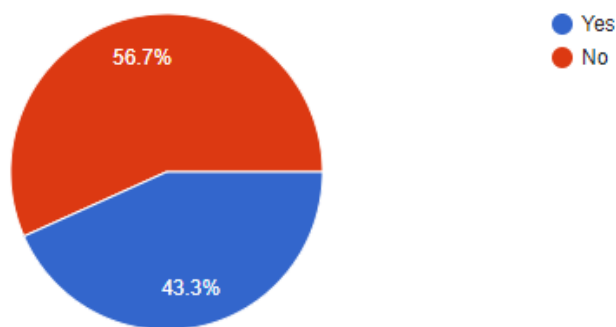
It is advisable to notify the insurer early enough when making a health insurance claim. One should ensure that he or she files a claim within the time stated in the policy. They guide filling out the claim form, the documents you need to submit, and your representatives who will walk you through this process with the insurer’s representatives. Failure to advise the insurer in time can lead to the rejection of a health insurance claim.

50% of the people think that filing insurance claims is a neutral process, 30% of the people think that it is a difficult process and 20% of the people think that it is an easy process



6.4 If you suspect a fraudulent medical insurance claim, would you know how to report it to the insurance company?

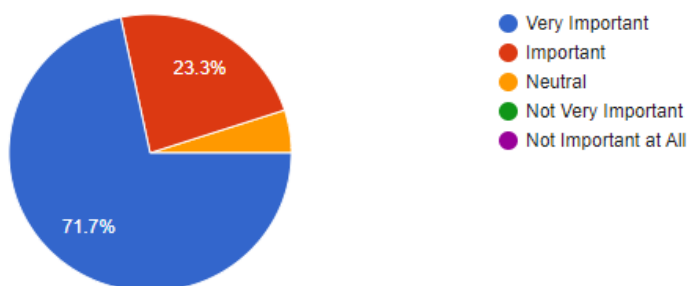
- Contact your insurance agent.
- Address the letter to an executive of the insurance company.
- Seek assistance by consulting any other person like an ombudsman in the settlement of your case.
- Submit a grievance to your state's Department of Insurance, as this is responsible for monitoring insurer activity and ensuring that they comply with state legislation.
- If seeking arbitration is an option in your policy, please proceed this way.
- Make a legal complaint against the insurer in court.
- 56.7% of the people do not have any idea about how to report it to the insurance company. And only 43.3% of people know how to report it to the insurance company.



6.5 How important do you think it is to detect and prevent fraud in insurance?

Insurance fraud detection and prevention is critical as it conserves funds and time, prevents loss, and settles disputes. Insurance fraud refers to falsified or exaggerated claims that annually cost billions to insurers. To counter this problem, insurers deploy AI, machine learning, and prediction model systems that are capable of going through the data and identifying fraud in real-time. Additionally, such technologies can be used to check the authenticity of the client's identity and improve a more friendly method of solving disputes between two parties.

71.7% of the people think that it is very important to detect and prevent fraud in medical insurance, 23.3% of the people think that it is important to detect and prevent fraud in insurance and only 5% of the people think that it is neutral, so the most of the people think that detect fraud in medical insurance is very important.



7 FUTURE SCOPE

The development of more interpretable machine learning models will be important. The use of explainable AI techniques can help understand why these models make certain decisions as well as ensure that they are compliant with regulations and ethics. The evolution and development of systems that can detect these frauds at the time they occur or immediately after will be due to advances in technology. Such adaptive systems capable of autonomous self-learning with immediate response in real-time will be a critical concern.

Development of cheap, low-cost fraud detection schemes that even small healthcare providers and insurance companies could apply with ease. Formulating a mixed approach with ambiguity-free man-machine integration for complex fraud detection.

8 CONCLUSION

Medical insurance fraud detection using machine learning is the process of using algorithms and data to identify and prevent fraudulent claims or behaviors in the healthcare sector. Machine learning can help detect patterns, anomalies, and relationships that are not easily visible to human analysts or rule-based systems. Machine learning can also handle large and complex datasets, and adapt to the ever-changing nature of fraud. The conclusion of using machine learning for fraud detection is that it can improve the accuracy, efficiency, and cost-effectiveness of identifying and preventing fraud in healthcare insurance claims. However, there are also some challenges and limitations, such as data quality, class imbalance, and ethical issues, that need to be addressed before machine learning can be fully implemented in this domain.

By leveraging advanced algorithms, such as anomaly detection and predictive modeling, insurers can enhance their ability to identify suspicious patterns and potential fraud cases. It's crucial to acknowledge the ongoing challenges, including the need for high-quality labeled data and continuous model refinement to adapt to evolving fraud schemes. Striking a balance between accuracy and interpretability is essential, ensuring that the implemented system is both effective and transparent. Overall, the integration of machine learning holds significant potential in mitigating medical insurance fraud, but a comprehensive and collaborative effort is required for sustained success in this complex domain.

9 REFERENCES

1. D. Vineela, P. Swathi, T. Sritha, K. Ashesh. Fraud Detection in Health Insurance Claims using Machine Learning Algorithms. International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8 Issue-5, January 2020 E6485018520.pdf (ijrte.org)
2. Shivani S. Waghade and Prof. Aarti M. Karandikar (2018). A Comprehensive Study of Healthcare Fraud Detection based on Machine Learning. International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 6 (2018) pp. 4175-4178 © Research India Publications. <http://www.ripublication.com>.
3. Rimantė Kunickaitė, Monika Zdanavičiūtė and Tomas Krilavičius. Fraud Detection in Health Insurance Using Ensemble Learning Methods. CEUR Workshop Proceedings (CEUR-WS.org) <https://ceur-ws.org/Vol-2698/p11.pdf>
4. Paresh Gohil, Dr. Sheshang Degadwala, Dhairya Vyas (2022). Fraud Detection in Medical Insurance Claim System using Machine Learning: A Review. November 2022 International Journal of Scientific Research in Computer Science Engineering and Information Technology DOI:10.32628/CSEIT228664
5. Hritik Kalra1 , Ranvir Singh , Dr.T. Senthil Kumar. (2022) Fraud Claims Detection in Insurance Using Machine Learning. Journal of Pharmaceutical Negative Results, Volume 13, Special Issue 3 <https://www.pnrjournal.com/index.php/home/article/download/498/351/638?shem=ssc> DOI: 10.47750/pnr.2022.13.S03.053
6. Arif Ismail Alrais. Fraudulent Insurance Claims Detection Using Machine Learning. A Capstone Submitted in Partial Fulfilment of the Requirements for the Degree of Master of Science in Professional Studies: Data Analytics. Rochester Institute of Technology RIT Scholar Works. Fraudulent Insurance Claims Detection Using Machine Learning (rit.edu)
7. Conghai Zhang, Xinyao Xiao, and Chao Wu (2020). Medical Fraud and Abuse Detection System Based on Machine Learning. Int J Environ Res Public Health. 2020 Oct; 17(19): 7265. Published online 2020 Oct 5. doi 10.3390/ijerph17197265 International Journal of environmental research and public health.