IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Impact Factor: 7.301

Volume 3, Issue 2, December 2023

A Comprehensive Analysis of Provider Fraud Detection through Machine Learning

Hole Prajakta Parshuram and Prof. S. G. Joshi
Department of Computer Engineering
Viswabharti College of Engineering, Ahmednagar, India
prajaktahole32@gmail.com and principal_vacoea@yahoo.com

Abstract*: This research paper presents a comprehensive analysis of healthcare provider fraud detection and analysis using machine learning, drawing insights from diverse literature surveys. The study employs a systematic approach to evaluate methodologies and insights from various academic fields. Leveraging the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement, the research synthesizes findings from 27 relevant studies out of 450 articles. The focus lies on characterizing healthcare fraud, emphasizing addressing the limitations and gaps identified in existing literature. The paper introduces a Sequential Forward Selection (SFS) method and SMOTE oversampling for fraud detection, utilizing K-Nearest Neighbors, Artificial Neural Network, Linear Discriminant Analysis, and Gradient Boosting Machine, Classification using a bagging classifier and a stacking meta-estimator. It is recommended to use the Stacking aggregator because it is statistically significant. This study intends to offer insightful information to researchers, legislators, and healthcare practitioners by resolving issues with and gaps in existing methodologies.*

Keywords: Fraud Detection, Machine Learning, Healthcare

I. INTRODUCTION

In the ever-changing world of healthcare, new technological advancements have opened doors to creative solutions that improve the effectiveness and reliability of healthcare systems. Healthcare provider fraud is a widespread problem with far-reaching effects on financial and patient-centric domains; detecting and analyzing this issue is a vital component demanding attention. This research study explores machine learning approaches for fraud detection and analysis of healthcare providers. Due to the growing complexity of healthcare ecosystems, an aggressive and complicated strategy is required to detect fraudulent activity successfully. Machine learning has emerged as a powerful technique for enhancing fraud detection capabilities since traditional methods are ill-equipped to handle complex fraudulent schemes. This research examines the use of machine learning algorithms in healthcare provider transactions, the algorithms' relative efficacy in detecting fraudulent patterns, and the algorithms' varied applications.

This study is based on the idea that a solid machine-learning system can spot outliers and change and adapt to new forms of fraud as they appear. Researchers hope to find signs of fraud by combing through massive databases that include provider invoices, claims, and transaction histories. To strengthen their defences against fraudulent practices, healthcare companies should adopt advanced analytics to acquire a more detailed knowledge of these trends. In addition, the article delves into the ethical concerns of using machine learning to identify healthcare fraud. With the growing importance of technology in healthcare, finding a middle ground between fighting fraud and protecting patient privacy and provider honesty is becoming more critical.

This research article delves into the critical function of machine learning in transforming the analysis and detection of healthcare provider fraud. The work adds to the growing conversation on protecting healthcare systems from fraud by delving into the complexities of algorithmic applications, ethical concerns, and practical implications; this will lead to a safer and more resilient healthcare system in the long run.

DOI: 10.48175/IJARSCT-14217

II. REVIEWS OF DIFFERENT TYPES OF FRAUD DETECTION AND TECHNIQUES

2.1 Systematic Review of Fraud Detection

This research aimed to catalogue, examine, and compile the several GBAD methods used in data mining fraud detection studies and constructed a classification framework to examine 39 scholarly articles through a systematic literature search. The questions used to do this research focused on various elements of GBAD-based fraud detection. Both theory and practice benefit significantly from this study's findings. Researchers can discover more about the implementation of GBAD approaches with the help of the suggested classification framework, which enables a methodical probe. The shortcomings that have been identified should motivate data scientists to conduct more empirical studies in this area. Similarly, this paper provides practitioners with a road map to better understand the relationship between their network's characteristics, various anomalies, and the ideal graph-based solutions for their specific requirements and use cases. [1]

Doctors and other medical professionals participating in medical schemes and those supplying health care are the most common fraud perpetrators. Insurance claim fraud is the most common type, including false claims and duplicates. The end goal of the fraud detection approach is to stop the loss of money and health services and inpatient treatment. Modifying procedures in response to the perceived condition or problem (fraud) is possible. The synthesis table results of previous reviews reveal that most research relies on secondary data to identify and attribute fraud events. For this reason, it would be beneficial for future studies to collect data on the types and regulations that apply to fraudsters in different countries and the method's limitations. This information could then be used to affirm the legitimacy of legal sanctions for fraud, regardless of the job or profession.[2]

The literature on Various detection techniques was explored based on attack intentions, including server, network, client software, client hardware, and user perspectives. The review identified future study directions and highlighted security challenges in cybersecurity planning. The analysis of existing literature revealed a growing number of studies on machine learning-based mobile malicious code detection. The classification of attack means into supervised and unsupervised learning showed a prevalence of studies on clients, with supervised learning algorithms proving more effective in detecting Android mobile malware. The comprehensive evaluation aims to guide future researchers in enhancing cybersecurity for mobile devices, particularly in the client section, using supervised learning algorithms. The synthesis of existing data provides valuable insights for ongoing and future research in machine learning-based mobile malware detection. [3]

The numerous methods employed for CCFD are examined in this review article. Analyses show that ML approaches significantly improve CCFD's accuracy. Nevertheless, to train the model and prevent the problem of data imbalance requires vast datasets. More diverse data can be provided with real-time information, but privacy is still a concern. This suggested approach allows us to train the model using real-time datasets while protecting users' privacy. The proposed strategy enables them to establish a more effective system for CCFD. The proposed method limits real-life deployment despite its effectiveness in CCFD employing privacy-preserving real-time datasets. Financial institutions and banks are notoriously stringent regarding their laws and regulations. Adapting the suggested method will be difficult because all financial institutions are unique and use their resources instead of a centralized strategy. Even without central data sharing, trained models will still learn patterns that hackers may potentially decipher. So, while the restrictions will remain in place, more effort is required to win over financial institutions and banks so that they can use this technology.[4]

This study systematically evaluates and synthesizes methodologies and data samples in peer-reviewed studies across various academic fields characterizing healthcare fraud. Employing the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement, 27 relevant studies out of 450 articles were analyzed using a qualitative case study approach. The review, utilizing 24 variables, revealed a challenge in quantitative comparison due to a lack of reported accuracy and overall fraud rates. Employing a validated approach, the qualitative assessment demonstrated high validity (r = 93%) in classifying fraud detection methods. Gaps identified include the need for method validation, proof of intent to commit fraud, absence of fraud rate estimates, and challenges in selecting the best detection method. Addressing these gaps through further research would benefit researchers, policymakers, and healthcare practitioners seeking optimal fraud detection methods. [5]

2.2 Machine Learning-Based Approach

This study article thoroughly evaluates machine learning methods for detecting medical fraud. According to a comprehensive experimental evaluation, the Multilayer Perceptron algorithm performs better than the competition regarding F-Measure, sensitivity, specificity, and accuracy.The algorithms used in the publications above were compared to enhance the accuracy of medical deceit detection. The goal was to identify and fix medical and allied healthcare fraud. The results demonstrate that, compared to the alternatives, Multilayer Perceptron Algorithm delivers substantially better performance. Deep learning algorithms will be used to do comparison analyses in the future. [6]

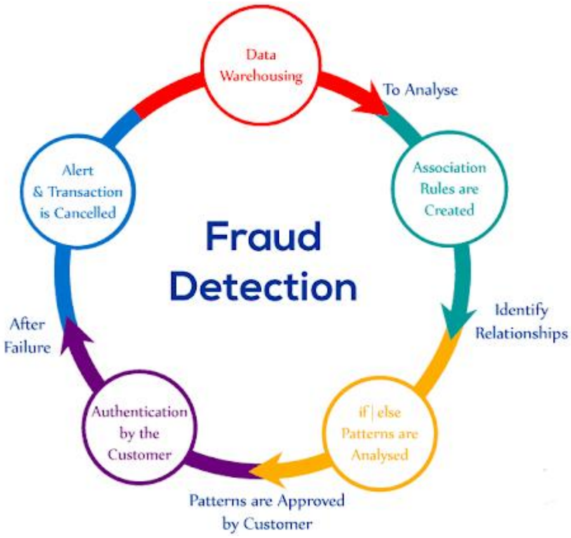
Data Warehousing
To Analyse
Association Rules are Created
Identify Relationships
if | else Patterns are Analysed
Patterns are Approved by Customer
Authentication by the Customer
After Failure
Alert & Transaction is Cancelled
Fraud Detection

Fig.1 Role of ML in Fraud Detection

Medical budgets have been steadily shrinking due to healthcare insurance fraud, and traditional, labour-intensive approaches to detecting fraud are inefficient. An efficient and reasonable way to detect healthcare insurance fraud is by using machine learning and deep learning techniques. They developed an algorithm to identify instances of healthcare claims fraud. This model achieved optimal accuracy and good assessment metrics in fraud detection using logistic regression, random forest, and artificial neural networks. In addition, the key characteristics that contributed to the result were highlighted by every model. Policy type, education level, and age were the most critical factors contributing to fraudulent behaviours. In future research, better generalization might be possible with larger datasets, additional variables, and various healthcare providers. [7]

Compared to the conventional rule-based approach, this model's main benefits include a simpler data type need, great practicality, and improved accuracy and recall, all of which make data analysts' jobs easier.Sought to improve the model's performance by limiting the range of dataand keeping the number of medications and disorders under 1000. This anomaly detection algorithm can only identify approximately 71% of data, a side consequence of improving accuracy. Check on the 1000 records dataset or use level-sample tests to validate the sampled result because of the large quantity of data and the limitations of the unlabeled dataset. These options need to be revised to cover every medicine and disease. However, they did not collect sufficient training data because of privacy and security concerns, which is a shortcoming of this article. There is significant noise in the information, and the medical history is incomplete despite best efforts to clear it multiple times. There is room for error in the results. Therefore, they are best used as a tool to help data analysts save time. The databases held by government agencies, hospitals, and related organizations are crucial for developing an ideal model; thus,we should make more of an effort to request access to them. [8]

A low false positive rate indicated cheap exactness, and the machine learning models applied to these datasets could identify the most erroneous cases. The relatively low levels of prediction were a consequence of specific knowledge sets' serious problems with data quality. Given the intrinsic peculiarities of different datasets, it would not bebrilliant to prescribe ideal algorithmic approaches or employ a feature engineering process to achieve much better performance. Afterwards, the models would be applied to specific business scenarios and user objectives. By focusing on certain types of fraud, loss management units may better ensure that their models are evolving to detect them. The models'

DOI: 10.48175/IJARSCT-14217

ability to detect new frauds and their success in back-testing make them an affordable option for usage in the insurance claims fraud detection area. [9]

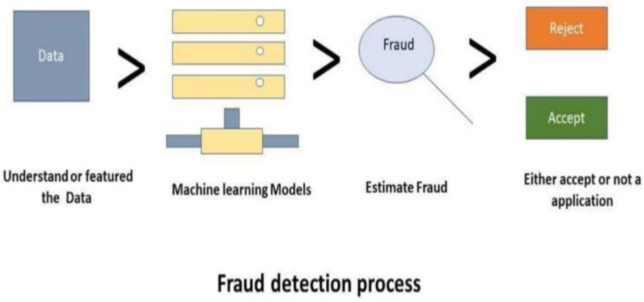

Fig.2 Process of Fraud Detection

Because of machine learning, healthcare has the potential to see numerous technological breakthroughs. It can help find trends and patterns in patient data, make diagnoses more accurate, streamline administrative processes, and allow personalized treatment plans. Data privacy concerns, ethical considerations, and stringent validation and regulation are obstacles to using machine learning in healthcare. To successfully integrate machine learning into healthcare, one must have an in-depth knowledge of the complex and ever-changing healthcare system, work closely with data scientists, and use machine learning ethically and responsiblyto benefit patients. [10]

The medical field has a long history of conducting hypothesis-driven studies that draw on epidemiologic and statistical expertise—developed multiple predictive ML algorithms thanks to the recent influx of massive amounts of varied data and the increased processing capacity of fast real and virtual machines. Many clinical decision support systems use these algorithms and can now forecast health metrics based on populations. As machine learning (ML) algorithms become more critical in healthcare research and clinical practice, healthcare providers must acquire the training to comprehend the specialized terminology. Equally crucial is for data scientists to be familiar with the conceptual similarities between data science and different ideas in epidemiology. When extrapolating results from these ML algorithms, keeping the ethical ideal of "no harm" in mind is crucial.[11]

This research utilizes a variety of classification algorithms to identify instances of fraud, including Random-Forest, decision tree, Support Vector Machine, K-Nearest Neighbor, Adaboost, Linear Regression, Naïve Bayes, and Multi-Linear Perceptron.Performed experiments on a trustworthy repository's auto-insurance dataset and audited numerous strategies to identify or modify the best classifier for the fraud detection system. In addition, the system has been evaluated for each method using precision, recall, and F1-score metrics.An ANFIS, a combination of neural networks and neuro-fuzzy algorithms, might be a future use case for this fraud detection technique. Using internal factors, the Hidden Markov Model (HMM) can detect fraud cases and increase prediction accuracy. [12]

Fraud detection is a crucial responsibility to safeguard financial systems and decrease the frequency of fraudulent activities. As the efficacy of long-standing procedures decreases, there is a rising desire for new fraud detection methods. According to the study, machine learning's ability to sift through large datasets in search of complex patterns makes it a promising tool for automated fraud detection. To offer a thorough understanding of its possibilities and constraints, this research investigated the current state of machine learning in fraud detection. Considering the ever-changing and intricate character of fraudulent operations, the research emphasized the significance of investigating machine learning methods for detecting fraud in financial transactions. The offered fraud detection task was best handled by the Random Forest model, according to evaluation criteria and overall performance. Its excellent recall rates, accuracy, and precision demonstrated its exceptional capacity to identify legitimate and fraudulent transactions accurately. [13]

Any medical professional, scientist, or researcher can benefit significantly from using ML. ML is seeing breakthroughs daily. New machine learning applications that address real healthcare issues arise with every new development. Many in the medical field are keeping a careful watch on the rapid development of ML. Concepts from machine learning are helping surgeons and doctors save lives, identify problems before they happen, improve patient management, get patients more involved in their rehabilitation, and a whole lot more. Organizations worldwide are enhancing healthcare delivery with the help of AI-powered technologies and ML models. Companies and pharmaceutical companies can use this technology to create cures for serious diseases more quickly and efficiently. Virtual clinical trials, sequencing, and pattern recognition have allowed businesses to speed up their observation and testing processes. More essential indicators of general health are health-promoting behaviours and socioeconomic characteristics such as income, social support networks, and level of education. Organizations in the health sector understand that improving people's health as a whole requires looking at their environment and lifestyle as well. ML models can detect individuals more likely to develop chronic diseases such as diabetes, heart disease, and others that can be prevented. [14]

Medical insurance fraud prediction is a hot area of study. Academic and corporate researchers alike find it to be an arduous undertaking. Health insurance firms' bottom lines have taken a hit due to the prevalence of anonymous actions in claims. This work examines a healthcare provider fraud detection dataset and analyses traditional Machine Learning (ML) classifiers such as k-mean clustering, Support Vector Machine (SVM), and Naive Bayes (NB). Methods of filtering are used to pre-process the data that has been acquired. Fraudulent claims can be detected from the various providers' diagnoses and the overall amount charged on a claim. Claim amounts, diagnoses, and providers make up the proposed claim data. Providethe accuracy, recall, and precision of the ML classifiers considered in the classification process, with the diagnosis attribute as the decision variable and the provider attribute as the target class. More research on the classifiers' FPR (False Positive Rate) is available. [15]

2.3 Feature Selection

A GA-based feature selection technique incorporating the RF, DT, ANN, NB, and LR was suggested in this study. The GA was applied with the RF as part of its fitness role. The fivebest feature vectors were produced after using Te GA to the dataset of credit card transactions made by European cardholders. The GA-RF (with v5) attained an optimal overall accuracy of 99.98%, according to the experimental results obtained using the GA-chosen qualities. Additional classifiers, like the GA-DT, used v1 and attained an impressive accuracy of 99.92%. The outcomes of this study outshone those of previously used approaches. In addition, the results obtained on the European credit card fraud dataset were validated by implementing the suggested framework on a synthetic dataset. The GA-DT achieved a perfect score of 100% accuracy and an area under the curve (AUC) of 1. Following closely behind with a perfect score and an area under the curve (AUC) of 0.94 is the GA-ANN. to employ additional datasets to evaluate the system in future research. [16]

Healthcare system frauds have led to increased expenses and a decline in patient service quality. This research introduces a Sequential Forward Selection (SFS) method and SMOTE oversampling for healthcare insurance fraud detection. Classification employs K-Nearest Neighbors (KNN), Artificial Neural Network (ANN), Linear Discriminant Analysis (LDA), Gradient Boosting Machine (GBM), Bagging classifier, and stacking meta-estimator, with Stacking aggregator preferred due to statistical significance. The study categorizes, compares, and summarizes healthcare fraud detection articles from the past decade, defining fraud types, subtypes, and data evidence nature. The proposed methodology achieves the highest accuracy (97.19%). TheStacking classifier's feature selection was confirmed with a literature-based real-world dataset on healthcare insurance fraud. [17]

Researchers have devised many approaches to the complex problem of detecting fraudulent credit card transactions. To improve the detection rate, this study suggested a hybrid method. Information gain, genetic algorithms, and extreme learning machines are just a few of the ML and feature-selection methods leveraged in the hybrid approach. Initially, features were selected using the IG technique, and the GA wrapper was fed the features that rated highest. As for the GA wrapper, it used the ELM as its learning algorithm. The suggested strategy achieved better results than competing baseline classifiers and approaches reported in the literature recently. In addition, the suggested method was used to validate its performance on two additional credit card datasets. In all datasets, it produced a good performance, proving its robustness. The suggested hybrid strategy is a reliable means of detecting credit card fraud. Furthermore, the possibility of acquiring more current datasets to train ML models should be investigated in future studies.[18]

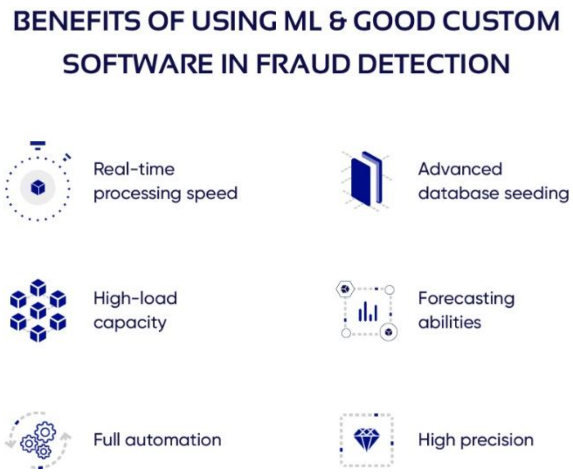
BENEFITS OF USING ML & GOOD CUSTOM SOFTWARE IN FRAUD DETECTION
Real-time processing speed
Advanced database seeding
High-load capacity
Forecasting abilities
Full automation
High precision

Fig. 3 Software in Fraud Detection

Financial fraud, including credit card fraud, has recently increased due to the rise of e-commerce and e-payment systems. Consequently, systems that can identify credit card fraud must be implemented. Selecting the right features of credit card fraud when using machine learning to detect them is crucial. This study provides a genetic algorithm (GA) feature selection engine based on machine learning (ML) that can detect credit card fraud. Decision Tree (DT), Random Forest (RF), Logistic Regression (LR), Artificial Neural Network (ANN), and Naive Bayes are some of the machine learning classifiers that the suggested detection engine employs after the optimum features have been selected (NB). A dataset created from European cardholders is used to test the suggested engine's ability to detect credit card fraud. This proposed solution outperformed previous systems, as shown by the result. [19]

2.4 Data Mining Method

This proposed technique is divided into two parts. The first part uses K to cluster providers to find similar and consistent hospital groups in treating a particular condition. The second part uses these groups to find outliers. The technique maintains high adaptability thanks to an initial grid search for the optimal feature selection (using Principal Feature Analysis) and the optimal number of local groups. Step two involves proposing a human-decision support system that auditors can use to cross-validate the outliers they have found, test them against fraud-related characteristics, and determine how complicated a casemix of patients they have handled they are. The suggested approach was evaluated using data collected over three years (2013–2015) on HDC by RegioneLombardia (Italy), emphasizing heart failure treatment. Results: From the 183 units, the model found 6 hospital clusters and 10 outliers. Detail the implementation of Step Two at three hospitals among these providers (two private and one public). Auditors found the two private providers intriguing and warranted additional research after cross-validating with the patient population and hospital features; the public hospital appeared to have good reason to be an outlier. [20]

There are victims of healthcare fraud, a costly white-collar crime in the US. The public pays the price for fraud through higher premiums or catastrophic losses beneficiaries suffer. Advancements in digital healthcare fraud detection systems are urgently required to address this pressing social concern. Digital healthcare innovations are challenging due to the heterogeneity and complexity of the United States' data systems and health models. Healthcare fraud detection aims to give investigators leads that can be further investigated to increase the likelihood of recoveries, recoupments, or referrals to relevant authorities or agencies. This page provides a synopsis of the literature on healthcare fraud detection systems and methodologies. Here is a table with all the peer-reviewed articles published in this field. Each article briefly summarises the study's goals, findings, and data. We will discuss the possible problems with using these technologies for healthcare data. The authors suggest other areas for further research to address these deficiencies.[21]

There are victims of healthcare fraud, a costly white-collar crime in the US. The public pays the price for fraud through higher premiums or catastrophic losses beneficiaries suffer. Advancements in digital healthcare fraud detection systems are urgently required to address this pressing social concern. Digital healthcare innovations are challenging due to the heterogeneity and complexity of the United States' data systems and health models. Healthcare fraud detection aims to give investigators leads that can be further investigated to increase the likelihood of recoveries, recoupments, or referrals to relevant authorities or agencies. This page provides a synopsis of the literature on healthcare fraud detection systems and methodologies. Find a table with all the peer-reviewed articles published in this field here. Each article briefly summarises the study's goals, findings, and data on the possible problems with using these technologies in actual healthcare data. The authors suggest many areas for further research to address these shortcomings. [22]

Supervised and Unsupervised Methods

This study demonstrates the efficacy of unsupervised machine learning, specifically using a tuned autoencoder, for detecting overutilization of procedure codes in healthcare claims. The model efficiently identifies outliers in millions of claims, serving as an automated pre-payment screening tool. The proposed approach involves flagging procedure codes that appear inconsistent with others in a claim, subject to subsequent verification by a human reviewer. The model's high F1-score (0.97) on the out-of-sample test dataset indicates low false positives and no false negatives for rare procedures in specific combinations. However, a lower F1-score (0.63) on the manually annotated test dataset suggests room for improvement in reducing false positives. Future efforts will focus on refining model precision to enhance performance. [23]

Health care in the United States costs over $4 trillion annually, with most of the spending going to private providers, who then compensate insurance companies. Overbilling, waste, and fraud by providers are significant issues with this system. They have an incentive to underreport their claims so they may get more money. Create new machine-learning technologies to find healthcare providers who overcharge insurance companies.Analysis of Medicare claims data reveals trends that may indicate inpatient hospitalization fraud or overbilling. Medicare is a government health insurance programme in the United States that covers the elderly and people with disabilities. With the suggested method, users can understand the reasons behind the flagged providers' possibly questionable conduct, and this approach to fraud detection is entirely unsupervised, meaning it does not depend on any labelled training data. Case studies of questionable providers and data from anti-fraud actions filed against providers by the Department of Justice supportmethodology and results. Hospital features not used for detection but associated with a high suspiciousness score can be better understood through a post-analysis, which also does. Public and private health insurance organizations can utilize the method to direct audits and investigations of questionable providers; it offers an 8-fold improvement over random targeting. [24]

Credit card fraud is at an all-time high, and con artists are actively targeting this industry. Data scientists and machine learning experts have created many algorithms to check for fraudulent transactions. This paper compares the accuracy of three ML models in identifying, predicting, and classifying fraudulent credit card transactions: logistic regression, random forest, and decision trees. By comparing the two models' outputs, we find that random forest is the most effective at predicting and detecting fraudulent credit card transactions, with an accuracy of 96% and an area under the curve value of 98.9%. Credit card fraud might be challenging to forecast and detect, but the random forest is the best machine-learning algorithm. A higher percentage of fraudulent transactions occurred between 22:00GMT and 4:00GMT, and the victims of these transactions tended to be credit card holders older than 60 years. [25]

2.5 Different Methods for Detection

Potentially saving billions of dollars in healthcare expenditures and improving overall patient care, automated technologies for spotting fraudulent healthcare providers offer great promise. This research proposes a data-centric strategy to enhance the accuracy and efficiency of healthcare fraud classification using Medicare claims data. Ten massive supervised learning datasets were built using publicly available Centers for Medicare & Medicaid Services (CMS) data. The first step is to use CMS data to compile the Medicare fraud classification sets for Parts B, D, and DMEPOS (Durable Medical Equipment, Prosthetics, Orthotics, and Supplies) from 2013 to 2019. presented an enhanced data labelling procedure and examined each dataset and data preparation method used to generate Medicare datasets for supervised learning. The next step is to add up to fifty-eight more provider summary attributes to the first Medicare fraud data sets. Lastly, fix a typical problem with model evaluations and provide a modified cross-validation method to reduce target leakage and produce trustworthy assessment outcomes.Using extreme gradient boosting, random forest learners, and several complementing performance indicators and 95% confidence intervals to assess each dataset on the Medicare fraud classification job. The results demonstrate that the new enhanced data sets routinely

DOI: 10.48175/IJARSCT-14217

surpass the performance of the original Medicare data sets utilized in related efforts. The findings support data-centric machine learning efforts and lay the groundwork for practical data interpretation and preparation methods for healthcare fraud machine learning applications. [26]

Improved provider fraud detection throughout the pre-processing and classification steps is one outcome of the suggested system. An orderly collection of links between diseases, patients, and claims variables is built initially through pre-processing the obtained datasets using a Relative Risk-based MapReduce architecture. The classification step suggests A Recurrent Neural Network as a workaround (RNN). It is an intricate process that uses hyperparameter optimization to consideressential qualities. One of RNNs' most robust features, recalling ability, defines the network's history and current state. Ultimately, the provider's Recurrent Neural Networks (RNNs) input layer is used to identify anomalies based on the best qualities with the chosen hyperparameters. The public repositories test and validate the proposed PF ADS architecture. The suggested framework beats the previous methods in terms of accuracy (88.09%), precision (14.15%), recall (32.80%), and computing time (92.30 seconds), according to the experimental data. [27]

The challenge of detecting health insurance fraud was the focus of this research. To investigate various behavioural correlations of patient visits, a genuine healthcare dataset. An analysis was conducted to determine the effect of fraud detection on interactions between various items in a healthcare scenario. An AHIN in a model named MHAMFD captured these interactions. To increase the quality of neighbour nodes, leverage different degrees of behavioural relationships to choose suitable neighbours.The resulting composite semantic information is taken into account here. Accumulating behavioural interactions at various levels allows for thoroughly learning the embedding representation of target nodes within and between them. Experiments using real-world medical data confirmed that MHAMFD effectively detects health insurance fraud. Explanatory research on using MHAMFD to detect health insurance fraud in various contexts will be conducted in future work. [28]

Using four separate datasets, this research suggested a way to improve recall in detecting credit card fraud. When tested against competing models,This method proved to be competitively accurate. Future fraud detection algorithms may be guided by the datasetsused on the Kaggle platform, which are publicly available. Medical diagnostics, catastrophe prediction, and airport security are just a few examples of the areas where the approach could be helpful. There is much room to grow the approach in the future. Hope to include it in a flexible and evolving architecture that can detect fraud in online banking and other areas in real time. This methodology's inherent flexibility makes it applicable to various domains, such as online banking, e-commerce platforms, and the ever-changing mobile payment systems. [29]

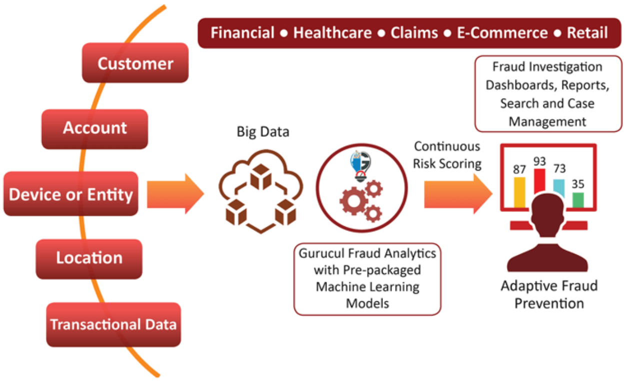
Financial ● Healthcare ● Claims ● E-Commerce ● Retail
Customer
Account
Device or Entity
Location
Transactional Data
Big Data
Gurucul Fraud Analytics with Pre-packaged Machine Learning Models
Continuous Risk Scoring
Fraud Investigation Dashboards, Reports, Search and Case Management
87
93
73
35
Adaptive Fraud Prevention

Fig.5 Fraud Monitoring

Insurance fraud is constantly expanding in both scope and quantity. There will inevitably be strategies based on artificial intelligence for embezzlement detection since manual methods are not practical. Researchers are developing Machine Learning methods to tackle the challenges in this field. To enhance the XGBoost model's output, this study utilized the Bayesian Optimization (BO) strategy. With a 98% success rate, the suggested BOXGBoost model has been deemed the top choice for deception prediction. Predicting health insurance fraud and minimizing claim adjudication time could be made easier using the BOXGBoost-based system. This long-term goal is to improve the model's performance by implementing balancing strategies. Furthermore, intend to investigate the potential of transfer learning techniques, enabling us to evaluate and contrast the model's efficiency with other health insurance claim databases. Additionally,it intended to build a unified platform for the health insurance industry that would house a protected database with fraud detection algorithms powered by Machine Learning. This project would also investigate the practical and legal obstacles in the actual world. In the fight against health insurance fraud, pray that the platform based on machine learning will be victorious. [30]

To develop a healthcare fraud detection model, this article reviews numerous ML and DL models. On the Ayushman Bharat healthcare dataset, 26 different models were evaluated (PM-JAY). SMOTE, ADASYN, and TGANs were employed to address the dataset's imbalance. The neural networks' performance improved when trained on an undersampled dataset compared to other classification models. The F1-score, a measure of the model's accuracy and recall, was 0.95, the highest value. To determine the most effective deep learning or machine learning models, a comparable study can be conducted with different business lines in the future. [31]

Even though everyone knows that healthcare fraud hurts, it keeps happening all around the globe. The abundance of reports and figures published in the literature provides undeniable evidence. The revolutionary potential of blockchain technology to supply modern solutions cannot be overstated. The most pressing issue is the urgent need for a new method to handle health insurance claims, given the enormous sums of money wasted each year due to fraud. This study proposes a new, safe, data-driven method for health insurance claims processing and submission by building and testing a blockchain-based system that leverages domain data and machine learning to determine the likelihood of fraudulent claims. Results from the machine learning trials show that the suggested method achieved a claim data classification accuracy of about 98%. Similarly, a 2% mistake rate will be used for future claim classifications. Considering the yearly amounts lost to fraud worldwide, the long-term benefits of adopting the proposed approach outweigh the short-term costs. Transitioning from a centralized system to a decentralized Blockchain-based system ensures security, efficiency, and high data integrity in claims processing and significantly increases efforts against fraud. [32]

III. CONCLUSION

In conclusion, this research paper offers a nuanced understanding of healthcare provider fraud detection and analysis through meticulously examining existing literature surveys. Utilizing machine learning techniques, including the innovative application of the Stacking aggregator, showcases promising avenues for enhancing fraud detection capabilities. The study underscores the importance of addressing challenges such as method validation, intent proof, and the absence of fraud rate estimates in current research. The proposed methodology, demonstrated with a real-world healthcare insurance fraud dataset, achieves a notable accuracy of 97.19%, emphasizing its practical efficacy. The research consolidates and compares methodologies from the last decade, highlighting the need for future studies to validate existing methods and bridge the identified gaps. This paper contributes to the ongoing discourse on fortifying healthcare systems against fraudulent activities, providing a foundation for future research in this critical domain.

REFERENCES

[1]. Tahereh Pourhabibia,, Kok-Leong Ongb, Booi H. Kama, Yee Ling Boo, "Fraud detection: A systematic literature review of graph-based anomaly detection approaches", Decision Support Systems, 2020

[2]. Andi Yaumil Bay R. Thaifur a,b,∗, M. Alimin Maidinc, Andi Indahwaty Sidinc, Amran Razak, "How to detect healthcare fraud? "A systematic review", Gac Sanit. 2021

[3]. Yu-kyung Kim , Jemin Justin Lee , Myong-Hyun Go ,Hae Young Kang and Kyungho Lee , “A Systematic Overview of the Machine Learning Methods for Mobile Malware Detection”, Hindawi Security and Communication Networks Volume 2022

[4]. Rejwan Bin Sulaiman, Vitaly Schetinin, Paul Sant, "Review of Machine Learning Approach on Credit Card Fraud Detection", Human-Centric Intelligent Systems, 2022

[5]. Jing Ai,Jennifer Russomanno,Skyla Guigou&Rachel Allan, "A Systematic Review and Qualitative Assessment of Fraud Detection Methodologies in Health Care", 2022

[6]. S. Lavanya,S.ManojKumar ,P.Mohan Kumar, "Machine Learning Based Approaches for Healthcare Fraud Detection: A Comparative Analysis", Annals of RSCB, ISSN:1583-6258, Vol. 25, Issue 3, 2021

[7]. Eman Nabrawi and Abdullah Alanazi, "Fraud Detection in Healthcare Insurance Claims Using Machine Learning", MDPI Risks 2023

DOI: 10.48175/IJARSCT-14217

[8]. Conghai Zhang, Xinyao Xiao and Chao Wu, "Medical Fraud and Abuse Detection System Based on Machine Learning", Int. J. Environ. Res. Public Health 2020

[9]. Abhijeet Urunkar, Amruta Khot, Rashmi Bhat, Nandinee Mudegol, "Fraud Detection and Analysis for Insurance Claim using Machine Learning", EEE International Conference on Signal Processing, Informatics, Communication and Energy Systems (SPICES), 2022

[10]. Qi An, Saifur Rahman, Jingwen Zhou and James Jin Kang, "A Comprehensive Review on Machine Learning in Healthcare Industry: Classification, Restrictions, Opportunities and Challenges", MDPI Sensors 2023

[11]. Abdullah Alanazi, "Using machine learning for healthcare challenges and opportunities", Informatics in Medicine Unlocked, 2022

[12]. Laiqa Rukhsar, Waqas Haider Bangyal, Kashif Nisar, Sana Nisar, "Prediction of Insurance Fraud Detection using Machine Learning Algorithms", Mehran University Research Journal of Engineering and Technology, 2022

[13]. Md Sumon Gazi, Rejon Kumar Ray, "Exploring Machine Learning Techniques for Fraud Detection in Financial Transactions", Chinese Journal of Geotechnical Engineering, 2023

[14]. Mohd Javaid, Abid Haleem, Ravi Pratap Singh, Rajiv Suman, Shanay Rab, "Significance of machine learning in healthcare: Features, pillars and applications", International Journal of Intelligent Networks, 2022

[15]. A. Jenita Mary;S. P. Angelin Claret, "Analytical study on fraud detection in healthcare insurance claim data using machine learning classifiers", AIP Conf. Proc. 2516, 240006, 2022

[16]. Emmanuel Ileberi1*, Yanxia Sun1 and Zenghui Wang, "A machine learning based credit card fraud detection using the GA algorithm for feature selection", Journal of Big Data, 2022

[17]. Anuradha Mohanta & Suvasini Panigrahi, "Health Insurance Fraud Detection Using Feature Selection and Ensemble Machine Learning Techniques", Advances in Distributed Computing and Machine Learning, 2023

[18]. Ibomoiye Domor Mienye and Yanxia Sun, "A Machine Learning Method with Hybrid Feature Selection for Improved Credit Card Fraud Detection", Appl. Sci. 2023

[19]. Emmanuel Ileberi, Yanxia Sun & Zenghui Wang, "A machine learning based credit card fraud detection using the GA algorithm for feature selection", Journal of Big Data, 2022

[20]. Michela Carlotta Massi, Francesca Ieva and Emanuele Lettieri, "Data mining application to healthcare fraud detection: a two-step unsupervised clustering method for outlier detection with administrative databases", BMC Medical Informatics and Decision Making (2020

[21]. Nishamathi Kumaraswamy, MS; Mia K. Markey, PhD; Tahir Ekin, PhD; Jamie C. Barner, PhD, FAACP, FAPhA; and Karen Rascati, PhD, "Healthcare Fraud Data Mining Methods: A Look Back and Look Ahead", Perspectives in health information management / AHIMA, American Health Information Management Association, 2022

[22]. Nishamathi Kumaraswamy, MS, Mia K. Markey, PhD, Tahir Ekin, PhD, Jamie C. Barner, PhD, FAACP, FAPhA, and Karen Rascati, "Healthcare Fraud Data Mining Methods: A Look Back and Look Ahead", Perspect Health Inf Manag. 2022

[23]. Michael Sussman, Samantha Gorny, Daniel Lasaga, John Helms, Dan Olson, Edward Bowen and Sanmitra Bhattacharya, "Procedure code overutilization detection from healthcare claims using unsupervised deep learning methods", BMC Medical Informatics and Decision Making, 2023

[24]. Shubhranshu Shekhar, Jetson Leder-Luis, Leman Akoglu, “Unsupervised Machine Learning for Explainable Health Care Fraud Detection”, arXiv:2211.02927v3, 2023

[25]. Jonathan Kwaku Afriyie, Kassim Tawiah, Wilhemina Adoma Pels, Sandra Addai-Henne, Harriet Achiaa Dwamena, Emmanuel Odame Owiredu, Samuel Amening Ayeh, John Eshun, "A supervised machine learning algorithm for detecting and predicting fraud in credit card transactions", Decision Analytics Journal, 2023

[26]. Justin M. Johnson, Taghi M. Khoshgoftaar,"Data-CentricAI for Healthcare Fraud Detection", SN Computer Science, 2023

[27]. Jenita Mary & S. P. Angelin Claret, "Design and development of big data-based model for detecting fraud in healthcare insurance industry", Application of soft computing, 2023

[28]. Jiangtao Lu, Kaibiao Lin, Ruicong Chen, Min Lin, Xin Chen and Ping Lu, "Health insurance fraud detection by using an attributed heterogeneous information network with a hierarchical attention mechanism", BMC Medical Informatics and Decision Making, 2023

[29]. Jiwon Chung and Kyungho Lee, "Credit Card Fraud Detection: An Improved Strategy for High Recall Using KNN, LDA, and Linear Regression", MDPI Sensors 2023

[30]. Saravanan Parthasarathy, Arun Raj Lakshminarayanan, A. Abdul Azeez Khan, K. Javubar Sathick, Vaishnavi Jayaraman, "Detection of Health Insurance Fraud using Bayesian Optimized XGBoost", International Journal of Safety and Security Engineering, 2023

[31]. Rohan Yashraj Gupta, Satya Sai Mudigonda, Pallav Kumar Baruah, "A Comparative Study of Using Various Machine Learning and Deep Learning-Based Fraud Detection Models For Universal Health Coverage Schemes", International Journal of Engineering Trends and Technology, 2021

[32]. Anokye Acheampong Amponsah, Adebayo Felix Adekoya, Benjamin Asubam Weyori, "A novel fraud detection and prevention method for healthcare claim processing using machine learning and blockchain technology", Decision Analytics Journal, 2022