IJARSCT
ISSN: 2581-9429
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Volume 5, Issue 2, May 2025
Impact Factor: 7.67

Analysis of Data Balancing and Advanced Machine Learning Techniques

Mr. P. Manikanda Prabu[1], R. Dharaniga[2], S. Kuralarasi[3], S. Meenalochani[4]
Professor, Department of Computer Science and Engineering[1]
Student, Department of Computer Science and Engineering[2,3,4]
Anjalai Ammal Mahalingam Engineering College, Thiruvarur, Tamilnadu, India
maniadt2006@gmail.com, dharanigaravi2004@gmail.com, kuralarasics5@gmail.com ,
meenalochanisakthivel@gmail.com

Abstract: *Medical insurance fraud poses a gigantic threat to healthcare systems globally with enormous financial losses and discrediting legitimate claimants. Below is the implementation of a machine learning model for medical insurance fraud detection with focus on addressing class imbalance using Synthetic Minority Over-sampling Technique (SMOTE). The system suggested here is having multiple modules like data gathering, preprocessing, feature extraction, model training, and real-time fraud detection. Various machine learning algorithms like Decision Trees and XGBoost were compared on the efficiency level to identify sophisticated patterns of fraud. The SMOTE technique was utilized to balance the data, and there was significant improvement in model performance by generating synthetic samples for the minority (fraud) class. The model was trained and validated on actual healthcare claim data, and performance was measured in terms of accuracy, precision, recall, and F1-score. Results indicate improved detection rates and reduced false positives compared to traditional models. This implementation is designed to help healthcare providers and payers minimize financial risk and improve claims handling integrity using a clever, scalable, and comprehensible solution for fraud detection.*

Keywords: Medical Insurance Fraud, Machine Learning , SMOTE, Fraud Detection, XGBoost

I. INTRODUCTION

With the technologically driven healthcare system in the modern world, enhanced cases of bogus medical insurance claims are one of the significant problems. In addition to incurring colossal financial losses—more than billions of dollars annually—the fake activities further damage the credibility of healthcare systems and raise insurance premiums for honest policyholders. Traditional rule-based and manual audit systems, as effective as they are against explicit fraud, do not detect complex and evolving schemes of modern-day fraudsters, especially those that are conspiring or colluding through covert networks.

To address such issues, this project proposes a strong, intelligent fraud detection system based on advanced machine learning (ML) techniques in combination with balancing data strategies. Central to the system is the application of Synthetic Minority Over-sampling Technique (SMOTE) and hybrid balancing techniques to address the class imbalance that is generally present in health-related data, where authentic claims dominate fraudulently issued claims to an unprecedented degree.

The architecture also utilizes ensemble learning models such as XGBoost and Random Forest that possess high predictability and interpretability ability. The models are trained on fully preprocessed and feature-engineered data in the effort to uncover statistical and context patterns imbued with fraud claims. Simple classifiers such as Logistic Regression are also examined in an effort to compare performances against model complexity.

The architecture of the framework is targeted towards the most important modules like gathering of data from different sources of healthcare, preprocessing for redundancy removal and noise, feature extraction that is specific towards billing fraud and provider-patient relationship, and real-time fraud detection. Modular in architecture, a framework like this supports the possibility of handling varying fraud patterns, minimizes false positives, and enables timely detection.

DOI: 10.48175/IJARSCT-26247

With the application of hybrid balancing techniques combined with ensemble machine learning and high-level analytics, the project is a paradigm shift towards more automated fraud detection towards more financial strength and operating efficiency in healthcare systems.

Medical Insurance Fraud

Health insurance fraud is intentional misrepresentation or dishonesty that results in the unauthorized receipt of benefits from health insurance companies. Normal fraudulent tactics encompass billing for services not delivered, submitting fictitious patient records, abuse of upcoding for services, or double billing. All these frauds have significant economic consequences, with billions of dollars lost within healthcare systems worldwide annually. This, in turn, generates higher costs and depletes the resources for actual patient care. With increasingly computerized medical files, criminals also take advantage of loopholes in data systems. Rule-based fraud detection systems and manual audit systems cannot adapt to fast-changing medical fraud. Dynamic and fluid nature of such fraud requires smart, scalable, and automated systems. Therefore, contemporary anti-fraud systems of the current times pay special attention to utilizing artificial intelligence (AI) and machine learning (ML) in an attempt to handle the huge volumes of healthcare claims data efficiently. The idea is to identify and flag suspicious patterns of claim behavior in real-time. Also, through the use of advanced algorithms, healthcare organizations are able to facilitate early detection and mitigate operating losses. Thus, medical fraud detection is vital to health insurance companies' financial well-being and the improvement of healthcare claims verification process efficiency. With evidence-based initiatives, organizations will be able to expect fraud schemes in the future and react reactively against them. The current project focuses particularly on enhancing fraud detection accuracy through hybrid data balancing and machine learning. The main aim is reducing false positives and maximizing fraud detection accuracy. Second, real-time analyzing and alerting also assist with timely action as well as investigation. Lastly, fraud tackled at a large magnitude significantly increases the public's level of trust as well as the provision for sustainability in the health insurance industry.

Machine Learning (ML)

Machine Learning (ML) is the key driver of present day data analytics and fraud prevention. It is an area of artificial intelligence (AI) that provides mechanisms for systems to learn from pattern data and improve their performance in incremental steps without explicit coding. ML models depend on large sets of data to detect sophisticated patterns, relationships, and outliers. Along the way, ML is used to create predictive models to flag true and false healthcare insurance claims. The models are created using labeled data and tuned by performance measures such as precision, recall, and accuracy. Applications of ML include supervised learning, where historical data is employed as the basis for training classifiers such as logistic regression, decision trees, and XGBoost. All such models learn via projecting labeled output on input feature, thereby supporting easy classification of future data. The system further supports unsupervised learning practices, which detect latent patterns hidden in unlabeled data, hence revealing previously unimaginable fraud tendencies. Reinforcement learning, one of the other ML practices, allows systems to learn and take actions in a surrounding environment with feedback received after discovering optimal action. In the detection of health care fraud, ML increases the scalability, efficiency, and effectiveness of fraud detection procedures. ML models above regular systems are more capable of adapting to changing fraud tactics and exhibit greater generalization between different data sets. Additionally, with regular retraining, the models adjust to change to accommodate changing fraudulent tactics. The use of ensemble techniques such as Random Forest and boosting techniques also guarantee reliability of predictions. By automatically making decisions, ML minimizes the workload of investigators and simplifies fraud detection workflow. Python is used as the modeling implementation language in the project, and the development focus area is PyCharm. The application of ML is crucial in the automation and making healthcare systems intelligent in fraud detection.

SMOTE (Synthetic Minority Over-sampling Technique)

SMOTE or Synthetic Minority Over-sampling Technique is a machine learning resampling algorithm addressing the issue of class imbalance. In the case of medical fraud data, fraud claims are significantly fewer in number than valid

DOI: 10.48175/IJARSCT-26247

claims. Such an imbalanced ratio can affect the model's performance until the model gets biased in favor of the majority class and the minority instances, i.e., frauds, are neglected. SMOTE remedies this by creating examples of the minority class from the provided instances. Synthetic instances are created through interpolation among a nearest neighbor's feature vectors of the minority class. SMOTE balances the training set in a way such that patterns of both classes are accurately learned by the model. SMOTE is used in this project before training classification models for best sensitivity to detect fraud claims. The method protects the machine learning algorithms from being biased towards real statements, which is most of the time accurate in imbalanced datasets. SMOTE improves other important key performance metrics such as recall and F1 score, which are extremely crucial in fraud detection scenarios. The classifier gains the ability to identify weak fraud patterns that would otherwise go undetected while training on the balanced set. Used in combination with ensemble learners like Random Forest or XGBoost, SMOTE increases depth of learning and accuracy of decisions. Being simple, effective, and even versatile enough to be used with other preprocessing methods, it is an excellent addition to data-driven anti-fraud procedures. SMOTE can even be used with healthcare data where high dimensionality and high variability necessitate strong balancing methods. Application of SMOTE becomes critical in directing the predictive model to become fair and trustworthy. The paper further investigates applying SMOTE alongside undersampling and enhanced resampling strategies to improve general detection performance. SMOTE hence is part of our data preprocessing process, which enables us to successfully train the model for fraud classification.

XGBoost

XGBoost is short for Extreme Gradient Boosting and is a highly efficient and scalable machine learning algorithm that can be used for classification and regression. It is a variant of the boosting concept in which multiple weak predictors, usually decision trees, are sequentially summed up to create a strong predictor.Every next model in the cascade aims to do better than the last one's errors and therefore the overall predictive capacity. XGBoost is employed in this project for the detection of spurious medical insurance claims from large and intricate health care data sets. Among the benefits of using XGBoost is that it facilitates imbalanced and high-dimensional data processing, which is common in fraud detection issues. Compared to most other older algorithms, XGBoost is correct and efficient rather than never both of these, making it perfect for real-time analytical applications. XGBoost applies regularization techniques in overfitting prevention in a manner that the model is able to generalize very well to new data. The algorithm is tree pruning enabled and parallel processing enabled as well, which further improves training efficiency. Its capacity to detect fine patterns in the data, i.e., billing anomalies or treatment claim discrepancies, is one of the strengths of XGBoost. In terms of interpretability, XGBoost provides feature importance scores that allow decision-makers to make decisions on features with large effects on the fraud prediction. Interpretability is one of the strengths in private contexts like healthcare, where transparency and accountability are strongly desired. The reason why XGBoost has been selected for this system is that it already has experience of success in competitions and real-world projects on noisy or incomplete data. Preprocessing methods like SMOTE and feature engineering greatly improve the performance of XGBoost in fraud detection in medical insurance. The very high values of Area Under Curve (AUC) and the removal of false positives at a constant proportion are the reasons why the model's performance is justified. Therefore, XGBoost is a key element of the proposed predictive model architecture that serves significant functions in the detection of medical insurance fraud with high efficiency and high accuracy.

Data Balancing

Balancing data is among the most significant preprocessing techniques embraced in addressing the class imbalance issue in data sets, especially in fraud detection. In health insurance, spurious claims are a minority fraction of the whole data, and therefore they are difficult to detect for machine learning algorithms. Unless balanced, models will over-predict the majority class—true claims—and hence become insensitive to fraud and contain too many false negatives. To counteract this, the project utilizes a variety of data balancing techniques, including oversampling, undersampling, and hybrids. Oversampling methods like SMOTE create extra samples of the minority class artificially to balance the dataset. Whereas oversampling trashes examples of the majority class for the purpose of balance but losing information, the hybrid does the two as it attempts to reap the strengths of one without providing vulnerabilities to either. Achieving balance within data means that fraud and clean claims are presented similarly well both when training, hence the trained model being prepared for discriminative characteristics by both classes. This translates into improved performance metrics, particularly recall and F1 score, that are paramount in correct fraud detection. Balanced datasets also reduce model bias and ensure maximum fairness of prediction. Balanced treatment is required in attaining maximum utility of classifiers like XGBoost and Random Forest, which depend on data quality for best performance. Balancing also results in stable training and prevention of imbalanced decision boundaries. Data balancing is included as a necessary preprocessing step in this project that directly influences the system to detect sophisticated and subtle fraudulent activity. The method avoids fraud detection models from neglecting outlier but significant fraud signals. Through ensuring adequate experimentation and validation, the optimal balancing technique is selected for final model training. Overall, data balancing enhances the stability of the model, promotes genuine fraud classification, and enables real-time fraud detection with minimal false alarms.

II. RELATED WORK

Detection of Medical Fraud through Rule-Based and Statistical Approaches

Early medical fraud detection systems relied heavily on sign-rule systems and statistical techniques for identifying variations in healthcare claims. Such systems utilized pre-programmed fraud rules and flagged patterns, such as duplicate billing, unusually high charges for a category of service, or accumulating service frequency irregularities. Although beneficial in highlighting previously known patterns of fraud, these methods were too rigid to identify new fraud patterns of fraud mechanisms. One of the key weaknesses of these tools of detection was inflexibility as well as not learning from new fraud cases; further, such systems have high true positive rates and had high need for human participation to make verifications, which extended investigation and created inefficiency. These limitations created the demand for more powerful and smarter solutions that could process big, complex, and high-dimensional data in real-time.

Fraud Detection via Logistic Regression and Decision Trees

Some of the research papers augmented some type of machine learning classifiers, such as Logistic Regression and Decision Trees, to form predictive models to detect fraud. These were trained models based on historical claims data, providing better accuracy and automation compared to rules-based systems. Logistic Regression had transparency and explainability, thus appropriate for highly regulated industries, like healthcare. Decision Trees, conversely, facilitated hierarchical decision-making based on claim attributes like provider ID, claim value, and diagnosis code. Nonetheless, both models were prone to class imbalance and overfitting on noisy or unbalanced datasets. Consequently, their fraud detection was primarily limited to known or recurring fraud patterns.

Ensemble-Based Techniques for Improved Fraud Detection

To address the constraints of individual classifiers, researchers delved into ensemble learning techniques like Random Forest and XGBoost. These are models that combine several weak learners (usually decision trees) to enhance accuracy and resilience. Ensemble models in fraud detection proved effective in detecting complex patterns of fraud by taking advantage of feature importance and dealing with noisy data. XGBoost, in specific, attracted attention for its scalability and class imbalance handling capability, thus making it ready for real-time fraud detection. The models consistently outperformed baseline classifiers with regard to precision, recall, and F1 score. Their performance still relied on the quality of data preprocessing and feature engineering

Data Balancing Approaches in Imbalanced Fraud Datasets

Class imbalance poses a serious problem in fraud detection, where legitimate claims far outnumber fraudulent claims. Researchers proposed several data balancing methods, including undersampling, oversampling, and SMOTE (Synthetic Minority Over-sampling Technique) to enhance model performance. SMOTE became one of the most popular methods as it can create synthetic fraud instances by interpolating existing samples. Experiments demonstrated that the use of SMOTE prior to training of models dramatically improved fraud detection sensitivity. Hybrid balancing approaches that merge SMOTE with random undersampling were also observed to minimize false positives and maximize classifier fairness. However, their misuse could result in overfitting or loss of information, highlighting the requirement for cautious application.

Recent Advances and System Integrations

New trends in medical fraud detection prioritize real-time analysis, auto- alerting, and adaptive learning. Some approaches incorporate electronic health records (EHRs) and insurance claims platforms to review and authenticate claims on a constant basis. Other systems apply deep learning to advanced feature extraction but at the expense of heavy data requirements and computing resources. Solutions involving data balancing, ensemble modeling, and real-time detection pipelines indicate great success in reducing fraudulent schemes. Yet, the majority of these works do not have comparative studies across several classifiers under conditions of balanced data, which this work remedies.

III. SYSTEM MODEL

The designed system is for detecting medical insurance fraud claims via machine learning and data balancing methodologies. The architecture of the system involves several linked modules that provide data collection, preprocessing, feature extraction, training of models, fraud prediction, and user interfaces. Each module serves a purpose in the pipeline and together achieves accurate and timely fraud detection.

Data Collection Module

The data module is the bedrock of the system, tasked with collecting rich healthcare data. Patient demographics, diagnosis codes, procedure information, claim values, and provider details are some of the data that are collected.The data source is insurance repositories, claim history files, and healthcare organizations. The module brings together structured (table) and unstructured (text) data in various forms to a common repository. This allows access to a rich, heterogeneous dataset for model training and testing.

Data Preprocessing Module

This module deals with raw healthcare data transformation to cleaned and usable data. The imputation method is used to handle missing values, the duplicate records are removed, and numerical and categorical fields are normalised. The categorical values are encoded through one-hot or label encoding. Normalization and outlier handling are performed to ensure data consistency. The preprocessing step plays a critical role in enhancing the quality and reliability of the input data before the modeling process.

Feature Extraction Module

The feature extraction module discovers and picks out the relevant predictors that affect fraud classification.These are temporal features such as claim submission date, financial outliers such as abnormal billing amount, and relational features between patients and providers. The aim is to minimize data dimensionality without compromising important information. Feature engineering is used to build new variables from available data to enhance model performance, such as visit frequency, billing patterns, and treatment behavior.

Data Balancing Module

To counter class imbalance in detecting fraud, the system employs a hybrid data balancing module. The module employs methods like SMOTE (Synthetic Minority Over-sampling Technique) for oversampling the minority class (fraud cases) and random undersampling of the majority class (genuine cases). Balanced data is a prerequisite for training fair machine learning models. Through the hybrid model, representative patterns from both classes are learned, enhancing fraud case sensitivity.

Model Training Module

Model training module is designed to create prediction models based on machine learning models. It takes balanced and feature-processed data to train the classifiers such as Logistic Regression, Decision Tree, Random Forest, and XGBoost. Models are trained through cross-validation and hyperparameter optimization to achieve better performance. Evaluation metrics like accuracy, precision, recall, and F1 score are computed for comparison of the effectiveness of the models. Stored models are used for deployment in the fraud detection module

DOI: 10.48175/IJARSCT-26247

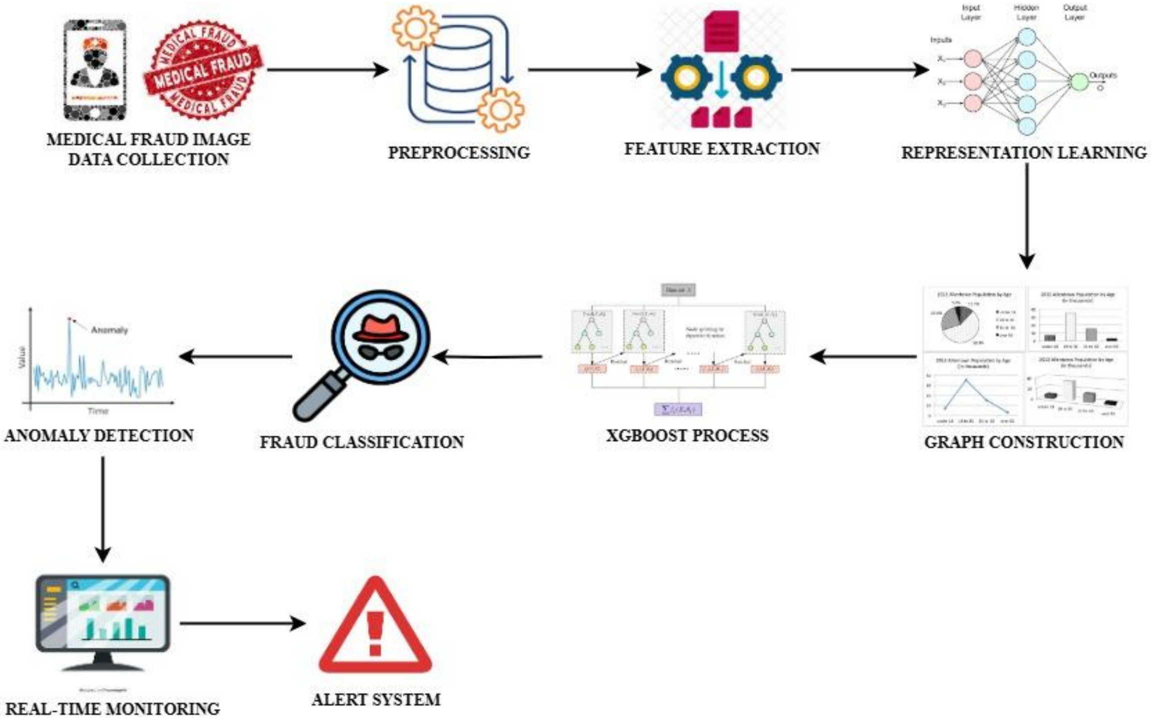
Fraud Detection and Prediction Module

This module applies the trained models to examine real-time claim data and identify the probability of fraud. It treats real-time claims similarly using the same preprocessing and feature extraction as it feeds the model. Depending on the output probability scores, claims are flagged for review or processed routinely. High-confidence fraud scores generate alerts that are sent to be investigated. The system constantly watches new data and adjusts its behavior in response to new fraud patterns unfolding.

User Interface Module

The user interface component provides an insurance auditor and health authority dashboard to communicate with. The UI displays fraud probability scores, model results, flagged claims, and performance metrics. Trends in data can be graphed, reports can be exported, and users can post feedback for the model choices. The UI allows model threshold settings and running of datasets, adding to an interactive and controllable fraud detection setting.

IV.SYSTEM ARCHITECTURE



V. IMPLEMENTATION

Algorithm

Step 1: Retrieve the healthcare insurance claims dataset from various sources like hospitals, clinics, and databases of insurance providers.

Step 2: Process the data by addressing missing values, eliminating duplicates, categorical variable encoding, and normalization of numerical features.

Step 3: Utilize feature extraction methods to obtain important variables like billing frequency, provider-patient relationship, trend in claim amount, and diagnosis patterns.

Step 4: Apply data balancing via SMOTE oversampling of the minority (fraudulent) class and undersampling of the majority (legitimate) class, if needed, to prepare a balanced dataset.

Step 5: Split the dataset into training and testing sets while keeping stratified sampling to preserve class distribution.

Step 6: Train various machine learning models such as Logistic Regression, Decision Tree, Random Forest, and XGBoost with the balanced training dataset.

Step 7: Compare all models based on performance metrics like Accuracy, Precision, Recall, and F1 Score on the test dataset.

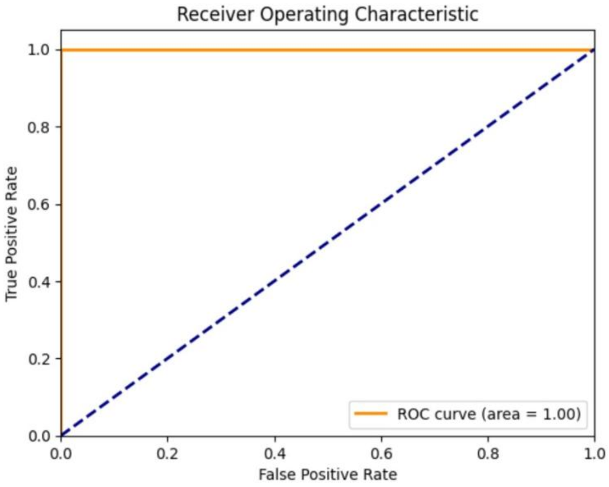
Step 8: Choose the top-performing model based on evaluation outcomes (XGBoost or Random Forest in most instances) for production deployment.

Step 9: Apply the trained model to predict fraud likelihood on real-time or batch claims data.

Step 10: Identify high-risk claims for manual investigation or review and render results in the user interface dashboard for healthcare auditors and insurance regulators.

Step 11: Recycle the process from time to time using updated datasets to keep learning and adapting to new fraud trends.

VI. GRAPH



The figure shows a Receiver Operating Characteristic (ROC) curve, which is a measure of a binary class model's performance. The plot shows the True Positive Rate on the y-axis versus the False Positive Rate on the x-axis.

This is what the graph illustrates:

The orange ROC curve clings to the top border of the graph, showing a perfect classifier with a perfect area under the curve (AUC) of 1.00.

The blue dashed line is a case of a zero discrimination power model—a random classifier, in effect.

Categories are:

Title: "Receiver Operating Characteristic"

Legend: "ROC curve (area = 1.00)"

X-axis: "False Positive Rate"

Y-axis: "True Positive Rate"

The plot is an idealized model, which is seldom seen in real applications. For the purpose of fraud detection or other classification problems, the ROC curve is a measure to assess the effectiveness of the model in distinguishing positive from negative cases.

VII. CONCLUSION

This paper proposes a comprehensive, acute fraudulent healthcare insurance claims detection system based on strong machine learning algorithms and extreme data balancing algorithms. Highly prevalent practice of medical fraud gravely endangers the sustainability of healthcare systems, weighing increasingly policyholders and insurers with rising costs in

DOI: 10.48175/IJARSCT-26247

terms of increasing premiums. Traditional human-based and rule-based approaches did not live up to the challenge of identifying complex, dynamic fraud patterns in heavily unbalanced and high-dimensional datasets.

The system proposed in this work effectively addresses all these issues through a systematic approach involving data collection, data preprocessing, feature extraction, data balancing, model training, and fraud prediction. Data is collected from various healthcare sources to ensure that the system is able to process actual claim scenarios. The preprocessing module maintains data consistency, and the feature engineering stage emphasizes the most significant fraud indicators.

One of the main highlights of this research is the application of hybrid data balancing strategies, specifically the combination of SMOTE with undersampling techniques, which addresses the problem of class imbalance—one of the most severe limitations in fraud detection models.This method significantly improves the sensitivity of the model toward identifying rare fraudulent patterns and prevents majority class bias.

For the task of classification, several models like Logistic Regression, Decision Tree, Random Forest, and XGBoost were trained and tested. Ensemble methods like Random Forest and XGBoost gave better performance as they can learn complicated non-linear data patterns without sacrificing much accuracy and interpretability. The system's feature of continuous fraud detection allows for real-time processing of incoming claims, auto-flagging and alerting suspicious activity for closer investigation by hand.

Besides, the user interface offers actionable insight and transparency so that auditors and health care analysts are able to see model predictions and adjust system parameters appropriately. Not only does the system enhance fraud detection effectiveness, but also operational efficiency, with the claims automatically being validated.

Overall, the project demonstrates the reason why a machine learning supported hybrid data balancing is an exciting method of combating medical insurance fraud. It is a fast, powerful, and adaptable option that is completely capable of protecting against traditional system vulnerabilities. In doing so, it is also playing a significant role in long-term financial integrity, reduced attempts at fraud, and enhanced credibility and trust across the health insurance sector.

REFERENCES

[1]. Matloob, I., Khan, S., ur Rahman, H., & Hussain, F. (2020). *Medical Health Benefit Management System for Real-Time Notification of Fraud Using Historical Medical Records*. Applied Sciences, 10(15), 5144. https://doi.org/10.3390/app10155144

[2]. Zhang, C., Xiao, X., & Wu, C. (2020). *Medical Fraud and Abuse Detection System Based on Machine Learning*. International Journal of Environmental Research and Public Health, 17(19), 7265. https://doi.org/10.3390/ijerph17197265

[3]. Zhang, D., Bhandari, B., & Black, D. (2020). *Credit Card Fraud Detection Using Weighted Support Vector Machine*. Applied Mathematics, 11(12), 1275–1291. https://doi.org/10.4236/am.2020.1112087

[4]. L. Bhavya, V. Sasidhar Reddy, U. Anjali Mohan, & S. Karishma. (2020). *Credit Card Fraud Detection using Classification, Unsupervised, Neural Networks Models*. International Journal of Engineering Research, 9(04). https://doi.org/10.17577/ijertv9is040749

[5]. Kolodiziev, O., Mints, A., Sidelov, P., Pleskun, I., & Lozynska, O. (2020). *Automatic machine learning algorithms for fraud detection in digital payment systems*. Eastern-European Journal of Enterprise Technologies, 5(9), 14–26. https://doi.org/10.15587/1729-4061.2020.212830

[6]. Sedik, A., Iliyasu, A. M., Abd El-Rahiem, B., Abdel Samea, M. E., Abdel-Raheem, A., Hammad, M., Peng, J., Abd El-Samie, F. E., & Abd El-Latif, A. A. (2020). *Deploying Machine and Deep Learning Models for Efficient Data-Augmented Detection of COVID-19 Infections*. Viruses, 12(7), 769. https://doi.org/10.3390/v12070769

[7]. Huang, J. (2020, published in 2019 conference). *Credit Card Transaction Fraud Using Machine Learning Algorithms*. Proceedings of ICESED 2019. https://doi.org/10.2991/icesed-19.2020.14

[8]. S P Maniraj, Aditya Saini, Shadab Ahmed, & Swarna Deep Sarkar. (2019). *Credit Card Fraud Detection using Machine Learning and Data Science*. International Journal of Engineering Research, 08(09). https://doi.org/10.17577/ijertv8is090031

[9]. (2019). *Solve fraud detection problem by using graph based learning methods*. Journal of Engineering and Science Research, 3(4), 28–31. https://doi.org/10.26666/rmp.jesr.2019.4.6

[10]. Lucas, Y., Portier, P.-E., Laporte, L., Calabretto, S., Caelen, O., He-Guelton, L., & Granitzer, M. (2019). *Multiple perspectives HMM-based feature engineering for credit card fraud detection*. Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing, 1359–1361. https://doi.org/10.1145/3297280.3297586

[11]. Verma, L. (2018). *Machine Learning Methods in Early Diagnosis of Coronary Artery Disease*. Journal of Cardiology & Cardiovascular Therapy, 11(1). https://doi.org/10.19080/jocct.2018.11.555801

[12]. Arora, M., Dhawan, S., & Singh, K. (2018). *Data Driven Prognosis of Cervical Cancer Using Class Balancing and Machine Learning Techniques*. EAI Endorsed Transactions on Energy Web. https://doi.org/10.4108/eai.13-7-2018.164264

[13]. Kavitha, M., & Suriakala, M. (2015). *Fraud Detection in Current Scenario, Sophistications and Directions: A Comprehensive Survey*. International Journal of Computer Applications, 111(5), 35–40. https://doi.org/10.5120/19538-1194

[14]. JacqulinMargret, J., & Sreenivasan, S. (2013). *Implementation of Data Mining in Medical Fraud Detection*. International Journal of Computer Applications, 69(5), 1–4. https://doi.org/10.5120/11835-7556