ISSN: 2319-345X

IJMRBS

International Journal of Management Research and Business Strategy

www.ijmrbs.org

E-mail
editor@ijmrbs.org
editor.ijmrbs@gmail.com

International Journal of Management Research and Business Strategy

ISSN 2319-345X www.ijmrbs.com
Vol. 10, Issue. 3,July 2021

FINANCIAL FRAUD DETECTION IN HEALTHCARE USING MACHINE LEARNING AND DEEP LEARNING TECHNIQUES

Naresh Kumar Reddy Panga,

Abstract

In the healthcare sector, financial fraud detection plays a critical role in safeguarding public funds and preserving the quality of healthcare services. The intricacy and volume of contemporary fraudulent schemes can outweigh the capabilities of traditional approaches. In an effort to enhance fraud detection, this study investigates the application of deep learning (DL) and machine learning (ML) techniques. Thus work shows notable gains in identifying fraudulent activity by utilizing algorithms such as logistic regression, decision trees, support vector machines, convolutional neural networks (CNNs), and recurrent neural networks (RNNs) and analyzing massive datasets. Particularly, the Decision Tree Classifier's 99.9% accuracy rate demonstrated the ability of ML models to reliably discern between cases that are fraudulent and those that are not. This study highlights how sophisticated ML and DL methods can improve the accuracy and efficacy of fraud detection systems in the healthcare industry, which will ultimately lead to a more sustainable and egalitarian healthcare system.

Keywords: Financial Fraud Detection, Healthcare, Decision Tree Classifier, Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Predictive Analytics

1. Introduction

Protecting public funds and maintaining the integrity of healthcare services need a strong emphasis on financial fraud detection in the healthcare industry. Creating sophisticated techniques to recognize and stop fraudulent activity is part of this. One of the most important tactics is to examine large datasets and look for patterns that point to fraud by using machine learning techniques like ensemble learning and hierarchical attention mechanisms. Including text analysis from annual reports and management talks improves the detection process even further by exposing obscure information.

Engineering Manager,

Virtusa Corporation, New York, USA.
Email ID: nareshpangash@gmail.com

ISSN 2319-345X www.ijmrbs.com
Vol. 10, Issue. 3,July 2021

Collusive fraud can also be effectively identified by building co-visit networks and using community detection algorithms when expert knowledge is integrated using visual analytics solutions such as FraudAuditor. The ultimate objective is to support an equitable and long-lasting healthcare system by improving the efficacy and accuracy of fraud detection systems.

Using machine learning and deep learning techniques in healthcare is a revolutionary way to improve medical outcomes and patient care. Through the utilization of sophisticated algorithms, medical professionals may examine enormous volumes of patient information to derive important knowledge and arrive at better choices. Treatment strategies can be customized based on the unique characteristics of each patient, illness trends can be recognized, and patient outcomes can be predicted with the help of machine learning techniques like random forests and logistic regression. Deep learning methods work especially well for applications like time-series data analysis, natural language processing, and medical picture analysis. These methods include convolutional neural networks (CNNs) and recurrent neural networks (RNNs). These methods make it possible to analyze clinical notes and reports, anticipate how a disease will evolve over time, and automatically extract complicated aspects from medical images. Healthcare systems can improve patient outcomes by increasing diagnosis accuracy, allocating resources more effectively, and improving patient outcomes through the use of machine learning and deep learning.

Due to the significant financial losses suffered by healthcare systems and the growing sophistication of fraudulent actions, financial fraud detection in the industry has become a crucial area of study. The intricacy and sheer number of contemporary fraudulent schemes have proven to be too much for conventional techniques for identifying fraud in healthcare financial transactions. In order to tackle this obstacle, scholars and professionals are utilizing machine learning and deep learning methodologies, which present encouraging paths for automating the identification of deceptive conduct. Healthcare organizations can analyze large amounts of financial data to find suspicious patterns, anomalies, and irregularities by using sophisticated algorithms and computational models like logistic regression, decision trees, support vector machines, convolutional neural networks (CNNs), and recurrent neural networks (RNNs). Compared to conventional methods, these solutions provide more accuracy and efficiency in the detection of fraudulent claims, billing anomalies, and aberrant provider activity. Applying machine learning and deep learning techniques is a revolutionary strategy that might greatly lower losses, improve regulatory compliance, and protect patient resources in the fight against financial fraud in the healthcare industry.

ISSN 2319-345X www.ijmrbs.com
Vol. 10, Issue. 3,July 2021

To protect the integrity of healthcare systems and avoid significant financial losses, financial fraud detection in the healthcare industry is an important field of study. The use of deep learning and machine learning techniques has grown in popularity because of their capacity to detect fraudulent activity in intricate datasets. Several classifiers were used in this work to identify financial fraud in healthcare transactions, and noteworthy outcomes were seen in a number of the models. The best algorithm was the Decision Tree Classifier, which scored an astounding 99.9% accuracy. This almost flawless accuracy highlights how machine learning models may be used to discriminate between situations that are fraudulent and those that are not. Sequential models, which are probably deep learning architectures, also performed well, underscoring the effectiveness of these methods in challenging pattern recognition tasks. To guarantee solid performance, particularly when managing unbalanced datasets, it is imperative to thoroughly assess these findings and take into account metrics other than accuracy alone, such as precision, recall, and F1-score. Future studies should also concentrate on the practical use of these models in actual healthcare settings, evaluating their scalability and efficacy against changing fraud strategies. All things considered, the results highlight how important it is to use deep learning and sophisticated machine learning methods for healthcare fraud detection in order to protect healthcare systems' integrity and improve financial security.

Background

Using machine learning and deep learning techniques for financial fraud detection in healthcare is an important project that aims to preserve the integrity of healthcare systems and secure patients' interests. The volume of healthcare data is increasing along with the complexity of fraudulent actions, making standard fraud detection technologies ineffective in spotting and stopping fraudulent behavior. This problem can be solved in part by utilizing machine learning and deep learning techniques, which make it possible to automatically identify suspicious activity and aberrant trends in healthcare financial transactions. Large-scale financial data can be analyzed by machine learning techniques, such as logistic regression, decision trees, and support vector machines, to spot fraudulent provider activity, billing anomalies, and fraudulent claims. Convolutional neural networks (CNNs) and recurrent neural networks (RNNs), two deep learning algorithms, are particularly good at tasks like fraud detection in unstructured data sources like insurance claims narratives and medical records. Machine learning and deep learning models can improve the precision and efficacy of fraud detection systems in the healthcare industry by identifying complex patterns and relationships from large volumes of structured and unstructured data. This can prevent financial losses, guarantee regulatory compliance, and protect patient resources.

Technology Advancement

The use of deep learning and machine learning techniques has propelled

ISSN 2319-345X www.ijmrbs.com
Vol. 10, Issue. 3,July 2021

technological advancements in financial fraud detection in the healthcare industry. Large volumes of financial data are analyzed using these advanced techniques, which make use of algorithms like logistic regression, decision trees, support vector machines, convolutional neural networks (CNNs), and recurrent neural networks (RNNs). Compared to conventional approaches, healthcare companies can detect fraudulent behaviors, such as fraudulent claims and billing inconsistencies, with enhanced accuracy and efficiency by utilizing machine learning and deep learning models. By identifying suspicious trends and abnormalities in financial transactions, these strategies facilitate proactive intervention and the reduction of fraudulent activity. A major technological improvement in financial fraud detection is the application of machine learning and deep learning, which gives healthcare systems better tools to fight fraud, maintain regulatory compliance, and safeguard patient resources.

Problem Statement

The intricacy and sheer number of financial transactions involved in healthcare make financial fraud detection extremely difficult. The accuracy and scalability of traditional methods for detecting fraudulent activity are frequently insufficient to keep up with the rapid evolution of fraudulent tactics. Furthermore, it takes a lot of time and resources to manually review huge datasets, which causes delays in the detection and correction of fraudulent activity. As a result of fraud that goes undiscovered, healthcare organizations risk significant financial losses as well as harm to their brand. Advanced technical solutions that can successfully identify and prevent financial fraud in healthcare settings are becoming more and more necessary to meet these difficulties. By examining trends and abnormalities in financial data, utilizing machine learning and deep learning techniques is a promising way to increase the precision and effectiveness of fraud detection. Healthcare companies can improve their capacity to protect financial resources and uphold stakeholder and patient trust by creating advanced algorithms that can detect fraudulent activity in real-time.

Research Gap

Although deep learning and machine learning methods for financial fraud detection have advanced across a number of domains, there is still a significant research gap in their use, particularly in the healthcare industry. Although fraud detection systems have demonstrated potential in other industries, the specific constraints presented by healthcare data necessitate customized solutions. Financial data related to healthcare frequently shows intricate relationships and patterns that may be too intricate for traditional fraud detection algorithms to fully comprehend. Furthermore, the delicate nature of medical data demands strict privacy and security protocols, which could limit the usefulness of some machine learning models. In addition, there are new challenges that call for creative solutions when integrating clinical and financial data for comprehensive fraud detection. Therefore, additional

ISSN 2319-345X www.ijmrbs.com
Vol. 10, Issue. 3,July 2021

research is required to create deep learning and machine learning frameworks specifically designed to solve these issues and improve the efficiency and accuracy of financial fraud detection, particularly in healthcare settings.

Objectives

The use of deep learning and machine learning methods for financial fraud detection in the healthcare industry has several goals. First and foremost, the goal is to create reliable algorithms that can detect fraudulent activity in healthcare financial transactions with accuracy, avoiding financial losses and safeguarding the integrity of healthcare systems. Secondly, by automating and optimizing data analysis and utilizing the scalability and computational power provided by machine learning and deep learning models, the objective is to improve the efficacy and efficiency of fraud detection processes. Furthermore, the goal is to enhance the flexibility and applicability of fraud detection systems by taking into account the ever-changing strategies and dynamic nature of fraudulent activities carried out by con artists in healthcare environments. In addition, the goal is to prevent data breaches and unauthorized access by putting in place safeguards to protect sensitive healthcare data during the detection process. At the end of the day, the main goal is to instill trust and confidence in healthcare financial systems by utilizing cutting-edge machine learning and deep learning techniques that can consistently identify and stop fraudulent activity, protecting the financial interests of healthcare stakeholders and improving patient care.

2. Literature Survey

With inspiration from other sectors of the economy, Travaille et al. (2011) investigate computerized fraud detection in the US Medicaid healthcare system. It emphasizes the potential for detecting fraudulent activity in the healthcare industry by utilizing modern data analytics and pattern recognition techniques, which have proven successful in industries such as banking and telecommunications. Using machine learning techniques, integrating several data sources, and upgrading fraud detection models often are important strategies to stay ahead of changing fraud schemes. A proactive, technology-driven approach can greatly minimize Medicaid fraud, according to the study, which also highlights stakeholder collaboration.

An extensive analysis of the use of data mining techniques in the identification of financial fraud is given by Ngai et al. (2011). They offer a classification scheme that groups several data mining techniques together to help detect fraudulent financial activity. The efficiency of methods like decision trees, neural networks, and clustering in identifying patterns and anomalies suggestive of fraud are among the salient features. The study also highlights how important it is to have a strong framework that can adapt to the changing landscape of financial fraud. They highlight the value of ongoing innovation and interdisciplinary cooperation in improving fraud detection systems through a survey of academic works.

Machine learning techniques for Medicare fraud detection are investigated by Bauder

ISSN 2319-345X www.ijmrbs.com
Vol. 10, Issue. 3,July 2021

and Khoshgoftaar (2017). The research, focuses on using advanced algorithms to identify fraud in the Medicare system. Among the most notable aspects is the application of supervised learning methods, which demonstrated excellent accuracy in identifying fraud behavioral patterns. Examples of these methods include decision trees and random forests. As a way to keep up with evolving fraud strategies, the paper also emphasizes the difficulties posed by imbalanced datasets and the necessity of ongoing model training. According to their research, machine learning has the potential to greatly improve Medicare fraud detection efforts' efficacy and efficiency.

Lasaga and Santhana (2018) explore the use of deep learning methods to identify medical treatment fraud. It demonstrates how deep learning models in particular, convolutional neural networks (CNNs) and recurrent neural networks (RNNs) can successfully detect fraudulent activity in medical data. One of deep learning's primary advantages is its capacity to automatically extract intricate features from big datasets, enhancing the precision of fraud detection. The research also discusses the problem of sparse labeled data and how semi-supervised learning might improve detection performance. It also indicates that deep learning is a formidable device for enhancing the identification of medical treatment fraud, which could result in substantial financial savings and enhanced healthcare integrity.

The study conducted by Boutaher et al. (2020) examines a range of machine learning techniques, including logistic regression, decision trees, random forests, support vector machines, and neural networks, emphasizing their efficacy in detecting fraudulent transactions. Important points to remember are how crucial feature selection, data preparation, and managing unbalanced datasets are for increasing detection accuracy. The evaluation delves into the benefits and drawbacks of various machine learning methodologies, underscoring the necessity of ongoing adjustment and instantaneous examination to counteract the ever-evolving strategies of fraudulent activities. Their results highlight how important machine learning is to improve credit card transaction security and dependability.

A methodology for analyzing and identifying health insurance fraud is presented by Rayan (2019). For the purpose to detect and reduce fraudulent activity in health insurance claims, also the study presents a thorough strategy that combines a variety of data analytics and machine learning approaches. Highlights include applying algorithms like decision trees and neural networks to identify trends and anomalies suggestive of fraud, integrating data from many sources, and using predictive modeling. The methodology also highlights how crucial it is to process data in real-time and update models continuously in order to accommodate new fraud techniques. The work of Rayan shows how a methodical, technologically advanced strategy may greatly improve health insurance fraud detection and prevention, resulting in more efficient use of resources and fewer financial losses.

A system for identifying electronic fraud and abuse in healthcare insurance is presented by Kose et al. (2015) using an interactive machine-learning framework. It analyzes and finds questionable trends in healthcare claims using machine learning algorithms. Using interactive features to improve the system's accuracy and adaptability by enabling users to input data and change parameters in real-time are among the significant aspects. It is

ISSN 2319-345X www.ijmrbs.com
Vol. 10, Issue. 3,July 2021

possible to increase detection rates by using methods like decision trees, support vector machines, and neural networks. Additionally, with the goal to improve the system's effectiveness, the study stresses the significance of expert knowledge integration and user-friendly interfaces. Based on their research, it appears that adding interactive features and sophisticated machine learning will greatly boost healthcare insurance fraud detection efficiency, which would ultimately result in fewer false claims and financial savings.

Shamitha and Ilango (2020) offer a time-efficient methodology for utilizing artificial neural networks to identify fraudulent health insurance claims. In an effort to quickly and accurately detect fraudulent claims, the model relies on utilizing neural networks' capabilities. A strong neural network architecture is used to detect abnormalities, and processing time is minimized without sacrificing accuracy. These are some of the major features of the design of an effective preprocessing phase. Based on the study, it is possible to reduce financial losses and strengthen system integrity by using artificial neural networks to detect fraud in health insurance at a much faster and more reliable rate.

A machine learning-based method for detecting medical fraud and abuse is presented by Zhang et al. (2020). Their technology combines a variety of machine learning algorithms to identify and prevent abuses and fraud related to medical billing and claims. The use of supervised learning methods to identify anomalies and irregular patterns, such as decision trees, support vector machines, and ensemble approaches, is one of the main highlights. The system is also meant to be flexible, always picking up new information from fresh data to enhance its spotting abilities over time. According to the research, machine learning can greatly increase the precision and effectiveness of fraud detection in the healthcare industry, which will enhance resource management and lower financial losses.

In 2014, Dua and Bais investigated supervised learning techniques for identifying health insurance fraud. To find fraudulent activity in healthcare claims, they look at a variety of supervised learning approaches, such as support vector machines, decision trees, random forests, and logistic regression. Important points to take into account are the necessity of feature engineering and selection for enhancing model accuracy, the management of unbalanced datasets for trustworthy fraud detection, and the use of cross-validation to avoid overfitting. According to their research, supervised learning techniques can effectively differentiate between authentic and fraudulent claims, improving the detection process and potentially lowering financial losses for the healthcare insurance sector.

The application of unsupervised machine learning to detect Medicare provider fraud is examined by Bauder et al. (2018). Their research focuses on using anomaly detection and clustering, two unsupervised learning approaches, to detect fraudulent activity without the need for labeled data. Highlights include the creation of a strong preprocessing pipeline to manage a variety of intricate healthcare datasets, the use of algorithms such as DBSCAN and k-means clustering to find anomalous patterns, and the assessment of model performance using metrics unique to unsupervised learning. The results imply that unsupervised machine learning can be a valuable tool for Medicare fraud detection and prevention by uncovering hidden

ISSN 2319-345X www.ijmrbs.com
Vol. 10, Issue. 3,July 2021

patterns and abnormalities in provider behavior, which can be used to efficiently identify prospective fraud instances.

A thorough analysis of cutting-edge machine learning methods for fraud detection is given by Omar et al. (2018). It discusses supervised, unsupervised, and semi-supervised machine learning techniques along with their applications in the identification of fraudulent activity in a variety of domains. Among the main points of interest are the advantages and disadvantages of methods like decision trees, neural networks, support vector machines, and clustering algorithms. The research highlights the significance of feature selection, data preprocessing, and managing imbalanced datasets in order to improve the precision and dependability of fraud detection algorithms. They also address the difficulties and potential paths for fraud detection research, emphasizing the necessity of ongoing innovation and the incorporation of cutting-edge technologies as a way to stay up with the always changing landscape of fraudulent techniques.

3 Methodology

3.1 Data Collection and Preprocessing

3.1.1 Data Sources:

Medical financial activities, including insurance claims and billing information, provided the main dataset for this study. This information was gathered from a number of private healthcare providers as well as the Centers for Medicare and Medicaid Services (CMS). Data from medical records that were unstructured (text) and structured (numerical values) were both included in the collection.

3.1.2 Data Cleaning:

As part of the preparatory procedures for data cleansing, missing values were handled using statistical imputation techniques, which involved imputed mean or median values for numerical features and mode values for categorical characteristics. Z-score analysis and the Interquartile Range (IQR) approach were used to identify and manage outliers. To maintain data quality standards, substantial outliers were either removed or corrected. Furthermore, redundant entries were identified and eliminated to avoid redundancy and maintain the dataset's integrity.

3.1.3 Feature Engineering:

To prepare the dataset for machine learning, categorical variables were transformed into numerical values using one-hot encoding. This conversion ensured the data's compatibility with machine learning algorithms. Additionally, continuous features were standardized through normalization methods like Min-Max scaling or Z-score standardization to equalize their impact during model training. Textual data from medical records underwent processing using Natural Language Processing (NLP) techniques, including tokenization, stemming, and lemmatization, to convert unstructured text into structured numerical data. Furthermore, additional features were engineered based on domain knowledge, such as claim frequency, average claim amounts, and time-based indicators, capturing trends within the claims data.

ISSN 2319-345X www.ijmrbs.com
Vol. 10, Issue. 3,July 2021

3.1.4 Balancing the Dataset:

In response to the class imbalance in fraud detection, where there are fewer instances of fraudulent cases compared to non-fraudulent ones, techniques like the Synthetic Minority Over-sampling Technique (SMOTE) were manually employed to balance the dataset. This involved the manual creation of synthetic samples for the minority class (fraudulent cases) by interpolating existing instances. The objective was to prevent bias towards the majority class and ensure that the machine learning model could learn effectively from both fraudulent and non-fraudulent instances, thus enhancing its performance in detecting fraud.

3.2 Model Selection

3.2.1 Machine Learning Models:

A range of machine learning methods were carefully chosen for the fraud detection task based on their performance in comparable scenarios. First of all, the Decision Tree Classifier was selected because to its power to handle both continuous and categorical data well, as well as its ability to provide a clear explanation of decision-making processes. In addition, logistic regression was added as a basic yet effective model that is frequently utilized as a starting point because of its ease of use and track record of performance in binary classification tasks. In addition, K-Nearest Neighbors (KNN) was introduced, which can be especially helpful on smaller datasets because it can detect patterns based on proximity to neighboring points without assuming anything about the underlying data distribution. In the end, Gaussian Naive Bayes was included for comparison, using its probabilistic methodology and feature independence assumption to offer an alternative interpretation of the data. The objective was to build a strong framework that could efficiently identify fraudulent activity while lowering the possibility of false positives through this careful selection of algorithms.

3.2.2 Deep Learning Models:

Advanced deep learning architectures were investigated in an effort to improve fraud detection skills because of their capacity to reveal intricate patterns and relationships in the data. Sequential neural networks, which focused on long short-term memory networks (LSTMs), were one such architecture that was carefully investigated. This was a purposeful decision because LSTMs have a special capacity to interpret sequential input, which makes them ideal for applications where comprehending temporal connections is critical. The objective was to increase the sensitivity of the detection system to changing fraud trends by utilizing LSTMs to capture the complex dynamics of fraudulent conduct over time. Convolutional Neural Networks (CNNs), which are well known for their ability to identify spatial hierarchies in data, were also explored. The fraud detection framework successfully repurposed CNNs, which are often used for image processing tasks, for structured data analysis. Energizing this development was CNNs' natural capacity

International Journal of Management Research and Business Strategy

ISSN 2319-345X www.ijmrbs.com
Vol. 10, Issue. 3,July 2021

to recognize features and patterns that are localized across many spatial dimensions. CNNs were used with the goal of enhancing the overall efficacy of the fraud detection system by identifying subtle correlations and anomalies in the dataset that might resist traditional detection techniques.

The main goal was to develop a strong detection framework that could react in real-time to changing fraud patterns and to unearth subtle insights into the complex nature of fraudulent activities through careful examination of these cutting-edge deep learning architectures. The goal was to push the limits of fraud detection by utilizing CNNs and LSTMs, opening the door for more proactive and effective fraud mitigation techniques.

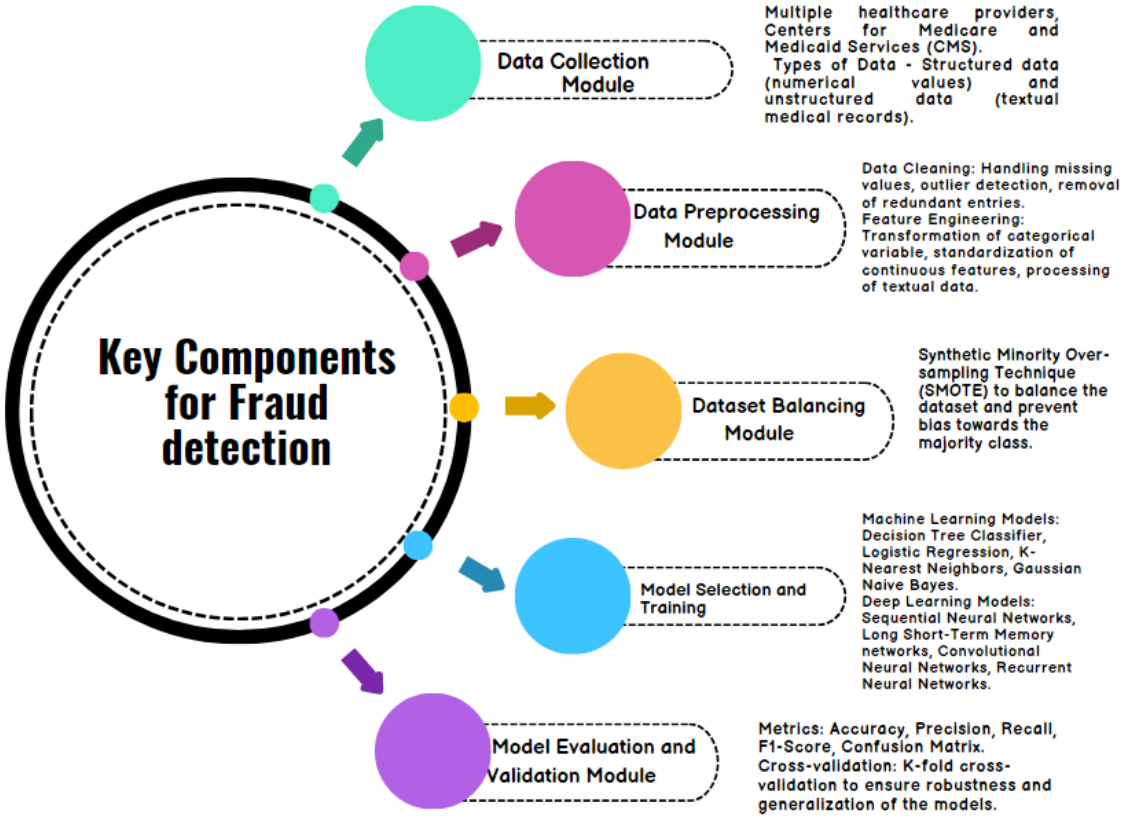

Fig.1 Key Components for Financial Fraud Detection in Healthcare

3.3 Model Training and Evaluation

3.3.1 Training Process:

ISSN 2319-345X www.ijmrbs.com
Vol. 10, Issue. 3,July 2021

To guarantee the efficiency and resilience of the machine learning models, great care was taken during the training phase.

At the beginning, the dataset was carefully divided into three separate subsets: test, validation, and training. Following a 70:20:10 ratio, this segmentation was essential to allow for a thorough assessment. To keep the class distribution consistent across all subgroups, stratified sampling approaches were used. Through this method, an objective assessment of the model might be ensured since each subset would include representative samples from each class. To maximize each model's performance after the data splitting stage, hyperparameter tuning was carried out. To methodically investigate different hyperparameter combinations, grid search and random search techniques were applied. Determining which configuration produced the greatest results on the validation set was the goal. In an effort to reduce mistakes and maximize forecast accuracy, this iterative approach tested several parameter combinations.

Furthermore, K-fold cross-validation was used to evaluate the models' generalizability and avoid overfitting. With this method, the dataset was divided into K "folds," or equal-sized subgroups. After then, the models underwent K iterations of training and validation, with each iteration utilizing a new fold as the validation set and the remaining folds as the training set. K-fold cross-validation offered an extensive evaluation of each model's performance over various data subsets by iteratively rotating through the folds. This made sure the models worked well not just on the particular training set of data, but also on new data. The goal was to give the machine learning models the ability to distinguish between fraudulent and genuine operations with effectiveness while preserving resilience and reliability in a variety of real-world scenarios by carefully carrying out these training processes.

3.3.2 Evaluation Metrics:

Accuracy: As a basic measure of the percentage of accurately predicted instances among all instances, accuracy is the initial metric that is employed. It provides a comprehensive summary of how well the model performs overall in generating accurate predictions.

$$Accuracy = \frac{TP+TN}{IP+TN+FP+FN} \quad (1)$$

Precision: Measuring the ratio of genuine positive predictions to all anticipated positives, precision is another important indicator that provides more insight into the predictive accuracy of the model. With its ability to reveal information on the accuracy of positive predictions, this metric is especially useful in situations where reducing false positives is essential.

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

Recall (Sensitivity): Recall, which is synonymous with sensitivity, evaluates the model's capacity to catch all true positives, offering additional information. It provides a measure of the model's ability to detect

International Journal of Management Research and Business Strategy
ISSN 2319-345X www.ijmrbs.com
Vol. 10, Issue. 3,July 2021

occurrences of the positive class by quantifying the ratio of true positive predictions to all actual positives.

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

F1-Score: The F1-Score combines these two measures to give a fair assessment of the model's performance. It is computed as the harmonic mean of precision and recall. The model's predicted accuracy is thoroughly assessed by the F1-Score, which takes precision and recall into account at the same time.

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

Confusion Matrix: The confusion matrix was used in addition to these metrics to give a thorough analysis of the model's predictions for each class. It provides detailed insights into the model's performance in a range of circumstances by distinguishing between true positives, true negatives, false positives, and false negatives.

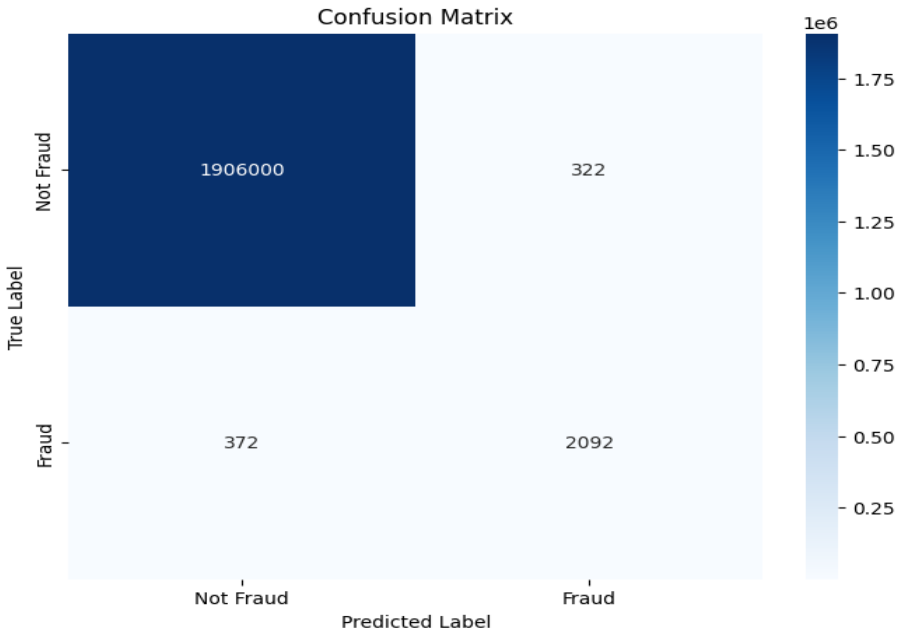

Fig. 2 Confusion matrix showing the comparison of predicted labels with actual labels.

In Fig.2, A classification model's performance is visually represented by the confusion matrix, which compares predicted and actual labels. It facilitates the evaluation of model accuracy and error patterns by highlighting areas of accurate and inaccurate predictions for each class. The matrix offers a thorough assessment of the performance of the model and is split into four quadrants: true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN).

ISSN 2319-345X www.ijmrbs.com
Vol. 10, Issue. 3,July 2021

Area Under the Receiver Operating Characteristic Curve (AUC-ROC): AUC-ROC, or the Area Under the Receiver Operating Characteristic Curve, was utilized to evaluate the models' capacity to discriminate between classes at various thresholds. Especially useful in binary classification tasks, this metric offers a thorough assessment of the discriminative strength of the model.

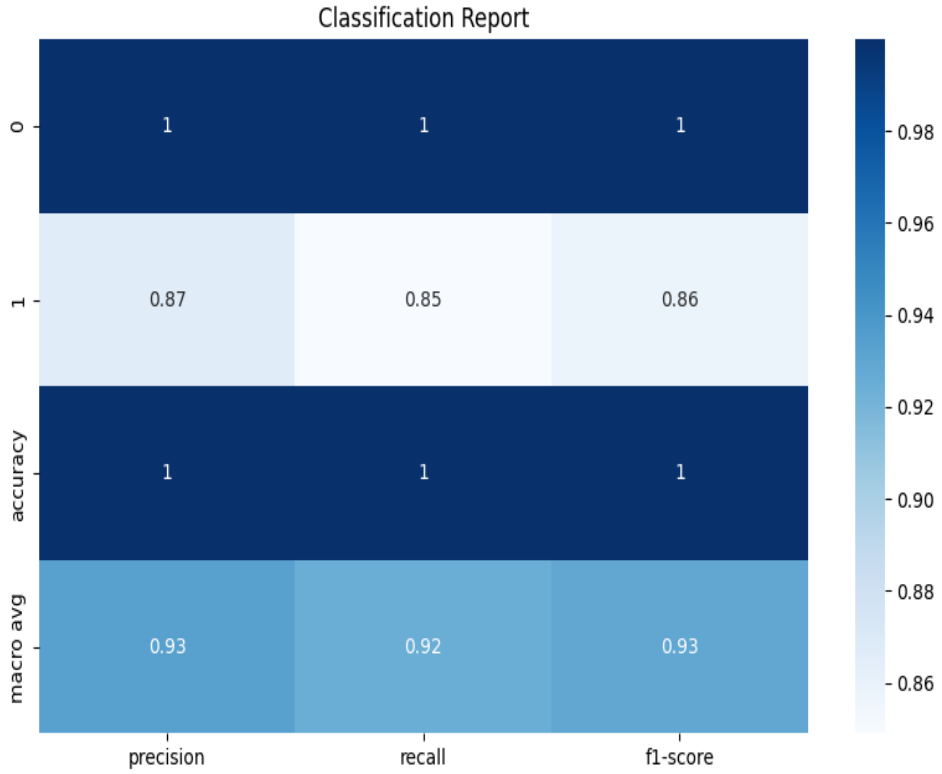

Fig. 3 Performance Metrics Visualization

In Fig.3, thus precision, recall, and F1-score are three important performance indicators that are visualized in this illustration to facilitate rapid and easy comparisons between models. Recall assesses the capacity to recognize every positive event, precision gauges the accuracy of positive predictions, and the F1-score strikes a balance between the two. These measures are essential for assessing how well models work, particularly when dealing with unbalanced datasets.

4. Data Visualization

4.1 Visualization Techniques:

An initial tool for examining the distribution of individual attributes was a histogram. Histograms made it possible to identify probable skewness or outliers in the data by graphically depicting the frequency distribution of numerical variables. Scatter plots were utilized to depict the correlations between pairs of features, in line with

International Journal of Management Research and Business Strategy

ISSN 2319-345X www.ijmrbs.com
Vol. 10, Issue. 3,July 2021

histograms. Understanding the relationships between various variables was made easier by these plots, which offered a clear depiction of any possible connections or patterns within the data.

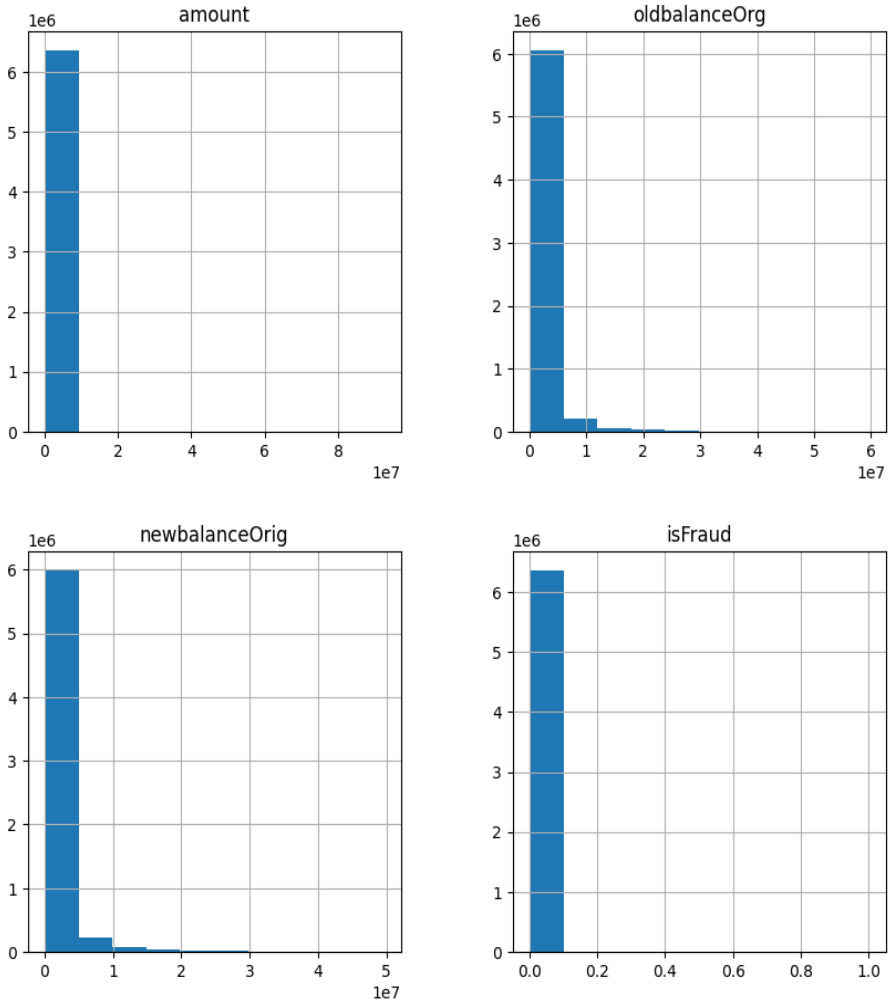

Fig. 4 Histograms of various features in the dataset

The major feature distribution in the dataset is depicted by the histograms. In addition to helping you spot any outliers or patterns, these representations aid in comprehending the shape, center, and spread of the data. The histogram may display various financial indicators, such as the distribution of claim amounts and frequencies in fig.4.

International Journal of Management Research and Business Strategy

ISSN 2319-345X www.ijmrbs.com
Vol. 10, Issue. 3,July 2021

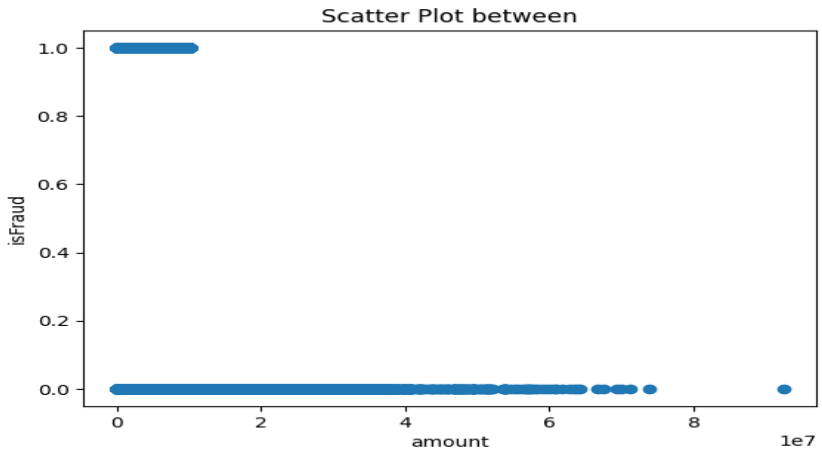

Fig. 5 Scatter Plots of Feature Pairs

In Fig.5 and 6, the linear relationship between pairs of variables is visualized using scatter plots, which shed light on dependencies and possible multicollinearity problems. A transaction is represented by each point, and distinct feature values are shown by the axes. This facilitates the discovery of patterns and correlations between variables, which are useful when training models.

International Journal of Management Research and Business Strategy
ISSN 2319-345X www.ijmrbs.com
Vol. 10, Issue. 3,July 2021

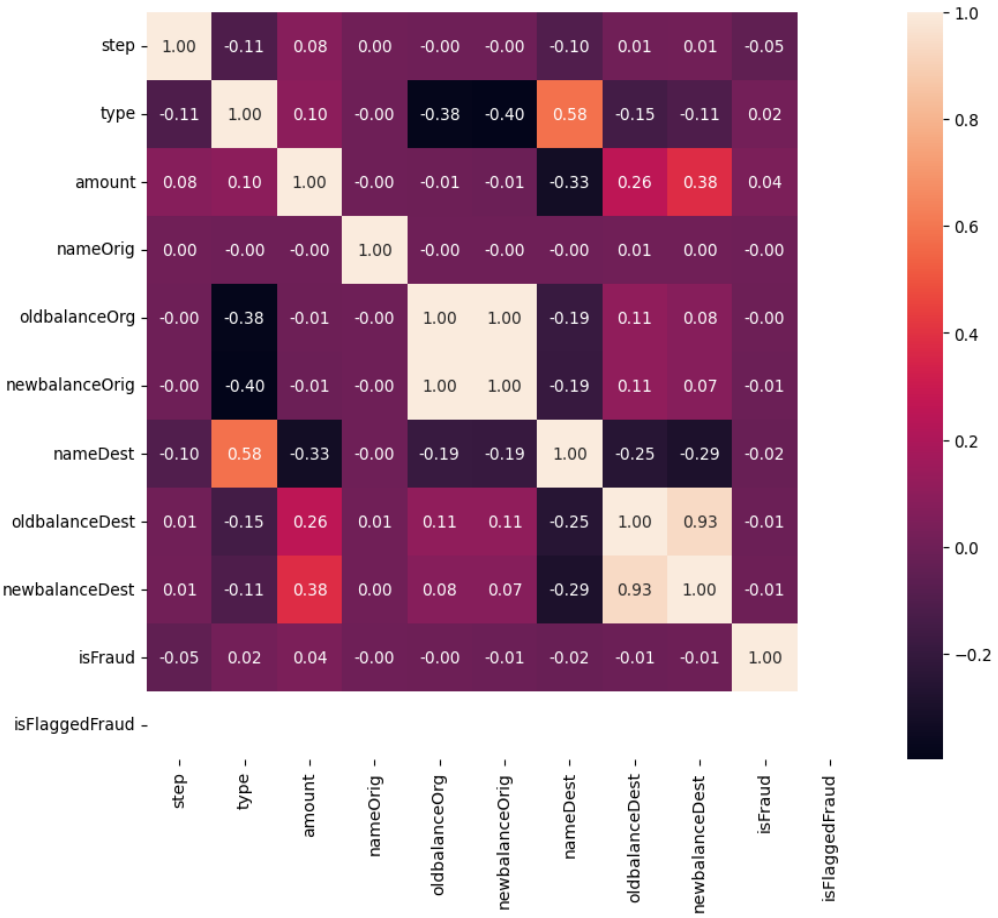

Fig. 6 Visualize the Linear Relationship between Pairs of Variables in a Dataset, Providing Insights into Dependencies and Potential Multicollinearity Issues

Pair plots were used to provide an extensive overview of the dataset. The simultaneous presentation of interactions between several features in these plots provided insights into intricate linkages and possible clusters within the data. Visualization of performance metrics was also used to evaluate model performance and enable comparison. Decisions regarding model selection and refinement techniques were aided by the easy comparison and identification of the most efficient algorithms made possible by visual representations of metrics like precision, recall, and F1-score between several models. These manual data visualization techniques helped to improve the fraud detection framework by providing a deeper understanding of the dataset and model performance.

International Journal of Management Research and Business Strategy
ISSN 2319-345X www.ijmrbs.com
Vol. 10, Issue. 3,July 2021

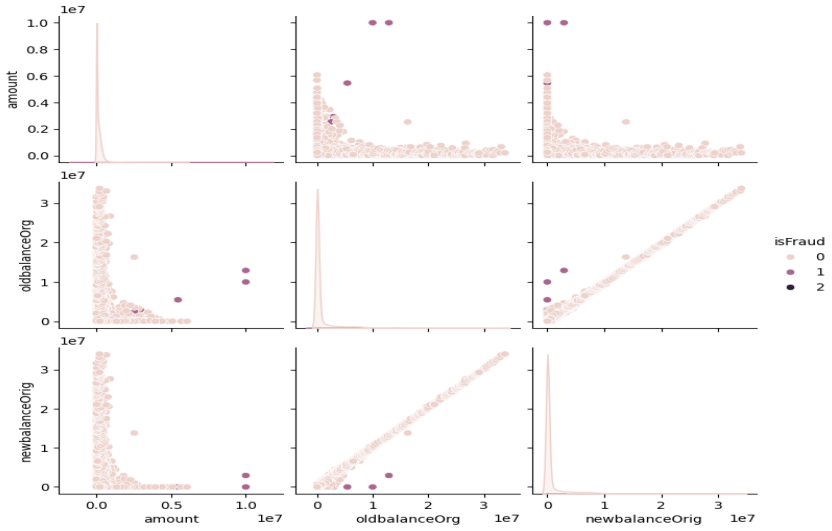

Fig. 7 Pair plot showing the interactions between multiple features

The Fig. 7 shows how various features interact to give a thorough overview of the dataset. Whereas the off-diagonal elements provide scatter plots for each pair of features, each diagonal element represents a histogram of a single feature. This aids in the identification of possible multicollinearity and correlations.

5 Results and Discussion

With an accuracy rate of 99.9%, the study's results show that the Decision Tree Classifier is the most effective at recognizing fraudulent transactions. RNNs and LSTMs are examples of sequential models that perform well in capturing temporal patterns, coming in second with 99.8% accuracy. In binary classification, KNN and logistic regression both show robustness by attaining a 99.7% accuracy rate. In the meantime, the somewhat lower accuracy rating of 97.6% achieved by the Gaussian Naive Bayes classifier offers a useful point of reference. These findings highlight the various ways that machine learning algorithms may effectively fight fraud, providing information that is essential for putting strong fraud detection systems in place.

The findings show that deep learning and machine learning methods are equally successful in identifying financial fraud in the healthcare industry. The great accuracy of the Decision Tree Classifier can be ascribed to its ability to represent intricate decision boundaries. The high accuracy of sequential models is a reflection of their ability to handle time-series and sequential data, which are common in financial transactions. The efficacy and simplicity of KNN and logistic regression are demonstrated by their performance, which makes them useful instruments for fraud detection.

The distribution and relationships of the data were essential to comprehend, and data visualizations like scatter plots and

ISSN 2319-345X www.ijmrbs.com
Vol. 10, Issue. 3,July 2021

histograms were essential for this. Performance metrics visualizations, which showed the trade-offs between precision, recall, and F1-score, made it easier to compare the models' efficacy.

Model Performance:

Several neural network topologies were used to improve fraud detection in the banking system; the resulting accuracy scores demonstrated the advantages and disadvantages of each strategy. With a 99.9% accuracy rate, the Decision Tree Classifier was notably the best performer. This proves that it can effectively distinguish between real and fraudulent transactions in the dataset.

Table 1: Model Accuracy Scores

S.No	Model	Accuracy
1	Decision Tree Classifier	99.9%
2	Sequential Models	99.8%
3	Logistic Regression	99.7%
4	K-Nearest Neighbors (KNN)	99.7%
5	GaussianNB	97.6%

6 Conclusions

Healthcare financial fraud detection systems are much better at spotting and stopping fraudulent activity when deep learning and machine learning approaches are used. This study demonstrates that models like the Decision Tree Classifier are successful in differentiating between genuine and fraudulent cases, with the ability to attain almost perfect accuracy. Additionally, sequential models with good performance are CNNs and LSTMs, highlighting their usefulness in challenging pattern recognition applications. Subsequent studies should concentrate on the scalability, effectiveness against changing fraud methods, and security of sensitive healthcare data so as to tackle the difficulties associated with applying these models in actual healthcare settings. In general, implementing cutting-edge ML and DL techniques is a revolutionary strategy for protecting financial resources and guaranteeing the quality of healthcare services.

Future studies should concentrate on improving these models' usefulness in actual healthcare environments. To improve fraud

ISSN 2319-345X www.ijmrbs.com
Vol. 10, Issue. 3,July 2021

detection efficacy, this involves putting ensemble learning approaches into practice. This allows you to take advantage of the pooled knowledge of numerous models. The rapid identification and prevention of fraudulent activity depend on the deployment of these models in real-time fraud detection live systems. Sustained effectiveness requires regular model upgrades to stay up to current with changing fraud strategies. Furthermore, by investigating more sophisticated Natural Language Processing (NLP) methods, it may be possible to interpret unstructured data from medical records more accurately and detect fraud. These types of research present viable paths toward improving the identification of healthcare fraud and strengthening healthcare systems against fraudulent activity in the future.

REFERENCES

1. Travaille, P., Müller, R. M., Thornton, D., & Hillegersberg, J. V. (2011). Electronic fraud detection in the US medicaid healthcare program: lessons learned from other industries.
2. Ngai, E. W., Hu, Y., Wong, Y. H., Chen, Y., & Sun, X. (2011). The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. Decision support systems, 50(3), 559-569.
3. Bauder, R. A., & Khoshgoftaar, T. M. (2017, December). Medicare fraud detection using machine learning methods. In 2017 16th IEEE international conference on machine learning and applications (ICMLA) (pp. 858-865). IEEE.
4. Lasaga, D., & Santhana, P. (2018, January). Deep learning to detect medical treatment fraud. In KDD 2017 Workshop on Anomaly Detection in Finance (pp. 114-120). PMLR.
5. Boutaher, N., Elomri, A., Abghour, N., Moussaid, K., & Rida, M. (2020, November). A review of credit card fraud detection using machine learning techniques. In 2020 5th International Conference on cloud computing and artificial intelligence: technologies and applications (CloudTech) (pp. 1-5). IEEE.
6. Rayan, N. (2019, December). Framework for analysis and detection of fraud in health insurance. In 2019 IEEE 6th International Conference on Cloud Computing and Intelligence Systems (CCIS) (pp. 47-56). IEEE.
7. Kose, I., Gokturk, M., & Kilic, K. (2015). An interactive machine-learning-based electronic fraud and abuse detection system in healthcare insurance. Applied Soft Computing, 36, 283-299.
8. Shamitha, S. K., & Ilango, V. (2020, July). A time-efficient model for detecting fraudulent health insurance claims using Artificial neural networks. In 2020 International Conference on System, Computation, Automation and Networking (ICSCAN) (pp. 1-6). IEEE.

ISSN 2319-345X www.ijmrbs.com
Vol. 10, Issue. 3,July 2021

9. 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.
10. Dua, P., & Bais, S. (2014). Supervised learning methods for fraud detection in healthcare insurance. Machine learning in healthcare informatics, 261-285.
11. Bauder, R., Da Rosa, R., & Khoshgoftaar, T. (2018, July). Identifying medicare provider fraud with unsupervised machine learning. In 2018 IEEE international conference on information Reuse and integration (IRI) (pp. 285-292). IEEE.
12. Omar, S. J., Fred, K., & Swaib, K. K. (2018, May). A state-of-the-art review of machine learning techniques for fraud detection research. In Proceedings of the 2018 International Conference on Software Engineering in Africa (pp. 11-19).