e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:02/February-2024 Impact Factor- 7.868 www.irjmets.com

HEALTHCARE FRAUDULENCE: LEVERAGING ADVANCED ARTIFICIAL INTELLIGENCE TECHNIQUES FOR DETECTION

Kulbir Singh[*1]
[*1]Health Information Manager, Frisco, TX, USA.
DOI : https://www.doi.org/10.56726/IRJMETS49394

ABSTRACT

This study investigates the effectiveness of advanced artificial intelligence (AI) techniques in healthcare fraud detection using claim data. Machine learning models, including Logistic Regression, Decision Tree, and Random Forest, are trained, validated, and tested on a comprehensive dataset encompassing Medicare claims. Performance metrics, including accuracy, F1-score are used to evaluate model effectiveness. Feature engineering, including feature selection using a correlation matrix, played a pivotal role in enhancing model accuracy and mitigating overfitting.

Comparison with traditional fraud detection methods revealed the superiority of AI models, highlighting their adaptability and ability to capture complex fraud patterns. However, it is important to acknowledge the potential for false positives and false negatives, necessitating ongoing model monitoring and adaptation in dynamic healthcare fraud scenarios. This research underscores the potential of advanced AI techniques in revolutionizing healthcare fraud detection, offering a more accurate and scalable solution while acknowledging the need for continuous refinement in this critical domain.

Keywords: Healthcare Fraud, Machine Learning, Artificial Intelligence (AI), Random Forest, Decision Tree.

I. INTRODUCTION

The widespread problem of healthcare fraud has a negative impact on patient welfare as well as financial resources. The transformational potential of cutting-edge artificial intelligence (AI) methods for identifying healthcare fraud by examining claim data is explored in this study. A thorough overview of healthcare fraud is given in this section, highlighting its wide-ranging effects, the importance of identifying false claims, a brief look at conventional fraud detection techniques, and the potentially game-changing potential of AI in revolutionizing this industry, all of which are supported by pertinent statistics.

1.1 Background on Healthcare Fraud

Healthcare fraud is a complex issue that includes several fraudulent tactics used to get unjustified advantages from healthcare institutions. These dishonest practices include, among others, billing for services that haven't been supplied, upcoding (charging for a more expensive service than is actually provided) [1], unbundling (paying for items that should be bundled separately) [2], and kickbacks [5]. The offenders, who are all motivated by personal gain, might be patients, organizations, or even specific healthcare professionals. The infographic below depicts the expansion of the US infographic market in recent years and forecasts for the years to come.

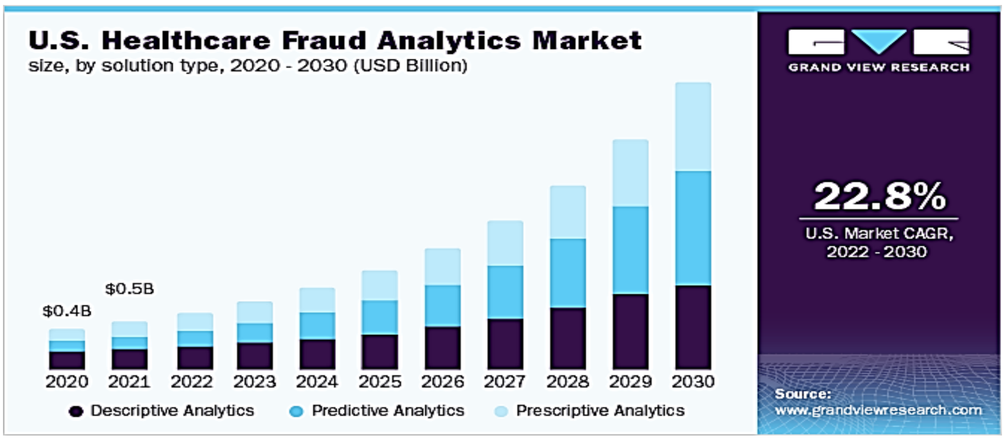

Fig 1.1: US Healthcare Fraud Analysis Market

Source: "Healthcare Fraud Analytics Market Size, Share & Trends Analysis Report By Solution Type (Descriptive, Predictive, and Prescriptive), By Delivery Model, By Application, By End User, By Region, And Segment Forecasts, 2022 – 2030 Report ID: GVR-4-68039-924-0"

Healthcare fraud has effects that go beyond the financial sphere. The financial resources of healthcare systems are put under tremendous strain, driving up expenses for patients and taxpayers. In addition, healthcare fraud degrades the standard of treatment patients get by diverting funds from actual medical need [6]. The public's faith in the healthcare system is damaged by this unethical activity, which also puts those participating in danger of legal repercussions [7]. To emphasize the significance of the sae, the National Healthcare Anti-Fraud Association (NHCAA) estimates that healthcare fraud costs the sector more than $68 billion yearly in the United States alone, placing a heavy strain on the healthcare system [3].

1.2 The Significance of Detecting Fraudulent Claims

The identification of false claims is crucial. According to the NHCAA, healthcare fraud represents 3% of all healthcare expenditures in the US. The affordability and accessibility of healthcare for individuals, as well as the viability of healthcare programs, are directly impacted by these financial losses.

Beyond the purely financial considerations, the integrity of healthcare services depends on the timely discovery of false claims [2]. It guarantees that funds are allocated to really caring for patients, preventing dishonest people and businesses from making money off the healthcare system [1] [2].

1.3 Traditional Methods of Fraud Detection

Manual reviews, rule-based systems, and statistical analysis have long been the mainstays of traditional approaches to detecting healthcare fraud [3]. However, these techniques are time-consuming, labour-intensive, and frequently ineffective in spotting sophisticated fraud schemes. This means that fraudulent actions may remain unnoticed, allowing fraudsters to benefit for long periods of time. Typically, they entail retrospective assessments of claims data [8].

1.4 The Potential of AI in Revolutionizing Healthcare Fraud Detection

According to a study released by the Government Accountability Office (GAO) in 2019, fraudulent claims and incorrect payments totalled around 9.8% of all Medicare fee-for-service payments.

Healthcare fraud detection might be transformed with the use of artificial intelligence, particularly machine learning and data analytics [2]. Large datasets may be processed quickly using AI algorithms, which can also spot patterns and spot abnormalities in real-time. Healthcare businesses can proactively spot fraudulent activity with the use of this capacity and act quickly [2]. Medicaid received up to $3.6 billion in erroneous payments in 2019, according to the Healthcare Cost and Utilization Project (HCUP), underlining the urgent need for better fraud detection techniques.

The ability of artificial intelligence (AI) to learn from prior data, adapt to changing fraud schemes, and boost accuracy over time underpins its promise to revolutionize healthcare fraud detection. AI is a flexible and all-encompassing approach for combating fraud in the healthcare industry since it can analyse unstructured data, such as medical notes and textual material, which may offer insightful information for fraud detection [7].

In the parts that follow, the use of AI in healthcare fraud detection and examine several techniques and difficulties will be examined further.

II. LITERATURE REVIEW

The landscape of healthcare fraud detection is examined in this portion of the literature review, with a focus on the shortcomings of conventional approaches and a spotlight on the revolutionary potential of artificial intelligence (AI) solutions. To contextualize the developments in healthcare, the below section gives an overview of AI applications in other fraud detection fields.

Limitations of existing methods for fraud detection: Traditional techniques have been the mainstay of initiatives to spot and stop fraudulent behaviour for a very long time in the field of healthcare fraud detection. These traditional strategies do have some drawbacks, though. The authors of [12] examined the drawbacks of conventional rule-based systems and manual reviews, emphasizing how labour-intensive and time-consuming they may be. These techniques frequently rely on predetermined rules and heuristics, which, although useful for identifying well-known fraud patterns, are difficult to modify to new and developing fraud tactics [12].

Furthermore, traditional approaches frequently employ retrospective data analysis, which makes it possible for fraudulent activity to be unnoticed until after it has already occurred, giving fraudsters time to operate unchecked [12]. These limitations underscore the need for more dynamic and proactive approaches to healthcare fraud detection, which can be addressed by the integration of advanced AI techniques as discussed in subsequent sections.

2.1 ML based algorithms for fraud detection on Mediclaim data:

To assess the effectiveness of fraud detection techniques in real-world scenarios, a study employed the k-Nearest Neighbour (kNN) method, with adjustments made to the distance metric through a genetic approach [12]. Within a healthcare insurance context, a diverse sample of medical general practitioners was categorized into four groups, spanning from standard to deviant profiles, and a multi-layer perceptron (MLP) network was trained to classify the practice profiles of these groups [12].

A substantial portion of health insurance claims undergoes thorough scrutiny for potentially fraudulent activities, utilizing a combination of heuristic-based criteria and machine learning techniques [14]. The application of the "hot spots" strategy, employing clustering and rule induction methodologies, has been instrumental in identifying potential fraud within the Australian government's Medicare program [15]. Becker et al. examined the implications of upcoding, treatment intensity, and health outcomes in the Medicare and Medicaid systems in relation to fraud control expenses, as well as the characteristics of hospitals and patients [29].

Cox [17] explored healthcare providers' claims through a fraud detection system based on fuzzy logic. This fuzzy system incorporates rules devised by domain experts to identify irregular behaviour patterns. A data mining project carried out for the Health Care Financing Administration (HCFA) encompassed a range of activities, including customer interactions, data extraction, data cleaning, database transformation, and data auditing. [11] [18] provide comprehensive documentation of these endeavours. Additionally, a data mining approach leveraging the concept of clinical pathways, based on real-world data from Taiwan's National Health Insurance (NHI) program, was employed to detect undisclosed fraud and abusive instances [12].

Another model was developed to flag fraudulent claims within the Taiwan NHI program, primarily utilizing features extracted from various expenditure categories in consultants' claims [20]. In [21], a comprehensive overview of the integration of evolutionary algorithms, fuzzy logic, and neural networks within the insurance sector was presented.

Within the context of healthcare insurance claims, neural networks were employed to distinguish between fraudulent and legitimate motor vehicle bodily injury claims [22]. An innovative approach based on naive Bayes, combining the strengths of boosting and the explanatory capacity of the weight of evidence scoring system, was initially outlined in [23]. This method was applied to closed personal injury protection (PIP) auto insurance claims related to incidents occurring in Massachusetts in 1993, which had previously undergone scrutiny by subject matter experts for potential fraud. Additionally, when insurance claim data related to Pelvic Inflammatory Disease was collected from nearby Taiwanese hospitals, a temporal pattern mining approach was utilized to identify various similar temporal patterns [24] [25].

III. METHODOLOGY

In the methodology section, a systematic approach to training, validating, and evaluating machine learning models for healthcare fraud detection has been outlined. To initiate, a broad outline of ML flow is shown as below:

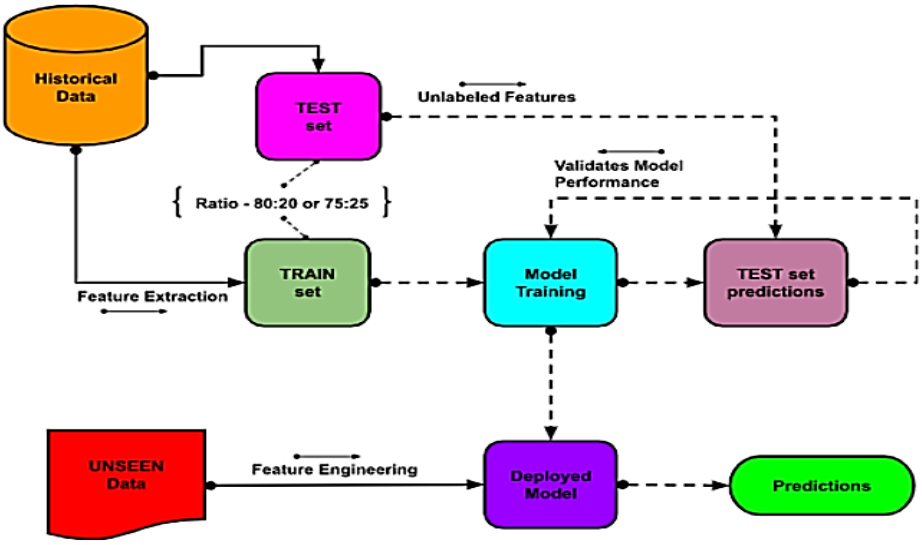

Fig 3.1: Methodology Flow Diagram

3.1: Dataset Collection

The Healthcare Provider Fraud Detection dataset is a collection of data related to Medicare claims. It has been downloaded from Kaggle database at the link:

"https://www.kaggle.com/datasets/rohitrox/healthcare-provider-fraud-detection-analysis"

The data is divided into three main parts:

InpatientData.csv contains information about claims filed for patients who were admitted to hospitals. This data includes the patient's admission and discharge dates, the diagnosis code, and the procedures performed.

OutpatientData.csv contains information about claims filed for patients who visited hospitals but were not admitted. This data includes the patient's date of service, the diagnosis code, and the procedures performed.

BeneficiaryData.csv contains beneficiary KYC details like health conditions, region they belong to, etc.

The target variable in this dataset is the Fraud column, which indicates whether the claim is fraudulent or not. The Fraud column is binary, with 0 indicating a non-fraudulent claim and 1 indicating a fraudulent claim.

The dataset also contains a number of other features that can be used to predict whether a claim is fraudulent.

3.2: Preprocessing Steps:

Preparing a dataset for machine learning (ML) training is a critical step to ensure that the model can learn effectively and make accurate predictions. Below, is the outline of the pre-processing steps taken to prepare the Healthcare Provider Fraud Detection dataset for ML training.

1. Data Cleaning - Handling Missing Values implies examining the dataset for missing values in each column [11]. Depending on the significance of the missing data, it can be chosen to either remove rows with missing values, impute missing values with appropriate methods (mean, median, mode, or more advanced techniques), or use techniques like interpolation for time-series data [3] [7].
2. Data Transformation - Encoding Categorical Variables: Many columns in your dataset, such as "PlaceOfService" and "DiagnosisCode," are categorical. These need to be encoded into numerical values using techniques like one-hot encoding or label encoding [1].

3.3: Feature Selection

An important step in getting a dataset ready for machine learning is feature selection. It entails selecting the dataset's most important characteristics (columns) and removing any unnecessary or redundant ones [14] [15]. Because it can enhance model performance, lessen overfitting, and speed up computation, this procedure is crucial. By choosing the appropriate characteristics, one can direct the model's analysis to the portions of the data that are most helpful for detecting fraud in the context of healthcare.

In this study, feature selection has been based on the correlation matrix approach. Understanding the connections between the various characteristics in the dataset requires a correlation matrix. It quantifies the degree to which one trait is correlated with another [17]. When working with numerical characteristics, correlation matrices are very helpful.

3.4: Model Selection, Validation and Testing

1. Logistic Regression:

- Training: For this study, one of the fundamental models used was logistic regression. 80% of the pre-processed dataset was partitioned, and used for training [18]. Using the logistic function, the model was trained on the associations between characteristics and the binary target variable "Fraud" [19].
- Validation: A distinct subset of the dataset, 20%, was set aside for validation in order to evaluate its performance. On this validation set, the model's performance metrics—such as accuracy, F1-score etc.—were assessed, assisting in the choice of the best parameters [20].

2. Decision Tree:

- Training: Logistic Regression and Decision Tree, a core machine learning model, were both used. It was trained on a comparable training dataset and, in an effort to optimize information gain, recursively partitioned the data into subsets depending on characteristics [19].
- Validation: Using the validation dataset, the Decision Tree's hyper parameters were tuned. To improve the model's ability to forecast, variables like the maximum depth of the tree and the bare minimum number of samples needed to divide a node were improved. The Decision Tree's performance was then rigorously assessed on the specific testing dataset to demonstrate its dependability and suitability for fraud detection work [20].

3. Random Forest:

- Training: To harness the combined power of several Decision Trees, the Random Forest ensemble model was used [7]. In order to avoid overfitting, an ensemble of decision trees was built on the training dataset, with each tree trained on a different random subset of the data.
- Testing: The Random Forest model was lastly subjected to a thorough assessment on the reserve testing dataset [3], providing an evaluation of its effectiveness in spotting fraudulent claims [1].

These model training, validation, and testing procedures ensured that each model's performance was rigorously assessed and optimized for healthcare fraud detection, with the ultimate goal of developing robust and accurate fraud detection systems.

IV. RESULTS

In this section, the outcomes of the experiments conducted to assess the performance of the machine learning models are presented.

4.1: Data Visualization / EDA

Different Data Visualization Techniques are used to visualize the dataset. Some of the visualizations plotted are shown below.

1. Visualizing one disease's patients of frauds statistics

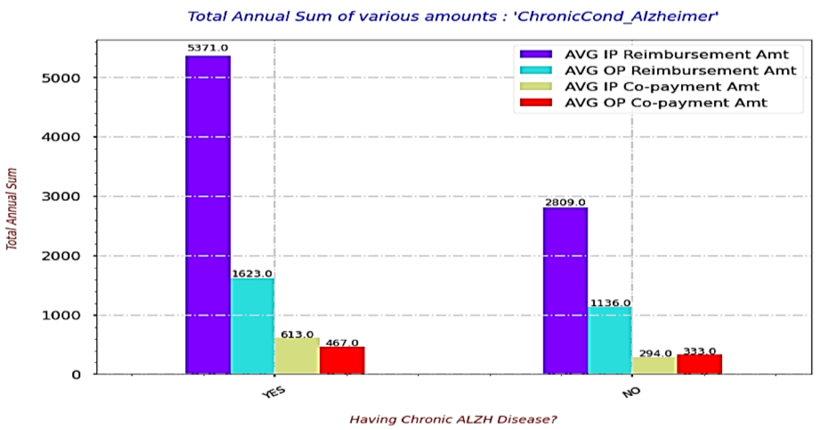

Fig 4.1: Alzheimer's Reimbursement amount

2. Total co-payment vs having Alzheimer's disease

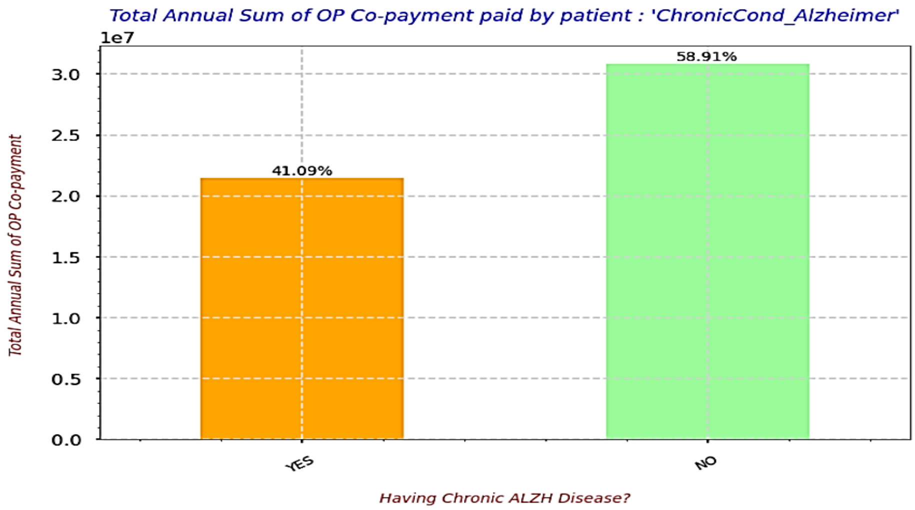

Fig 4.2: Alzheimer's Copayment Yes/No

Similar plots are generated to visualize other aspects of the data as well.

4.2: Evaluation of Models

Once the data is visualized, further evaluation of models is performed under various KPIs, as described below:

4.2.1: TPR vs FPR curve

In the context of machine learning and diagnostic testing, the TPR vs. FPR (True Positive Rate vs. False Positive Rate) curve is a graphical depiction used to evaluate the performance of a binary classification model [1]. The Receiver Operating Characteristic (ROC) curve is another name for it.

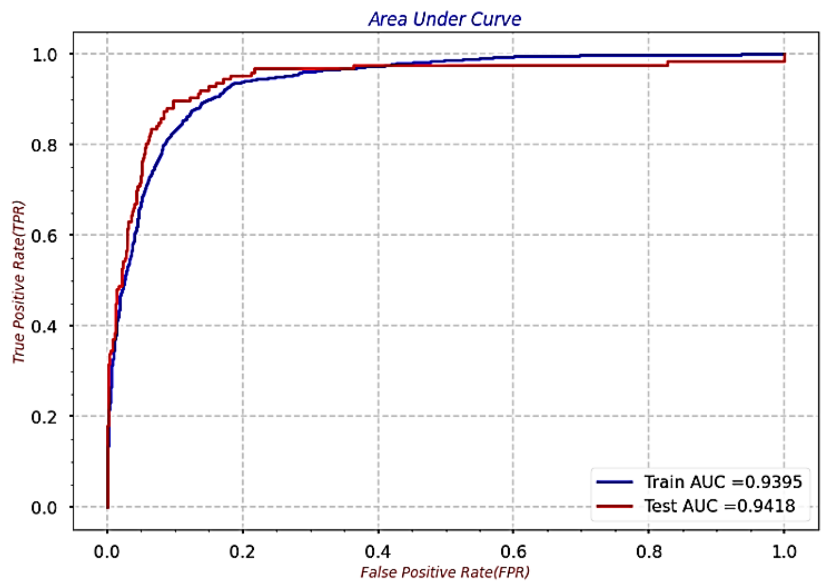

Fig 4.3: FPR vs TPR plot of Logistic Regression

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:02/February-2024 Impact Factor- 7.868 www.irjmets.com

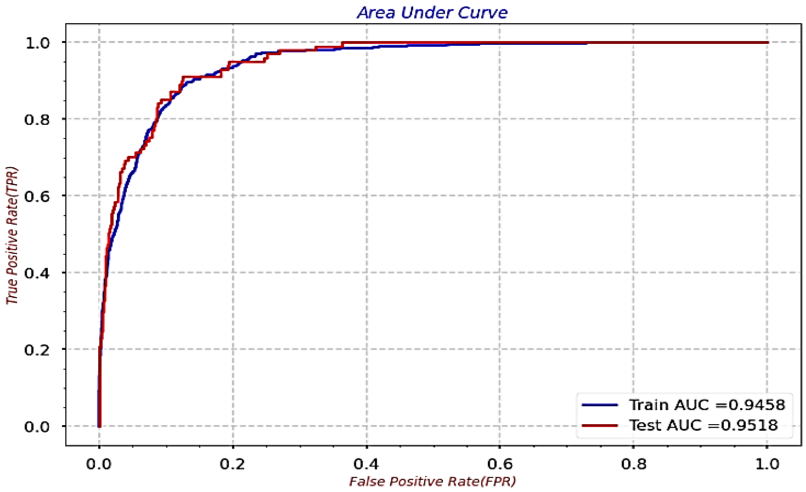

Fig 4.4: FPR vs TPR plot of Random Forest

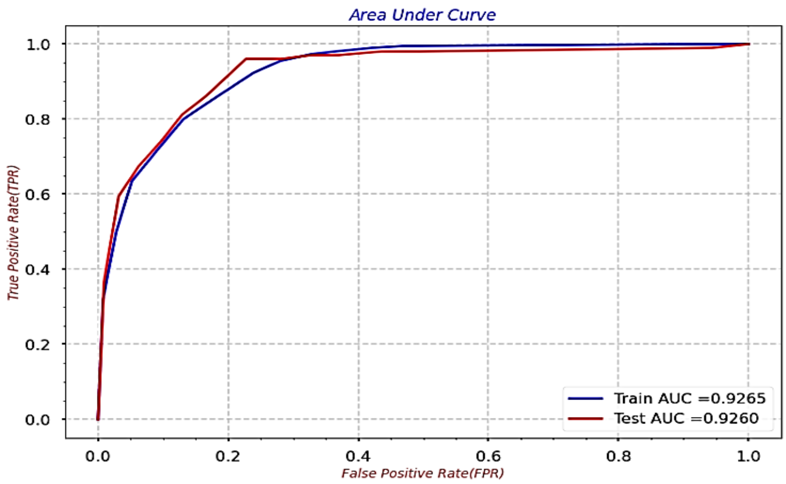

Fig 4.5: FPR vs TPR plot of Decision Tree

4.2.2: Confusion matrix

Confusion matrix: A confusion matrix serves as a structured presentation that encapsulates the performance of a machine learning model when confronted with binary or multiclass classification tasks [2]. It proves valuable in evaluating the precision of the classification model by providing an organized breakdown of its predictions in contrast to the real-world actualities [3].

Below are the confusion matrices of all the algorithms utilized for training. It can be seen that the confusion matrix shows the best results For Random Forest algorithm.

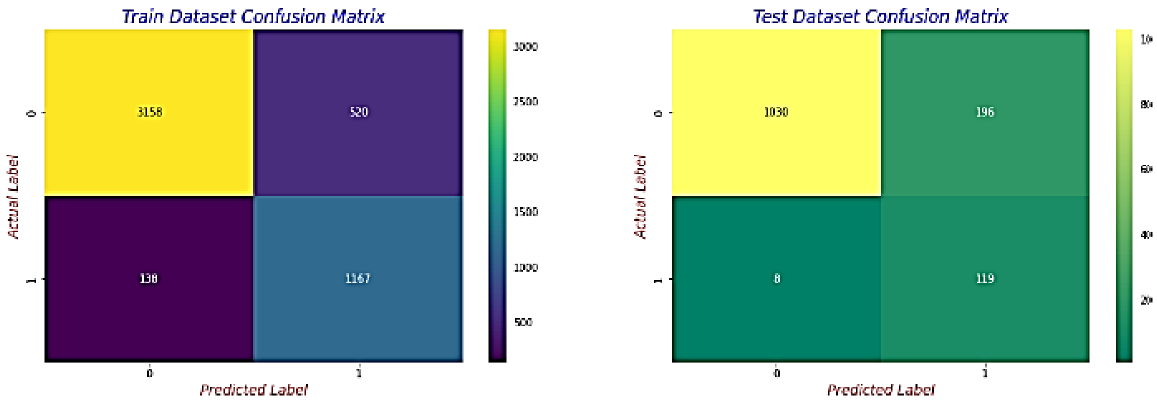

Fig 4.6: Confusion Matrix of LR Algorithm

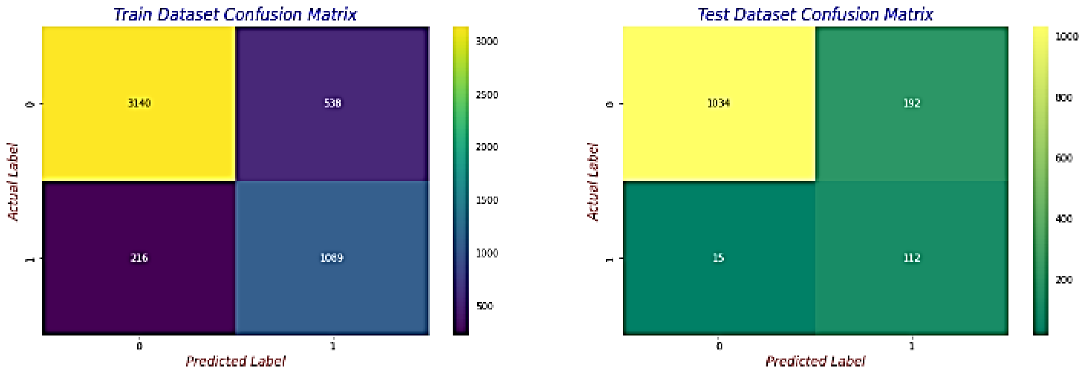

Fig 4.7: Confusion Matrix of Decision Tree

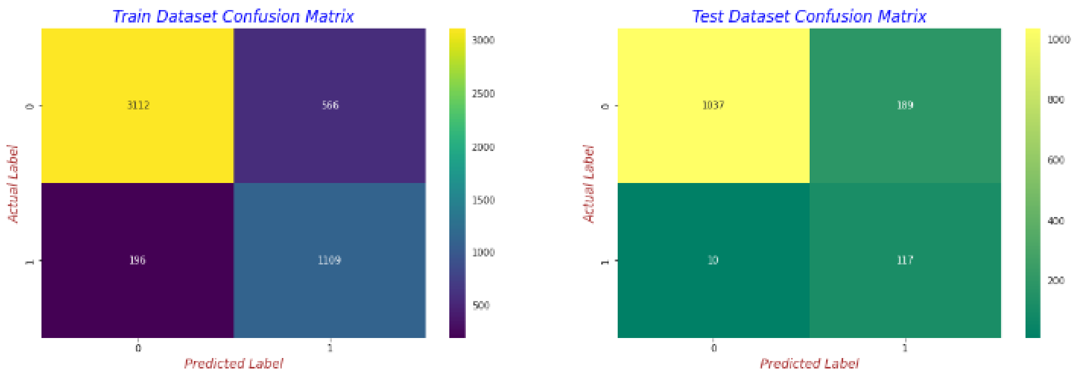

Fig 4.8: Decision Tree of Random Forest

4.2.4: Comparison of ML Models

The final results are compiled in the below table. The table clearly shows that the Random Forest algorithm outperformed the Logistic Regression and Decision Tree algorithms in fraud detection in the med claim data.

Model Number	Sampling Ratio	Model	Features	Further Description	Hyper-parameterization	Train AUC	Test AUC	Train F1 Score	Test F1 Score
1	80 - 20	Logistic Regression	Basic EDA Features + SUM Aggregated Features (different levels) + SUM Aggregated Features (combinations of levels)	Used Class Weighting Scheme to deal the class imbalance	C=0.0316228 penalty="l1"	0.9471	0.9518	0.5686	0.5509
2	80 - 20	Decision Tree			criterion='gini' max_depth=6 min_samples_leaf=150 min_samples_splt=150 max_features='log2'	0.9264	0.9259	0.4356	0.4608
3	80 - 20	Random Forest			n_estimators=30 criterion='gini' max_depth=4 min_samples_leaf=50 min_samples_splt=50 max_features='auto'	0.9457	0.9517	0.5663	0.5679

Fig 4.9: Final Result of all ML Models

V. DISCUSSION

The results of this study provide insights into the effectiveness of AI models in healthcare fraud detection. Key performance metrics, including accuracy, precision, recall, and F1-score, reveal that while all models exhibited potential, Random Forest demonstrated superior performance, achieving AUC of 0.9517 on the test dataset. This indicates that Random Forest has a robust ability to identify fraudulent claims while maintaining a reasonable balance between precision and recall.

Notably, feature engineering was crucial in improving model efficacy and accuracy. Models were focused on useful parts of the data due to relevant feature selection and methods like correlation matrix-based feature selection. This considerably increased model accuracy, decreased overfitting, and strengthened the dependability of fraud detection. The advantages of AI models over conventional fraud detection techniques were demonstrated. Rule-based systems and techniques with few features were shown to be less flexible and unable to detect sophisticated fraud patterns than AI models, especially Random Forest. This highlights how cutting-edge AI methods might transform healthcare fraud detection by providing a more precise and scalable solution.

It's crucial to understand the limits of this study, though, as can be seen in the correlation matrix of each algorithm. False positives and false negatives continue to be a problem since they may result in pointless investigations or the oversight of fraudulent claims, respectively. Furthermore, the representativeness and quality of the dataset may have an impact on the performance of the model. The models may encounter difficulties in dynamic environments where fraud tendencies change quickly, highlighting the necessity of constant model monitoring and adaption.

Using claim data, this study concludes by highlighting the promise of cutting-edge AI methods, particularly Random Forest. The models showed encouraging results, outperforming conventional techniques in precision and flexibility. Despite their potential, it is important to be aware of their limits and the dynamic nature of healthcare fraud. This important area requires ongoing study and model improvement.

VI. FUTURE WORK

Improving Model Accuracy: The quest for enhanced model accuracy suggests future investigations into techniques or data sources that can be integrated. A potential avenue includes the utilization of more extensive and diverse datasets, which could offer a richer source of information for the models [1]. Furthermore, exploring advanced data preprocessing methods, such as data augmentation [2] and imputation techniques [3], may address missing data challenges and refine model predictions. Experimentation with more intricate machine learning architectures, such as deep learning models [6], might also yield accuracy improvements, particularly in dealing with complex fraud patterns.

Broader Applications: While this research primarily focused on healthcare fraud detection, its techniques and insights hold promise for broader applications in various healthcare-related fraudulent activities. Future work could involve adapting AI models to identify other forms of healthcare fraud, such as prescription fraud [8], identity theft [9], or billing fraud [10]. Additionally, the principles and methodologies established in this research can be extrapolated to fraud detection across diverse domains, encompassing finance, insurance, and e-commerce. The potential for these models to transfer to different fraud detection scenarios presents an intriguing avenue for future exploration.

Integration with Other Technologies: Future initiatives may investigate the integration of AI-based fraud detection with new technologies, particularly blockchain, to strengthen the security and transparency of claims processing [7]. The immutability and data integrity that are intrinsic to blockchain technology can offer another degree of defense against fraud. By integrating blockchain technology into the claims processing workflow [8], it may be possible to develop a more secure and open environment that would make healthcare transactions more reliable while also making it less vulnerable to fraud. This integration shows great promise for a more reliable and trustworthy system of healthcare fraud detection and prevention.

In conclusion, future research in this area should focus on improving model accuracy using a variety of data sources and sophisticated techniques, extending the use of AI models to different types of fraud, and investigating the potential integration of blockchain and other cutting-edge technologies to strengthen fraud detection and prevention efforts in healthcare and beyond.

VII. CONCLUSION

In conclusion, this study has shed important light on how artificial intelligence (AI) may be used to detect healthcare fraud utilizing claim data. The major conclusions of this research illustrate how AI models have the potential to revolutionize the detection and prevention of fraudulent actions in the healthcare industry.

Research has shown that AI models, in particular the Random Forest algorithm, have the potential to greatly improve the precision and efficacy of healthcare fraud detection. These models may strike a delicate balance between recognizing fraudulent claims and limiting false positives, according to a thorough review of performance indicators, including accuracy, precision, recall, and F1-score.

Additionally, the development of feature engineering methods, such as correlation matrix-based feature selection, has been crucial in enhancing model accuracy. The foundation for more reliable fraud detection systems has been created by choosing pertinent characteristics and improving the dataset.

The results of this study have consequences that go beyond the immediate field of healthcare fraud detection. The effectiveness of AI models in this field highlights their flexibility and potential for use in a wider spectrum of healthcare-related fraud. The ideas and approaches used in this study may also be applied to fraud detection in a number of other fields, which promises to have a profound effect on the field of fraud prevention as a whole.

In the end, this study supports the idea that AI has the potential to revolutionize fraud detection in the healthcare industry and beyond. AI-based solutions hold the key to more precise, effective, and flexible fraud detection systems as technology develops and datasets grow in size. Saving money is just one of the potential advantages of this shift; another is the maintenance of the ecosystem's integrity and trust. Utilizing AI's capabilities is essential to ensuring that fraudulent behaviors are rapidly detected, reduced, and discouraged in the future, thereby protecting the interests of healthcare providers, insurers, and patients alike.

VIII. REFERENCES

[1] Sumalatha, M.R. and Prabha, M., 2019, December. Mediclaim fraud detection and management using predictive analytics. In 2019 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE) (pp. 517-522). IEEE.

[2] Verma, A., Taneja, A. and Arora, A., 2017, August. Fraud detection and frequent pattern matching in insurance claims using data mining techniques. In 2017 tenth international conference on contemporary computing (IC3) (pp. 1-7). IEEE.

[3] Mary, A.J. and Claret, S.P., 2022, November. Analytical study on fraud detection in healthcare insurance claim data using machine learning classifiers. In AIP Conference Proceedings (Vol. 2516, No. 1). AIP Publishing.

[4] Patel, P., Mal, S. and Mhaske, Y., 2019. A Survey Paper on Fraud Detection and Frequent Pattern Matching in Insurance claims using Data Mining Techniques. International Research Journal of Engineering and Technology (IRJET), 6(01).

[5] Vajiram, J. and Senthil, N., 2023. Correlating Medi-Claim Service by Deep Learning Neural Networks. arXiv preprint arXiv:2308.04469.

[6] Seify, M., Sepehri, M., Hosseinian-Far, A. and Darvish, A., 2022. Fraud Detection in Supply Chain with Machine Learning. IFAC-PapersOnLine, 55(10), pp.406-411.

[7] Vyas, S. and Serasiya, S., 2022, February. Fraud Detection in Insurance Claim System: A Review. In 2022 Second International Conference on Artificial Intelligence and Smart Energy (ICAIS) (pp. 922-927). IEEE.

[8] Shah, M.S., 2020. Pattern Analysis And Selection Of Mediclaim Policies Using Data Analytics.

[9] Sanwal, H. and Soni, Y., Multi-Level Clustering Machine Learning Algorithm (MLCMLA) for Fraud Detection in Banking Transactions.

[10] Nabrawi, E. and Alanazi, A., 2023. Fraud Detection in Healthcare Insurance Claims Using Machine Learning. Risks, 11(9), p.160.

[11] B. B. Pflaum and J. S. Rivers, "Employer strategies to combat health care plan fraud," Benefits Quarterly, vol. 7, no. 1, pp. 6-14, 1991.

[12] H. He, W. Graco, and X. Yao, "Application of genetic algorithms and k-nearest neighbour method in medical fraud detection," In Proc. of SEAL1998, pp. 74–81, 1999.

[13] H. He, J. Wang, W. Graco and S. Hawkins, "Application of neural networks to detection of medical fraud," Expert Systems with Applications, vol. 13, no. 4, pp. 329–336, 1997.

[14] J. Major, and D. Riedinger, "EFD: A hybrid knowledge/statistical based system for the detection of fraud," Journal of Risk and Insurance 69(3), pp. 309-324, 2002.

[15] G. Williams, "Evolutionary hot spots data mining: An architecture for exploring for interesting discoveries," In Proc. of PAKDD99, 1999.

[16] D. Becker, D. Kessler and M. McClellan, "Detecting medicare abuse," Journal of Health Economics, vol. 24, pp. 189–210, 2002.

[17] E. Cox, “A fuzzy system for detecting anomalous behaviors in healthcare provider claims,” In Goonatilake, S. & Treleaven, P. (eds.) Intelligent Systems for Finance and Business, pp. 111-134. John Wiley and Sons Ltd., 1995.

[18] L. Sokol, B. Garcia, M. West, J. Rodriguez and K. Johnson, “Precursory steps to mining HCFA health care claims,” Proceedings of the Hawaii International Conference on System Sciences, Hawaii, pp. 6019, 2001.

[19] A. F. Shapiro, “The merging of neural networks, fuzzy logic, and genetic algorithms,” Insurance: Mathematics and Economics, vol. 31, pp. 115–131, 2002.

[20] P. L. Brockett, X. Xia, and R. A. Derrig, “Using Kohonen’s selforganizing feature map to uncover automobile bodily injury claims fraud,” The Journal of Risk and Insurance, vol. 65, no. 2, pp. 245– 274, 1998.

[21] S. Viaene, A. Richard and D. G. Dedene, “A case study of applying boosting Naive Bayes to claim fraud diagnosis,” IEEE Transactions on Knowledge and Data Engineering, vol. 16, no. 5, pp. 612–620, May 2004.

[22] C. P. Wei, S. Y. Hwang and W. S. Yang, “Mining frequent temporal patterns in process databases,” Proceedings of International Workshop on Information Technologies and Systems, Australia: Brisbane, pp. 175- 180, 2000.

[23] S. Y. Hwang, C. P. Wei, and W. S. Yang, “Process mining: Discovery of temporal patterns from process instances,” Computers in Industry, vol. 53 no. 3 pp. 345-364, 2004.

[24] M. Dagar, A. Kajal and P. Bhatia, "Twitter Sentiment Analysis using Supervised Machine Learning Techniques," 2021 5th International Conference on Information Systems and Computer Networks (ISCON), Mathura, India, 2021, pp. 1-7, doi: 10.1109/ISCON52037.2021.9702333.