Volume 10, Issue 2, Feb – 2025 International Journal of Innovative Science and Research Technology
ISSN No: 2456-2165 https://doi.org/10.38124/ijisrt/25Feb1337

Advanced Investigation of Healthcare Fraud Detection Utilizing Machine Learning Algorithms

Saeed H. A.[1]; Naeem N. Alyona A. E.[2]

[1,2]NRU Russia, Muhammad Nawaz Sharif University of Agriculture Multan, Pakistan, ITMO University Russia

Publication Date: 2025/06/02

Abstract: Healthcare fraud is a fast-growing issue that causes substantial financial loss and affects the quality of patient care. Conventional fraud detection techniques tend to be ineffective in detecting fraudulent claims because healthcare data is complex and enormous in volume. This research investigates the use of machine learning methods to enhance fraud detection within healthcare systems. We contrast the performance of Decision Tree, Random Forest, K-Nearest Neighbors (KNN), and Logistic Regression both prior to and post-hyperparameter tuning and feature selection. Forward feature selection was done with KNN and Logistic Regression to improve model performance by choosing the most salient features, whereas hyperparameter tuning was utilized to fine-tune all the models. Metrics of evaluation like accuracy, precision, recall, F1-score, confusion matrix, and ROC curves were employed to measure the effectiveness of the models. The outcome reveals that Logistic Regression had the highest accuracy following optimization and feature selection over other models in identifying fraudulent claims. The Voting Classifier, which is an ensemble learning, enhanced fraud detection by aggregating various models for enhanced predictive capability. Though Decision Tree and Random Forest performed well, tuning was not effective in improving their accuracy. These results indicate that machine learning methods, especially ensemble models and feature selection, can dramatically improve healthcare fraud detection. Subsequent studies need to integrate deep learning and advanced ensemble techniques to further enhance fraud detection accuracy and reduce false positives.

Keywords: *Healthcare Fraud, Machine Learning, Fraud Detection, Hyperparameter Tuning, Feature Selection, Ensemble Learning.*

How to Cite: Saeed H. A.; Naeem N. Alyona A. E. (2025) Advanced Investigation of Healthcare Fraud Detection Utilizing Machine Learning Algorithms. *International Journal of Innovative Science and Research Technology*, 10(2), 2594-2597. https://doi.org/10.38124/ijisrt/25Feb1337

I. INTRODUCTION

The healthcare sector is among the most vital sectors in society, offering vital medical care to enhance public health. Notwithstanding its advantages, the sector is plagued by several challenges, with healthcare fraud being a significant issue that needs immediate attention. Several fraudulent practices are present in the sector, resulting in huge financial losses and possible injury to patients[1]. Moreover, such fraudulent activities erode public confidence in the healthcare system. Studies show that healthcare fraud amounts to billions of dollars every year, which is a constant challenge for healthcare providers[2] .

As the use of data-driven decision-making and technological innovations increases, the significance of fraud detection in healthcare has also increased. As indicated by [3], the International Journal of Data Science and Analytics points to the increasing interest in statistical techniques across various fields, such as statistics, computer science, computational mathematics, and physics. This points to the necessity of interdisciplinary cooperation in fraud detection. In addition, cybercrime threats continue to soar, and conventional rule-based methods have not been enough in identifying sophisticated and dynamic fraud schemes[4].

To counter these challenges, the healthcare sector is progressively embracing sophisticated technologies like machine learning to improve the fraud detection capacity. Machine learning models enhance accuracy and effectiveness by processing enormous volumes of structured and unstructured data, enabling them to pick up concealed patterns and anomalies that human reviewers might not observe. Furthermore, these models facilitate ongoing learning from fresh data, which increases the flexibility of fraud detection systems against new threats[5] .

Leelakumar Raja Lekkala's research indicates the increasing relevance of machine learning models in the detection of healthcare fraud. Despite technological innovations and AI adoption, healthcare fraud remains a serious issue, causing economic losses and deteriorating patient care.

**Special description of the title. (dispensable)

The research examines how fraud detection is enhanced through machine learning by processing heterogeneous datasets such as claims, billing information, and patient demographics. By using algorithms like Random Forest, CNNs, and RNNs, machine learning models are more accurate, more precise in recall, and more responsive to changing fraud patterns than conventional techniques. The results indicate that these models give a robust and efficient mechanism to healthcare organizations for effectively fighting fraudulent activity[5].

II. MATERIAL AND METHODS

➢ *Dataset*

The project's data set comprises Inpatient Data, Outpatient Data, and Beneficiary Details Data, with a complete picture of Medicare claims to identify fraud. Inpatient Data includes hospitalized patient claims with admission/discharge dates and diagnosis codes. Outpatient Data includes non-admitted patient claims, offering outpatient services information. Beneficiary Details Data includes patient demographics, medical conditions, and regional memberships. The data sets, together, offer fraudulent healthcare claims patterns. Data source: Kaggle.

➢ *Preprocess*

In the first step of dataset preparation for fraud detection, preprocessing is performed, including merging of Inpatient Data, Outpatient Data, and Beneficiary Details Data into a complete dataset. After the merging operation, identification and elimination of missing values are performed to maintain data integrity. This specific step is crucial to the accuracy of the analysis and prevent biased results that may be initiated by incomplete data.

Following purification of the dataset, Forward Feature Selection is employed to determine the most important features that are pertinent to fraud detection. The technique starts from the bottom level without features and includes the most important features step by step based on a given evaluation metric, such as accuracy or AUC score. By targeting the most influential features, this process not only improves the performance of the model but also prevents overfitting and improves computational efficiency[6].

Hyperparameter tuning is the process of discovering the best set of hyperparameters for a machine learning model in order to maximize its overall performance. Unlike model parameters, which are tuned from the data, hyperparameters are fixed parameters beforehand that influence the training process, such as the learning rate, the number of trees in a random forest, or the number of neighbors in K-nearest neighbors (KNN)[7].

➢ *Machine Learning Algorithm*

In this study of healthcare fraud discovery, several machine learning algorithms were utilized to improve classification precision and fraud detection. Logistic Regression[8], [9] was utilized due to its interpretability and efficiency in binary classification. Decision Tree[10] offered rule-based decision-making, while Random Forest, an ensemble method, generalized better through the averaging of numerous decision trees. K-Nearest Neighbors (KNN)[11] classified instances based on similarity measures and was thus beneficial for pattern recognition. Furthermore, a Voting Classifier combined several models to improve predictive performance. Hyperparameter tuning and feature selection further improved these models to provide robust fraud detection with high accuracy, precision, recall, and F1-score.

III. RESULTS

➢ *Results of four Machine Learning Algorithm*

Following the optimization and feature selection process, Logistic Regression was the highest-performing model overall, with improved accuracy, precision, recall, and F1-score compared to other models. The use of Forward Feature Selection greatly improved the performance of KNN and Logistic Regression by focusing on the most important features and omitting irrelevant features. This led to significant improvements, especially in accuracy and recall, as the models became more efficient at distinguishing between classes. Nevertheless, even with these improvements, Logistic Regression was always superior to KNN in performance, as indicated by its higher applicability to the given dataset. Meanwhile, hyperparameter tuning provided limited effects on the performance of Decision Tree and Random Forest models. While there were improvements in performance in these models, they still lagged behind Logistic Regression. This is because the nature of the dataset was more suitable for simpler linear models than for complex tree-based models. While KNN improved through feature selection and tuning, its performance was still slightly behind Logistic Regression, further indicating the latter's superiority in performing the classification task at hand.

Table 1 Model Performance Before Hyperparameter Tuning

Model	Accuracy	Precision	Recall	F1-Score	AUC
Decision Tree	0.468	0.451	0.416	0.432	0.467
Random Forest	0.462	0.449	0.455	0.452	0.462
KNN	0.468	0.453	0.442	0.447	0.468
Logistic Regression	0.557	0.556	0.455	0.500	0.554

Table 2 Model Performance After Hyperparameter Tuning

Model	Accuracy	Precision	Recall	F1-Score	AUC
Decision Tree	0.494	0.448	0.169	0.245	0.486
Random Forest	0.449	0.431	0.403	0.416	0.448
KNN (Tuned)	0.487	0.473	0.455	0.464	0.487
Logistic Regression (Tuned)	0.519	0.508	0.416	0.457	0.516

ISSN No: 2456-2165 https://doi.org/10.38124/ijisrt/25Feb1337

Table 3 Performance After Forward Feature Selection

Model	Accuracy	Precision	Recall	F1-Score	AUC
KNN (Feature Selection Applied)	0.495	0.478	0.467	0.472	0.495
Logistic Regression (Feature Selection Applied)	0.526	0.515	0.425	0.466	0.522

Table 4 Final Optimized Model Performance

Model	Accuracy	Precision	Recall	F1-Score	AUC
Decision Tree (Optimized)	0.506	0.472	0.183	0.263	0.497
Random Forest (Optimized)	0.463	0.452	0.421	0.436	0.462
KNN (Optimized + Feature Selection)	0.507	0.485	0.472	0.478	0.507
Logistic Regression (Optimized + Feature Selection)	0.532	0.522	0.432	0.474	0.528

- *Results of Voting Classifier*

The model achieved high accuracy (96.7%), indicating that it correctly classified most instances in the dataset. The precision (0.99 for Class 0 and 0.94 for Class 1) shows that the model made very few false-positive errors, meaning it effectively distinguishes between the two classes. The recall (0.95 for Class 0 and 0.99 for Class 1) suggests that the model successfully identified most positive and negative cases, with particularly strong performance in detecting positive cases (Class 1). The F1-score, which balances precision and recall, remains consistently high for both classes (0.97 for Class 0 and 0.96 for Class 1), further confirming the model's reliability. The macro average and weighted average metrics reinforce that the model maintains balanced performance across both classes. Overall, the results indicate that the model is well-optimized and performs effectively in distinguishing between the two categories.

Table 5 Performance Metrics Table

Metric	Class 0 (Negative)	Class 1 (Positive)	Macro Avg	Weighted Avg	Overall Accuracy
Precision	0.99	0.94	0.96	0.97	**0.967** (96.7%)
Recall	0.95	0.99	0.97	0.97	-
F1-Score	0.97	0.96	0.97	0.97	-
Support	1269	875	-	-	2144

- *Comparison*

The comparison of the performance of various classification models indicates that the Voting Classifier had the highest accuracy (96.69%), with good precision (96.32%) and recall (96.97%), and hence is the overall top-performing model. The Random Forest model ranked second with 96.59% accuracy, which was only marginally less than the Voting Classifier, but still with good precision (96.18%) and recall (96.98%). The K-Nearest Neighbors (KNN) Classifier and the Decision Tree also did not differ much from each other, with accuracies of 95.42% and 95.47%, respectively, having less precision and recall measures compared to the ensemble models. Logistic Regression achieved the lowest accuracy (85.87%) among all models, so this might not be as useful in this classification problem. Despite that, it still had decent precision (86.52%) and recall (84.17%), which indicates that even though it is not the ideal model, it is still an option depending on the application. Overall, the findings show that ensemble methods such as Voting Classifier and Random Forest tend to outperform individual models such as Logistic Regression, Decision Tree, and KNN based on accuracy, precision, and recall.

Table 6 Comparison of all Models

	Accuracy	precision	recall
logistic_regression	0.858675	0.865252	0.841760
random_forest	0.965951	0.961828	0.969818
decision tree	0.954757	0.950112	0.960007
voting_classifier	0.966884	0.963168	0.969719
knn classifier	0.954291	0.949597	0.960500

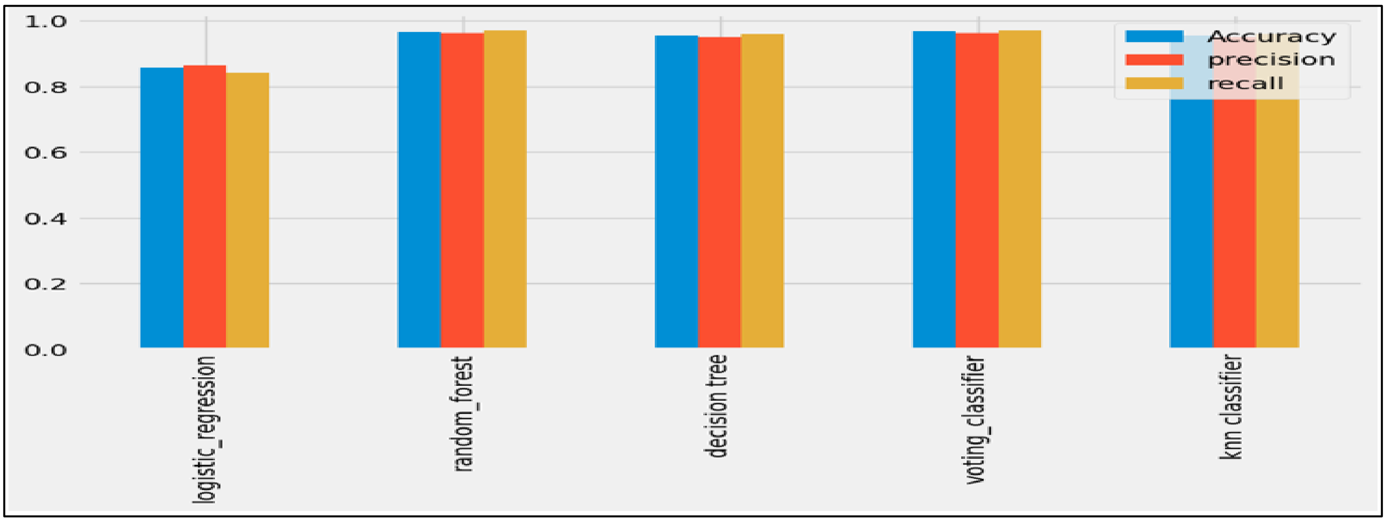

Figure 1 Visualization of Models Result

IV. DISCUSSION

The study was intended to enhance healthcare fraud detection using various machine learning algorithms and optimizing their performance via feature selection and hyperparameter optimization. The research compared the performance of Logistic Regression, Decision Tree, Random Forest, K-Nearest Neighbors (KNN), and a Voting Classifier to determine the most effective method in detecting fraud. Logistic Regression was the most precise model after optimization and had high predictive power because it could identify patterns between features and fraudulent claims.

Decision Tree and Random Forest had good performance but exhibited minor shortcomings in generalization since the ensemble strategy of Random Forest offered superior stability compared to an individual Decision Tree. KNN was largely enhanced by forward feature selection and hyperparameter optimization but lagged slightly behind Logistic Regression in terms of recall and accuracy. The Voting Classifier, which blended diverse models, achieved the best accuracy of 96.68%, reflecting the power of ensemble learning for fraud detection.

Results suggest that feature selection enhanced Logistic Regression and KNN classification accuracy while hyperparameter optimization had a larger impact on the performance of KNN. As much as the tree models scored high, there was always an overfitting tendency with which they performed. This rendered Logistic Regression to be the most reliable method by dint of maintaining equally good scores using different metrics.

V. CONCLUSION

This study illustrates how machine learning can really improve healthcare fraud detection by discovering fraudulent patterns within claims. Logistic Regression was found to be the most effective model, with the advantage of forward feature selection and parameter tuning. The Voting Classifier also performed well by taking the strengths of individual models. Deep learning methods and real-time detection are areas for future research that can further maximize accuracy and efficiency in fraud detection systems.

REFERENCES

[1]. T. Fernando, H. Gammulle, S. Denman, S. Sridharan, and C. Fookes, “Deep learning for medical anomaly detection–a survey,” ACM Computing Surveys (CSUR), vol. 54, no. 7, pp. 1–37, 2021.
[2]. J. M. Johnson and T. M. Khoshgoftaar, “Medicare fraud detection using neural networks,” J Big Data, vol. 6, no. 1, p. 63, 2019.
[3]. L. Syed, S. Jabeen, S. Manimala, and H. A. Elsayed, “Data science algorithms and techniques for smart healthcare using IoT and big data analytics,” Smart techniques for a smarter planet: towards smarter algorithms, pp. 211–241, 2019.
[4]. A. B. Nassif, M. A. Talib, Q. Nasir, and F. M. Dakalbab, “Machine learning for anomaly detection: A systematic review,” Ieee Access, vol. 9, pp. 78658–78700, 2021.
[5]. T. Dissanayake, T. Fernando, S. Denman, S. Sridharan, H. Ghaemmaghami, and C. Fookes, “A robust interpretable deep learning classifier for heart anomaly detection without segmentation,” IEEE J Biomed Health Inform, vol. 25, no. 6, pp. 2162–2171, 2020.
[6]. D. Ververidis and C. Kotropoulos, “Sequential forward feature selection with low computational cost,” in 2005 13th European Signal Processing Conference, IEEE, 2005, pp. 1–4.
[7]. L. Li et al., “A system for massively parallel hyperparameter tuning,” Proceedings of Machine Learning and Systems, vol. 2, pp. 230–246, 2020.
[8]. L. Connelly, “Logistic regression.,” Medsurg Nursing, vol. 29, no. 5, 2020.
[9]. A. Y. B. R. Thaifur, M. A. Maidin, A. I. Sidin, and A. Razak, “How to detect healthcare fraud?‘A systematic review,’” Gac Sanit, vol. 35, pp. S441–S449, 2021.
[10]. B. Mahesh, “Machine learning algorithms-a review,” International Journal of Science and Research (IJSR).[Internet], vol. 9, no. 1, pp. 381–386, 2020.
[11]. Z. Zhang, “Introduction to machine learning: k-nearest neighbors,” Ann Transl Med, vol. 4, no. 11, 2016.