Magna Scientia Advanced Research and Reviews
eISSN: 2582-9394
Cross Ref DOI: 10.30574/msarr
Journal homepage: https://msarr.magnascientiapub.com/

(REVIEW ARTICLE)

# ADVANCING FINANCIAL INTEGRITY IN U.S. HEALTHCARE THROUGH MACHINE LEARNING AND DATA ANALYTICS FOR FRAUD DETECTION

BAKARE Abolore Raliat [1,*], ODEYALE Kehinde Musiliudeen [2] and VISRAM Gabriel [1]

[1] Advanced Data Analytics, University of North Texas, Denton, Texas, USA.
[2] Centre for ICT and Training Services (CITS) Unit, Al-Hikmah University, Ilorin, Kwara State, Nigeria.

ORCID Details
BAKARE Abolore Raliat: https://orcid.org/0009-0004-3433-7813
ODEYALE Kehinde Musiliudeen: https://orcid.org/0009-0004-2462-3752
VISRAM Gabriel: https://orcid.org/https:/0009-0005-6744-8031

Magna Scientia Advanced Research and Review, 2026, 18(01), 074–082

Publication history: Received on 28 July 2026; revised on 07 September 2026; accepted on 09 September 2026

Article DOI: https://doi.org/10.30574/msarr.2026.18.1.0179



## ABSTRACT

The U.S. government has long recognized healthcare fraud as a serious concern and has introduced various measures over several decades to prevent and control fraudulent practices. Despite these efforts, healthcare fraud has become increasingly sophisticated and complex, continuing to create substantial financial pressures on the healthcare system and the wider economy. Recent developments in artificial intelligence (AI), particularly machine learning (ML), natural language processing (NLP), and neural networks, have strengthened the ability to analyze large-scale healthcare datasets and identify concealed patterns, anomalies, and potentially fraudulent activities. Collectively, the implementation of these AI-driven approaches offers considerable potential to safeguard public healthcare resources, reduce financial losses, and strengthen public confidence in the integrity of the U.S. healthcare system.

**Keywords:** Healthcare Fraud; Artificial Intelligence; Machine Learning; Natural Language Processing; Neural Networks; Fraud Detection.

## 1 INTRODUCTION

Healthcare fraud continues to impose a substantial financial burden on the U.S. healthcare system, resulting in billions of dollars in losses each year and placing additional pressure on public healthcare programmes, patients, and businesses (Malveen, 2026). On a broader scale, fraud, waste, and abuse (FWA) contribute to significant global financial losses while potentially affecting the quality of patient care. Traditional rule-based auditing systems may struggle to identify increasingly complex fraudulent activities because they are often rigid and labor-intensive (Patel, 2026).

The rapid digital transformation of healthcare has improved operational efficiency but has also created opportunities for more sophisticated forms of fraud, including phantom billing and claims manipulation. The growing volume and complexity of structured and unstructured healthcare data have further highlighted the limitations of conventional detection methods, increasing the importance of advanced artificial intelligence approaches for identifying suspicious patterns and supporting fraud prevention (Eniola, 2026). However, the adoption of such technologies also introduces

social and ethical concerns, particularly around privacy and potential job displacement, emphasizing the need for responsible implementation that balances technological progress with societal interests (Aqsa et al., 2026).

Healthcare fraud can take several forms, including **ghost billing, overbilling, duplicate billing, and medical identity theft** (Iweriebor, 2023). Unnecessary procedures and fraudulent claims can also place additional financial pressure on Medicare and Medicaid, contribute to rising healthcare and insurance costs, and potentially compromise patient safety through inaccurate medical records or unnecessary treatments (Foglia, 2020). Collectively, these challenges demonstrate the need for more adaptive and data-driven approaches to strengthen fraud detection while protecting healthcare resources and patients.

## 2 LITERATURE REVIEW

### 2.1 Introduction

Healthcare fraud represents a significant challenge to the economic sustainability and operational efficiency of healthcare systems, while also raising concerns about patient safety and public trust. Fraudulent practices, including ghost billing, overbilling, duplicate claims, upcoding, unnecessary medical procedures, and medical identity theft, can generate substantial financial losses and, in some cases, expose patients to inappropriate or unnecessary treatment. As healthcare datasets continue to increase in volume and complexity, conventional rule-based systems and manual auditing techniques may be insufficient for identifying sophisticated and continuously evolving patterns of fraudulent behaviour.

Recent advances in artificial intelligence (AI), particularly machine learning (ML), natural language processing (NLP), anomaly detection, deep learning, and predictive analytics, have created new opportunities for detecting suspicious claims and behavioural patterns within healthcare data. Nevertheless, the successful application of these technologies remains affected by challenges such as data quality, privacy, class imbalance, model interpretability, algorithmic bias, regulatory compliance, and the ability of models to generalise across different healthcare environments. This literature review therefore examines existing AI-driven approaches to healthcare fraud detection, their potential contribution to reducing financial losses and protecting patients, and the challenges associated with achieving scalable and sustainable implementation.

Patel (2026) proposed an optimised **Random Forest-based machine learning framework** designed to address class imbalance within healthcare claims data and identify potentially fraudulent providers. This approach demonstrates how machine learning can support fraud detection while contributing to financial risk reduction, cost efficiency, and improved decision-making.

Similarly, Aqsa et al. (2026) developed a **machine learning approach incorporating demographic characteristics**, particularly age and gender, across merged healthcare claims records. Their approach examined demographic patterns to support feature selection and develop more targeted strategies for identifying potentially fraudulent healthcare claims.

Malveen (2026) employed multiple **machine learning and deep learning algorithms** across large-scale inpatient, outpatient, and beneficiary datasets. The study incorporated SHAP-based model interpretation, automated model retraining, and a real-time detection pipeline, providing an approach that combines fraud detection performance with greater model explainability and adaptability.

Eniola (2026) presented a **real-time hybrid AI architecture** integrating machine learning, deep learning, NLP, and neuro-symbolic reasoning. By analyzing both structured healthcare data and unstructured clinical text, the framework demonstrates the potential value of combining multiple AI techniques to identify complex fraudulent patterns that may be difficult to detect using conventional approaches.

From a broader perspective, Ntiakoh et al. (2026) conducted a **systematic review** examining the role of AI, ML, and NLP in reducing financial fraud within U.S. healthcare systems. Their review emphasized real-world industry applications, human-in-the-loop approaches, and ethical governance, highlighting the importance of maintaining human oversight alongside increasingly automated fraud detection technologies.

Finally, Matloob et al. (2025) proposed an **adaptive, multi-actor healthcare fraud detection framework** combining an association-rule engine with an Anomaly Transformer. The framework examined interactions among doctors, patients, and medical services to identify suspicious relationships and sequence-based anomalies. This approach demonstrates the value of analyzing interconnected healthcare activities rather than evaluating individual claims in isolation.

Overall, the reviewed studies demonstrate a clear movement from traditional rule-based fraud detection towards **adaptive, explainable, real-time, and hybrid AI-driven approaches**. Although the studies differ in their methods and areas of emphasis, they collectively demonstrate the potential of AI and data analytics to improve fraud identification across increasingly complex healthcare datasets. At the same time, issues surrounding explainability, data quality, privacy, governance, model adaptability, and human oversight remain important considerations for the long-term implementation of AI-driven healthcare fraud detection systems.

## 3 COMPARATIVE ANALYSIS OF AI APPROACHES TO HEALTHCARE FRAUD DETECTION

### 3.1 Feature Engineering and Multimodal Data Representation

The reviewed studies demonstrate different approaches to feature engineering and healthcare data representation. Patel (2026) and Malveen (2026) primarily relied on structured financial, reimbursement, and billing attributes, whereas Eniola (2026) extended fraud analysis to unstructured clinical information through NLP-based representations. Matloob et al. (2025) adopted a different approach by transforming transactional healthcare records into longitudinal sequences, allowing relationships among patients, providers, and medical services to be examined over time.

Overall, these approaches indicate a gradual shift from conventional tabular claims analysis towards **multimodal and sequence-based representations**, which may provide a broader understanding of complex fraud patterns.

### 3.2 Class Imbalance Challenges in Healthcare Fraud Analytics

Class imbalance remains an important challenge because fraudulent claims typically represent a relatively small proportion of healthcare transactions. Patel (2026) addressed this issue using **SMOTE-Tomek**, combining synthetic oversampling with data cleaning before training the Random Forest model. In contrast, Matloob et al. (2025) avoided artificial resampling by applying an unsupervised Anomaly Transformer that identifies unusual patterns through anomaly-based attention scoring.

These contrasting approaches demonstrate that class imbalance can be addressed through either **data-level technique**, such as resampling, or **model-level approaches**, such as unsupervised anomaly detection.

### 3.3 Model Interpretability and Complexity in Healthcare Fraud Detection

As fraud-detection models become increasingly sophisticated, interpretability becomes particularly important in healthcare environments where decisions may require regulatory and professional justification. Malveen (2026) addressed this concern through **SHAP-based explanations**, while Eniola (2026) incorporated neuro-symbolic reasoning to combine predictive modelling with healthcare rules and domain knowledge. Ntiakoh et al. (2026), meanwhile, emphasized human auditor oversight as an important safeguard against the risks associated with black-box decision-making.

These findings suggest that improved predictive performance should not be considered independently of **transparency, accountability, and human oversight**, particularly when fraud-detection outcomes may affect providers or patients.

### 3.4 Comparative Model Performance

The studies reported strong but varying levels of predictive performance. Patel's (2026) optimized Random Forest achieved **94.4% accuracy, 93.8% precision, 95.1% recall, and an ROC-AUC of 0.961**. Aqsa et al. (2026) reported approximately **99.48% accuracy** after replacing an overfitted Random Forest model with Gradient Boosting. Matloob et al. (2025) achieved **97.0% accuracy and 91.3% recall** using an Association Rule Engine combined with an Anomaly Transformer.

Malveen's (2026) Stacking Ensemble achieved **92.79% accuracy, 93.63% precision, and an ROC-AUC of 0.9698**, while Eniola's (2026) hybrid neuro-symbolic architecture achieved **94.2% precision, 92.7% recall, and a 93.4% F1-score**, with a reported detection latency of 128 milliseconds per claim.

However, direct comparison of these results should be approached cautiously because the studies used **different datasets, sample sizes, fraud definitions, feature sets, validation procedures, and modelling techniques**. Therefore, a higher reported accuracy does not necessarily indicate that one model would perform better across all healthcare environments.

### 3.5 Adaptability and Real-Time Fraud Detection

Another important distinction among the studies concerns their ability to adapt to evolving fraud patterns. Patel (2026) and Aqsa et al. (2026) primarily employed static, retrospective approaches. Matloob et al. (2025), by contrast, incorporated an adaptive threshold mechanism, while Malveen (2026) proposed automated model retraining within a real-time detection pipeline. Eniola (2026) advanced this further through stream processing and continuous model and rule adaptation. Ntiakoh et al. (2026) emphasised a human-in-the-loop approach in which expert investigators remain involved in model evaluation and fraud investigation.

Collectively, the literature indicates a movement towards **adaptive and near-real-time fraud detection**, rather than relying exclusively on retrospective identification after fraudulent payments have already occurred.

### 3.6 Strengths and Limitations of the Reviewed Approaches

Despite their promising results, each approach presents limitations. Aqsa et al. (2026) initially experienced substantial Random Forest overfitting and relied heavily on demographic variables. Patel's (2026) framework concentrated mainly on provider-level billing information and used synthetic resampling, potentially limiting its ability to capture complex interactions among healthcare actors. Matloob et al. (2025) demonstrated adaptive capabilities but required considerable computational resources and expert oversight for threshold adjustment.

Similarly, Malveen's (2026) analysis was affected by hardware constraints and remained focused on tabular claims data, whereas Eniola's (2026) multimodal architecture introduced greater computational, maintenance, and data-governance requirements. Ntiakoh et al. (2026) further highlighted broader concerns surrounding algorithmic bias, fragmented healthcare IT infrastructure, model transparency, privacy, and the continuing need for human oversight.

### 3.7 Overall Synthesis

Taken together, the reviewed studies demonstrate that there is **no single universally superior model for healthcare fraud detection**. Random Forest and Gradient Boosting provide strong classification capabilities for structured claims data, while anomaly detection and transformer-based approaches offer advantages for identifying unusual or previously unseen behaviours. Ensemble and hybrid neuro-symbolic approaches extend these capabilities by combining multiple models and incorporating structured and unstructured healthcare information.

More importantly, the literature shows that healthcare fraud detection is progressing beyond simple claim classification towards **adaptive, explainable, multimodal, and human-supervised AI systems**. Future development should therefore focus not only on improving predictive accuracy but also on model generalizability, explainability, data quality, computational efficiency, privacy protection, regulatory compliance, and responsible human oversight. These considerations are particularly important for translating experimental fraud-detection models into sustainable solutions for real-world healthcare systems.

## 4 DISCUSSION

### 4.1 Data-Driven Strategies for Detecting, Preventing, and Mitigating Healthcare Fraud in the United States

The reviewed studies demonstrate that data analytics can address healthcare fraud at different levels of the U.S. healthcare system, ranging from individual claims and patient characteristics to provider behavior and wider fraud networks. Although the studies apply different analytical methods and datasets, they collectively demonstrate how data-driven approaches can improve the identification of suspicious patterns, prioritize investigations, and strengthen fraud-prevention strategies.

**Patel (2026)** focused primarily on **provider-level financial risk**, using aggregated billing indicators such as annual inpatient reimbursements and deductible amounts to identify and rank providers with potentially suspicious billing behavior. This approach is particularly useful for audit prioritization because it enables investigators to concentrate resources on providers presenting higher financial risks.

**Aqsa et al. (2026)** examined healthcare fraud from a **patient and claims-demographic perspective**, incorporating characteristics such as age and gender into the analytical process. This approach provides insight into whether particular demographic patterns are associated with fraudulent claims and demonstrates how demographic information can contribute to feature selection and targeted fraud-detection strategies.

**Matloob et al. (2025)** adopted a more detailed **multi-actor analytical approach**, examining interactions among doctors, patients, and medical services across 62 clinical specialties. Their findings indicated that 50% of detected fraud occurred at the patient level, 25% involved doctors, while the remaining cases involved mismatches between services,

patients, and doctors. This relational approach extends fraud detection beyond individual reimbursement anomalies by identifying suspicious interactions and potential scope-of-practice violations.

**Malveen (2026)** analyzed combined inpatient, outpatient, and beneficiary claims data, focusing on claim- and provider-level patterns within a dataset of 558,211 records. The study compared several machine learning approaches and incorporated SHAP analysis to improve the interpretation of fraud predictions. It also proposed automated retraining to improve the model's ability to respond to changing fraud patterns.

**Eniola (2026)** extended healthcare fraud analytics beyond structured claims by adopting a **multimodal approach** that combined billing information with unstructured clinical narratives. The framework used ML, deep learning, NLP, and symbolic reasoning to identify inconsistencies between structured claim information and clinical documentation. This provides a broader analytical perspective because suspicious claims can be evaluated against the clinical information supporting the billed treatment.

At a wider systemic level, **Ntiakoh et al. (2026)** examined AI-driven fraud detection across the U.S. healthcare ecosystem, including Medicare, Medicaid, healthcare providers, billing networks, and public and private payers. Their review highlighted the transition from retrospective **"pay-and-chase" approaches towards pre-payment fraud detection**, while also emphasising human oversight, explainability, and ethical governance.

Among the reviewed approaches, **Matloob et al. (2025)** provides an important example of adaptive fraud analytics because the framework analyses relationships among multiple healthcare actors rather than examining claims independently. Its adaptive threshold mechanism is designed to respond dynamically to changing fraud patterns and concept drift. The model was evaluated using 440,000 hospital transactions covering 62 specialties and benchmarked against MIMIC-IV, demonstrating its potential for identifying complex relational and sequence-based anomalies.

Collectively, these studies demonstrate that effective healthcare fraud detection requires analysis at **multiple levels rather than reliance on a single source of information**. Provider-level financial indicators can identify unusual reimbursement behaviour, demographic information can reveal patterns within claims, multi-actor models can detect suspicious relationships, and NLP can uncover inconsistencies within unstructured clinical documentation.

The evidence therefore suggests that the future of healthcare fraud analytics lies in integrating **structured and unstructured data, adaptive machine learning, anomaly detection, explainable AI, and human expertise**. Such integration could enable healthcare organizations to move beyond detecting fraud after financial losses have occurred towards earlier identification and prevention. However, achieving this transition requires careful consideration of model interpretability, data quality, computational requirements, privacy, regulatory compliance, and continued human oversight.

### 4.2 Discussion Based on Machine Learning

The reviewed studies demonstrate that **machine learning (ML) has become an important component of healthcare fraud detection**, offering greater capacity to identify complex patterns, classify suspicious claims, and support financial risk management. However, the studies differ considerably in their modelling approaches, levels of adaptability, interpretability, and ability to handle evolving fraud patterns.

**Patel (2026)** concluded that the proposed **Random Forest model** provided strong predictive performance across accuracy, precision, recall, and financial risk estimation when compared with the traditional baseline models evaluated in the study. The optimised model achieved 94.4% accuracy, 93.8% precision, 95.1% recall, and an ROC-AUC of 0.961. By identifying providers exhibiting unusual billing and reimbursement patterns, the approach can support healthcare administrators in prioritising audits, identifying financial risks, and allocating investigative resources more efficiently.

**Aqsa et al. (2026)** demonstrated the potential of **Random Forest and Gradient Boosting** for classifying fraudulent healthcare insurance claims. Although the initial Random Forest model achieved high training accuracy, its substantial decline in test performance indicated overfitting. Gradient Boosting subsequently produced stronger generalisation performance, with approximately 99.48% reported test accuracy. Nevertheless, future research could strengthen the approach through more recent and diverse datasets, improved techniques for managing class imbalance, and evaluation of additional models to determine whether performance can be maintained across changing fraud patterns.

**Matloob et al. (2025)** extended conventional classification by integrating machine learning and deep learning within a **multi-actor fraud detection framework**. The combination of an Association Rule Engine and an Anomaly Transformer enabled the identification of suspicious interactions and sequence-based anomalies involving doctors, patients, and medical services. The framework achieved 97.0% accuracy and 91.3% recall when evaluated using 440,000 transactions across 62 medical specialties and benchmarked against MIMIC-IV. Rather than relying exclusively

on individual claims, this approach demonstrates the potential value of examining relationships among multiple healthcare actors when detecting more complex forms of fraud.

**Malveen (2026)** found that the **Stacking Ensemble model** produced the strongest overall predictive performance among the models evaluated, achieving 92.79% accuracy, 93.63% precision, 86.94% recall, a 90.16% F1-score, and an ROC-AUC of 0.9698. The incorporation of SHAP further improved model interpretability by helping explain the features influencing fraud classifications. The proposed real-time detection pipeline and automated retraining framework also demonstrate how machine-learning systems could become more adaptive as fraud patterns change over time.

**Eniola (2026)** presented a broader **hybrid AI approach** combining machine learning, deep learning, NLP, and neuro-symbolic reasoning. Unlike models limited primarily to structured claims, this framework incorporated both structured billing information and unstructured clinical narratives. It achieved 94.2% precision, 92.7% recall, a 93.4% F1-score, and a reported processing latency of 128 milliseconds per claim in the simulated streaming pipeline. The findings demonstrate the potential of multimodal AI for detecting inconsistencies between claims and supporting clinical documentation, although its computational and governance requirements may present challenges for large-scale implementation.

**Ntiakoh et al. (2026)** provided a broader perspective through a systematic review of AI, ML, NLP, predictive analytics, and related approaches within U.S. healthcare fraud detection. The review highlighted a transition from retrospective **"pay-and-chase" auditing towards proactive and pre-payment detection strategies**. However, the authors also identified significant challenges, including algorithmic bias, privacy requirements, fragmented healthcare IT infrastructure, model interpretability, and the need for human oversight. These findings reinforce the argument that AI should function as a **decision-support mechanism rather than a complete replacement for professional judgement and fraud investigators**.

In summary, although some methodological overlap exists within healthcare fraud detection research, the reviewed studies make distinct contributions through different datasets, analytical techniques, levels of automation, and approaches to explainability. Collectively, they demonstrate that ML, deep learning, anomaly detection, NLP, ensemble learning, and hybrid AI can uncover complex patterns and anomalies that may be difficult to identify through conventional rule-based auditing alone.

The growing incorporation of **Explainable Artificial Intelligence (XAI)** is particularly important because predictive performance alone may not be sufficient in a highly regulated healthcare environment. Explainability can help investigators understand why claims or providers have been classified as suspicious, thereby supporting more transparent and defensible decision-making. At the same time, adaptive modelling and automated retraining may help detection systems respond to emerging fraud behaviours.

Nevertheless, stronger predictive performance should not be interpreted as evidence that AI can independently eliminate **Fraud, Waste, and Abuse (FWA)**. Effective implementation requires appropriate model validation, continuous performance monitoring, representative and high-quality data, safeguards against bias and overfitting, regulatory compliance, and meaningful human oversight. Consequently, the long-term value of machine learning in healthcare fraud detection will depend not only on increasingly sophisticated algorithms but also on the development of **reliable, explainable, adaptable, and responsibly governed systems**.

### 4.3 Findings

The reviewed studies identified several important findings regarding the application of Artificial Intelligence and machine learning to healthcare fraud detection. Although each study examined fraud from a different analytical perspective, the findings collectively demonstrate the value of financial, demographic, relational, and multimodal healthcare data in identifying potentially fraudulent activities.

**Patel (2026)** identified **annual inpatient reimbursements, total reimbursement amounts, total claims, and total beneficiaries** as some of the strongest indicators associated with fraudulent provider activity. The study also estimated potential cost-recovery opportunities among high-risk providers, with projected savings exceedingly approximately **$1.4 million to $1.79 million for individual providers**. These findings demonstrate the potential value of provider-level financial analytics for prioritizing audits and directing investigative resources towards cases presenting the greatest financial risk.

**Aqsa et al. (2026)** found through exploratory data analysis that fraudulent claims were relatively evenly distributed between male and female beneficiaries. However, higher frequencies of fraudulent claims were observed among older age groups, which the study associated with longer and more complex treatment patterns. These findings suggest that

demographic variables may contribute useful information to fraud-detection models, although they should be interpreted carefully to avoid inappropriate demographic profiling.

**Matloob et al. (2025)** demonstrated through benchmark testing on the **MIMIC-IV dataset** that the proposed framework could identify rare and suspicious caregiver–procedure sequences. This finding supports the value of analysing healthcare activities as interconnected and sequential events rather than treating individual transactions independently. The study's broader evaluation involved 440,000 transactions across 62 medical specialties.

**Malveen (2026)** identified **Feature 31** as the most influential attribute contributing to fraudulent claim classifications through SHAP-based interpretability analysis. This finding illustrates the value of explainable AI techniques in identifying which variables contribute most strongly to model predictions. However, the practical significance of Feature 31 depends on what the variable represents within the underlying dataset.

**Eniola (2026)** demonstrated the importance of incorporating unstructured healthcare information into fraud detection. Approximately **35% of the simulated dataset consisted of unstructured clinical information**, including physician notes and treatment summaries. The NLP components were incorporated to analyse these narratives alongside structured claims information, enabling the hybrid system to identify inconsistencies that may not be apparent from billing records alone.

At the broader economic level, **Ntiakoh et al. (2026)** reported estimated annual U.S. healthcare fraud-related losses ranging from approximately **$68 billion to $300 billion**. Their review also reported that the CMS Medicare Fraud Prevention System prevented more than **$1.5 billion in improper payments during its initial implementation**, while examining commercial applications involving organisations such as Optum and Cigna. The study highlighted the use of supervised and unsupervised machine-learning techniques alongside human investigator oversight as part of the transition towards proactive fraud prevention.

### 4.4 Overall Findings

Collectively, the findings indicate that healthcare fraud cannot be effectively understood through a single variable, dataset, or analytical technique. **Financial indicators** can reveal abnormal provider behaviour, **demographic analysis** can identify patterns across beneficiary groups, **sequence and relational modelling** can uncover suspicious interactions among healthcare actors, and **NLP and multimodal analysis** can identify inconsistencies between clinical documentation and billing information.

Most importantly, the findings demonstrate a broader transition from retrospective fraud investigation to **predictive, adaptive, explainable, and increasingly real-time fraud detection**. However, the effectiveness of these approaches remains dependent on data quality, model generalizability, appropriate validation, interpretability, regulatory compliance, and human oversight.

## 5 CONCLUSION

The increasing adoption of **Artificial Intelligence (AI), Machine Learning (ML), and Natural Language Processing (NLP)** has significantly changed the approach to healthcare fraud detection and investigation in the United States. These technologies provide the computational capability to analyse large and complex healthcare datasets, allowing investigators to identify subtle anomalies, behavioural patterns, and relationships that may be difficult to detect through conventional manual auditing or rule-based systems. By supporting the identification of suspicious claims and irregular provider, patient, and service activities, AI-driven systems have considerable potential to strengthen the prevention and mitigation of **Fraud, Waste, and Abuse (FWA)**.

The studies reviewed in this research demonstrate that different analytical approaches offer distinct advantages. Random Forest and Gradient Boosting provide effective classification of structured healthcare claims, while anomaly detection and transformer-based models can identify unusual patterns and relationships across healthcare transactions. Ensemble models can combine the strengths of several algorithms, whereas NLP and hybrid neuro-symbolic approaches extend fraud detection to unstructured clinical information. The literature therefore suggests that healthcare fraud detection is gradually progressing from static and retrospective analysis towards **adaptive, explainable, multimodal, and near-real-time approaches**.

Among the machine-learning techniques examined, **Random Forest has demonstrated strong potential for healthcare fraud detection** because it combines multiple decision trees within an ensemble framework. Rather than depending on the prediction of a single decision tree, Random Forest aggregates predictions across multiple trees, which can reduce variance and improve generalisation compared with an individual tree (Schonlau & Zou, 2020). The

findings reviewed in this study also demonstrate its practical application to healthcare claims, particularly for identifying suspicious provider billing and reimbursement patterns.

Another important advantage of Random Forest is its capacity to provide information about **feature importance**, allowing researchers and investigators to examine which variables contribute most strongly to fraud predictions. This capability can support greater understanding and transparency in healthcare fraud analytics and contribute to the broader development of explainable AI approaches (Sabat-Tomala & Raczko, 2020). However, Random Forest should not be regarded as inherently immune to overfitting. Its ensemble structure can reduce the variance associated with individual decision trees, but performance still depends on factors such as model configuration, data quality, class imbalance, feature selection, and appropriate validation.

Overall, the evidence reviewed suggests that **no single machine-learning algorithm provides a complete solution to healthcare fraud**. The effectiveness of fraud detection depends increasingly on combining appropriate analytical methods with high-quality and representative data, model explainability, continuous validation, regulatory compliance, and human expertise. Explainable Artificial Intelligence (XAI) is particularly important because healthcare investigators and decision-makers need not only to identify potentially fraudulent activities but also to understand the factors contributing to those predictions.

Consequently, the long-term contribution of AI to healthcare fraud prevention will depend on developing systems that are not only accurate but also **transparent, adaptable, scalable, and responsibly governed**. When supported by appropriate human oversight, these technologies can strengthen fraud investigations, improve the allocation of auditing resources, reduce financial losses, protect public healthcare funds, and contribute to greater accountability and confidence in the integrity of the U.S. healthcare system.

## STATEMENTS ON COMPLIANCE WITH ETHICAL STANDARDS

### *Disclosure of conflict of interest*

The authors declare no conflicts of interest regarding this manuscript.

## REFERENCES

[1] Aqsa, A., D'silva, N. S., & Rahaman, S. (2026). Leveraging machine learning in detection of healthcare insurance fraud. In J. Shreyas, G. H. L., S. Rahaman, K. Kaushik, A. Chaudhary, & D. P. (Eds.), Data science and exploration in artificial intelligence (pp. 220–230). Springer. https://doi.org/10.1007/978-3-032-19321-6_20

[2] Eniola, A. (2026). Hybrid AI models for real-time fraud detection in healthcare insurance systems [Unpublished manuscript].

[3] Foglia, B. (2020). Mitigating risk in health care through mandatory enterprise risk management programs. Loyola University Chicago Journal of Regulatory Compliance, 5, 78.

[4] Iweriebor, L. E. (2023). Approach to Medicare provider fraud detection and prevention [Doctoral dissertation, Capitol Technology University].

[5] Kapadiya, K., Patel, U., Gupta, R., Alshehri, M. D., Tanwar, S., Sharma, G., & Bokoro, P. N. (2022). Blockchain and AI-empowered healthcare insurance fraud detection: Analysis, architecture, and future prospects. IEEE Access, 10, 79606–79627.

[6] Malveen, Z. (2026). Healthcare insurance fraud detection using AI [Unpublished manuscript].

[7] Matloob, I., Khan, S., Rukaiya, R., Alfraihi, H., & Khan, J. A. (2025). Healthcare fraud detection using adaptive learning and deep learning techniques. Evolving Systems, 16(2), Article 72. https://doi.org/10.1007/s12530-025-09698-6

[8] Ntiakoh, A., Ocansey, I. T., & Amoakoh, C. (2026). The role of artificial intelligence in reducing financial fraud in U.S. healthcare systems. Magna Scientia Advanced Research and Reviews, 17(2), 128–134. https://doi.org/10.30574/msarr.2026.17.2.0130

[9] Patel, J. (2026). Implementing Random Forest method for healthcare provider fraud detection framework to mitigate financial risk and cost optimization in healthcare management [Preprint]. Research Square. https://doi.org/10.21203/rs.3.rs-10084737/v1

[10] Sabat-Tomala, A., Raczko, E., & Zagajewski, B. (2020). Comparison of support vector machine and Random Forest algorithms for invasive and expansive species classification using airborne hyperspectral data. Remote Sensing, 12(3), Article 516.

[11] https://doi.org/10.3390/rs12030516

[12] Schonlau, M., & Zou, R. Y. (2020). The random forest algorithm for statistical learning. The Stata Journal, 20(1), 3–29.

[13] https://doi.org/10.1177/1536867X20909688