RADemics

Healthcare Fraud Detection Using Data Mining and Machine Learning Techniques

Krutika Balram Kakpure, M. Chiranjivi
JSPM'S Jaywantrao Sawant College of Engineering,
Hyderabad Institute of Technology and Management

Healthcare Fraud Detection Using Data Mining and Machine Learning Techniques

[1]Krutika Balram Kakpure, JSPM'S Jaywantrao Sawant College of Engineering Hadpsar Pune Affiliated to Savitribai Phule Pune University, Maharashtra, India. krutika31.kakpure@gmail.com

[2]M. Chiranjivi, Associate Professor, Department of EEE, Hyderabad Institute of Technology and Management, Hyderabad, Telangana, India. chiranjivimadduluri@gmail.com

Abstract

Rapid digital transformation within healthcare infrastructures has significantly increased the complexity and volume of healthcare transactions, creating critical challenges associated with fraudulent insurance claims, abnormal billing practices, prescription misuse, identity manipulation, and financial exploitation within healthcare systems. Traditional fraud detection approaches based on manual auditing and static rule-based mechanisms encounter substantial limitations while processing large-scale healthcare datasets containing heterogeneous and continuously evolving fraud patterns. Advanced data mining and machine learning techniques provide intelligent, scalable, and automated solutions capable of identifying hidden fraudulent activities through predictive analytics, anomaly detection, classification models, clustering algorithms, and deep learning architectures. This book chapter presents a comprehensive investigation of healthcare fraud detection frameworks using data mining and machine learning methodologies within modern healthcare environments. Detailed discussion includes healthcare fraud classification, preprocessing strategies, feature engineering methods, imbalance handling techniques, predictive fraud analytics, scalable big data frameworks, and comparative analysis of machine learning algorithms including Decision Trees, Random Forests, Support Vector Machines, Neural Networks, and deep learning models. Critical challenges associated with healthcare fraud analytics such as data privacy, real-time fraud monitoring, interpretability limitations, computational scalability, and heterogeneous healthcare data processing receive significant attention throughout the chapter. Emerging technologies including explainable artificial intelligence, federated learning, blockchain-enabled healthcare security, cloud computing, and big data analytics frameworks also receive analytical consideration for improving intelligent fraud prevention systems. Integration of advanced analytical frameworks with healthcare security infrastructures strengthens fraud prediction capability, enhances operational transparency, reduces financial losses, and supports secure healthcare governance across digital healthcare ecosystems. The presented discussion contributes valuable insights for researchers, healthcare professionals, policymakers, and data scientists engaged in development of intelligent healthcare fraud detection and prevention frameworks for next-generation healthcare systems.

Keywords: Healthcare Fraud Detection, Data Mining, Machine Learning, Predictive Analytics, Big Data Analytics, Deep Learning

Introduction

Rapid technological advancements within healthcare infrastructures have transformed healthcare management systems through integration of electronic health records, insurance claim processing platforms, telemedicine services, cloud-based healthcare applications, and digital billing frameworks [1]. Continuous growth of healthcare data generated from hospitals, pharmacies, diagnostic laboratories, insurance organizations, and government healthcare programs has created new opportunities for efficient healthcare administration and intelligent medical decision-making [2]. Simultaneously, increasing digitalization within healthcare ecosystems has intensified security vulnerabilities associated with fraudulent healthcare activities, unauthorized insurance claims, prescription misuse, identity manipulation, and financial exploitation [3]. Healthcare fraud creates substantial economic burdens for healthcare organizations, insurance providers, and patients through unnecessary financial losses and resource misallocation [4]. Large-scale healthcare fraud activities also negatively influence healthcare service quality, operational transparency, and patient trust within healthcare environments. Growing complexity of healthcare transactions therefore demands advanced analytical systems capable of identifying suspicious healthcare activities accurately and efficiently within massive healthcare databases and distributed healthcare operational networks [5].

Traditional healthcare fraud detection mechanisms primarily depend upon manual auditing procedures, predefined business rules, statistical verification methods, and investigator-driven claim analysis processes [6]. Conventional fraud detection systems perform adequately for simple fraudulent activities involving duplicate billing, excessive reimbursement claims, and coding irregularities [7]. Increasing sophistication of fraudulent healthcare operations significantly reduces effectiveness of static rule-based detection frameworks because fraudsters continuously modify fraudulent strategies to bypass existing healthcare security controls [8]. Large healthcare datasets containing millions of transactions create additional difficulties for manual investigation processes due to computational complexity, operational delays, and limited scalability. Static analytical methods frequently produce high false-positive rates and inaccurate fraud prediction outcomes during real-world healthcare monitoring activities [9]. Continuous evolution of healthcare fraud patterns therefore requires intelligent analytical models capable of adaptive learning, predictive analysis, and automated healthcare transaction monitoring [10]. Modern healthcare fraud prevention systems increasingly depend upon computational intelligence techniques for efficient analysis of dynamic healthcare operational environments.