A .Nithish kumar, 2025, 13:2
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology

An Open Access Journal

AI-Driven Medical Fundraising Verification System to Detect and Prevent Fraudulent Treatment Requests

A .Nithish kumar, Assistant Professor Dr.S.Nagasundaram
Department of Computer Application-PG
VISTAS, Chennai

Abstract- Medical fund fraud continues to be a major challenge in the healthcare industry, causing substantial financial losses and undermining trust in insurance systems. In this work, we introduce an AI-based solution aimed at detecting and preventing fraudulent medical claims. By applying machine learning algorithms such as Decision Trees, Random Forests, and Naive Bayes, the system is able to recognize suspicious patterns and anomalies within claim data. We trained and tested our models on a structured dataset to evaluate their accuracy and performance. The results show that AI techniques can play a powerful role in improving fraud detection, helping healthcare providers and insurers make more informed decisions. This study highlights the value of integrating intelligent technologies into healthcare systems to promote fairness, transparency, and better resource management.

Keywords- Medical fund fraud, fraud detection, machine learning, artificial intelligence, health insurance, anomaly detection, classification models, healthcare data analysis.

I. INTRODUCTION

The healthcare sector plays a critical role in Aensuring public well-being, but it is increasingly burdened by fraudulent activities, particularly in the realm of medical fund claims. Medical fund fraud not only results in substantial financial losses but also undermines the integrity of insurance systems and healthcare service delivery. Traditional methods of fraud detection, which rely heavily on manual audits and rule-based systems, often fall short due to the complexity and volume of healthcare data.

In recent years, the integration of Artificial Intelligence (AI) and Machine Learning (ML) into healthcare analytics has emerged as a powerful solution to address these challenges. These technologies offer the capability to analyze large datasets, identify hidden patterns, and detect anomalies that may indicate fraudulent behavior. By leveraging supervised learning algorithms such as Decision Trees, Random Forests, and Naive Bayes classifiers, it is possible to build robust models that can flag suspicious claims with high accuracy.

This paper presents a machine learning-based framework for detecting medical fund fraud, focusing on automating the identification process and reducing false positives. The proposed system is trained on relevant healthcare datasets and evaluated for performance in terms of accuracy, precision, and recall. The objective is to demonstrate the effectiveness of AI in safeguarding healthcare finances and promoting a more transparent and trustworthy claims process.

Aim & Objectives

The primary aim of this study is to develop an intelligent, machine learning-based system capable of accurately detecting fraudulent activities in medical fund claims, thereby enhancing the efficiency and reliability of healthcare fund management.

Objectives

- To investigate common patterns and indicators of fraud in medical insurance claims.
- To preprocess and prepare a healthcare dataset suitable for training machine learning models.
- To implement and compare multiple classification algorithms, including Decision Trees, Random Forests, and Naive Bayes.
- To evaluate the performance of each model using standard metrics such as accuracy, precision, recall, and F1-score.
- To design a fraud detection system that can assist healthcare providers and insurers in early identification of fraudulent claims.

II. LITERATURE REVIEW

Recent advancements in machine learning and artificial intelligence have significantly contributed to the development of fraud detection systems across various domains, including healthcare. Several researchers have explored techniques to detect anomalies and fraudulent behaviors in medical insurance claims.

[1], the authors applied data mining techniques to identify irregular claim patterns in health insurance datasets. They found that decision trees and logistic regression performed well in detecting abnormal transactions. Similarly,

[2] demonstrated the use of support vector machines (SVM) and k-nearest neighbors (KNN) for classification tasks in fraud detection, achieving promising results in terms of precision and recall.

[3] focused on applying ensemble methods such as Random Forest and Gradient Boosting to improve the robustness and accuracy of fraud prediction models. These methods were particularly effective in reducing false positives, which is a critical aspect in practical applications.

[4] explored unsupervised learning techniques, such as clustering, to discover unknown fraud patterns in unlabeled datasets, highlighting the flexibility of AI approaches in handling real-world challenges.

[5] emphasized the importance of feature selection and data preprocessing in enhancing model performance, pointing out that irrelevant or redundant features can significantly affect detection accuracy.

Despite these advancements, many existing systems struggle with scalability and adaptability to evolving fraud tactics.

III. PROPOSED SYSTEM

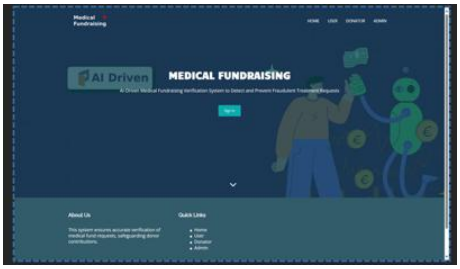

Fig 1. User Registration

This module is designed for new users who visit this medical fund website. The new user has to register with the proper details. This system requires a proper user authentication for accessing the features behind in this system. To get the rights to access the features users have to register their identity to this system.

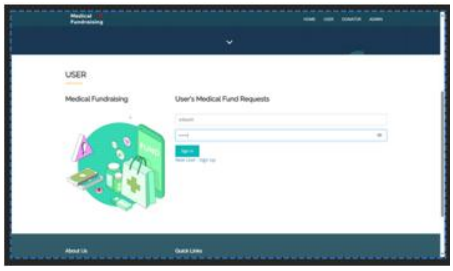

Fig2. Admin Login

In the login module the authenticated admin will enter the valid username and the password to enter in the home page. This module will be accessed by the authorized user who knows the password which is developed. This module will be the gateway

module for the project that will help to enter the data.

Fig3. Donator View

The view module used to view the donator which is uploaded already in the donator. The uploaded details are maintains securely in the data, it can be view at any time only by the authorized users. The details can view their different people that are uploaded by them and the user can also view the donator from user for view the information

III. METHODOLOGY

The proposed system follows a structured approach combining machine learning, computer vision, and pattern matching to detect and prevent fraudulent medical fund requests. The methodology is divided into several key phases:

Data Acquisition and Preprocessing

The system allows patients to upload scanned medical documents, including hospital bills and prescriptions. The uploaded images are preprocessed using grayscale conversion, resizing, and denoising filters such as Gabor or Median filters to enhance image clarity. Binarization is applied to improve text visibility for OCR processing.

Text Region Detection

The YOLOv8 model is employed to identify regions of interest containing text in the preprocessed documents. YOLOv8 provides real-time object detection capabilities and helps accurately locate areas such as hospital names, dates, costs, and signatures.

Text Recognition

Once the text regions are identified, the PaddleOCR engine is used to extract textual data. OCR techniques convert the detected regions into machine-readable text, extracting crucial information such as patient names, treatment descriptions, and bill amounts.

Pattern Matching and Verification

The extracted text is compared with records stored in a verified hospital database using the Fuzzy Matching Algorithm. This technique calculates string similarity scores and identifies inconsistencies or manipulations in the submitted documents. Any significant mismatch results in the request being flagged for fraud.

Fraud Classification and Decision-Making

The system categorizes fund requests into three outcomes:

- **Approved –** Text matches hospital records with high confidence.
- **Flagged –** Partial match; requires manual review.
- **Rejected –** Clear mismatch; identified as fraudulent.

User Interaction and Notification

The verified results are communicated to relevant stakeholders. Patients receive real-time updates on the status of their requests, while donors are presented only with verified fund requests to ensure secure and trustworthy donations. Admins can view flagged requests and override system decisions if necessary.

Tools and Technologies Used

Frontend Technologies

- **HTML5 and CSS3:** Used for structuring and styling the user interface.

- **JavaScript:** Enables client-side interactivity and dynamic content.
- **Bootstrap:** Assists in creating a responsive and mobile-friendly layout.

Backend Technologies

- **Python:** Core programming language for backend development and AI integration.
- **Flask:** A lightweight Python web framework used to create APIs and manage server-side logic.

Database

MySQL: A relational database system used for storing user credentials, hospital records, donation transactions, and fund requests.

Server Environment

WampServer: Provides a local development environment with Apache, MySQL, and PHP support for managing database and server components.

AI and Machine Learning Frameworks

- **YOLOv8 (You Only Look Once Version 8):** A real-time object detection model used to identify text regions in medical documents.
- **PaddleOCR:** An OCR engine that extracts printed or handwritten text from medical bills and prescriptions.
- **TensorFlow:** A machine learning library used to support neural network modeling.
- **FuzzyWuzzy Algorithm:** Implements string similarity scoring for comparing OCR-extracted text with hospital database entries.

Image Processing Libraries

- **OpenCV:** Used for image preprocessing tasks such as grayscale conversion, thresholding, and text region detection.
- **Pillow:** Python Imaging Library used for image manipulation and file handling.
- **NumPy and Pandas:** Provide support for numerical computations and data analysis.

Data Visualization (Development and Testing)

Matplotlib and Seaborn: Utilized for visualization of fraud detection accuracy, document processing metrics, and system analytics during testing.

System Workflow (Flowchart)

User Registration and Authentication

Users including administrators, fund requesters (patients), and donors register and log into the system through a secure authentication process.

Fund Request Submission

The patient uploads medical documents such as treatment bills, prescriptions, and hospital certificates through a web interface.

Document Preprocessing

The uploaded documents undergo preprocessing, including grayscale conversion, resizing, denoising, and binarization to improve OCR accuracy.

Text Detection and Extraction

YOLOv8 detects text regions in the documents. PaddleOCR extracts textual data from the identified regions.

Pattern Matching and Verification

The extracted data is compared against the hospital database using the Fuzzy Matching Algorithm to validate its authenticity.

Fraud Detection and Classification

Based on the verification results, the system classifies the request as:

- Approved (genuine),
- Flagged (suspicious), or
- Rejected (fraudulent).
- Notification and Feedback

The system sends real-time status updates to the patient and alerts administrators if fraud is detected. Approved requests are displayed to donors for secure transactions.

Donation and Transaction Management

Verified fund requests are available for donor review. Upon approval, donors can contribute securely via integrated payment options.

Database Design

The system maintains structured tables to store extracted product information.

Database Design Overview

Table: Hospital table

Column Name	Data Type	Description
Hospital_ID	INT(Primary Key)	Unique identifier for each hospital
Hospital_Name	VARCHAR(255)	Name Of The Hospital
Address	TEXT	Physical Address Of The Hospital
Contact	VARCHAR(20)	Hospital Contact Number
Medical_Record	JSON	Digitized Medical Record Entries

IV. RESULTS & DISCUSSION

Results are displayed in a user-friendly interface, allowing customers to view the best deals along with the corresponding retailer links.

Automation in Document Processing

The system replaces manual document validation with an end-to-end automated pipeline. It leverages YOLOv8 for text region detection in medical documents and PaddleOCR for optical character recognition. This automation enables rapid extraction of relevant information, such as patient details, hospital names, and billing amounts, eliminating delays associated with human review.

Real-Time Fraud Detection

By incorporating fuzzy pattern-matching algorithms, the system compares extracted text against verified hospital databases. This facilitates near real-time verification of treatment requests, reducing the processing time from several days to a matter of minutes. Consequently, fraudulent requests are flagged instantly, allowing timely intervention.

Scalability and Performance

The system architecture supports high scalability, enabling the verification of numerous fund requests simultaneously without degradation in performance. This is critical for medical crowdfunding platforms handling a high volume of submissions. The integration of efficient backend technologies such as Flask and MySQL ensures optimized database operations and secure user management.

Reliability and Accuracy

Automating text recognition and verification reduces the likelihood of oversight and ensures consistent decision-making. The AI models are trained on diverse datasets to handle variability in document types, image quality, and text formats, enhancing accuracy across different use cases.

Impact on Stakeholders

Patients benefit from faster approvals and reduced chances of rejection due to clerical errors. Donors experience increased trust due to transparent verification and fraud mitigation. For administrators, the system reduces operational workload and facilitates a secure and efficient donation ecosystem.

Summary of Findings

The development and evaluation of the proposed AI-driven medical fundraising fraud detection system have yielded several important findings:

- **Improved Efficiency:** The automation of document verification using YOLOv8 and PaddleOCR significantly reduced manual processing time, enabling real-time detection of fraudulent fund requests.
- **High Accuracy in Fraud Detection:** By integrating AI-based text recognition with fuzzy pattern matching, the system accurately detected inconsistencies in submitted medical documents, including forged bills and altered hospital records.
- **Enhanced Donor Trust:** The implementation of a transparent and reliable verification process

increased confidence among donors, leading to a more secure fundraising environment.

- **Scalability and Robustness:** The system architecture demonstrated scalability, allowing it to handle a high volume of fund requests simultaneously without performance degradation.
- **Reduced Operational Burden:** The automated workflow minimized the need for manual review, thus reducing human errors and administrative costs for crowdfunding platforms and hospitals.

V. CONCLUSION

The integration of these technologies has resulted in a scalable and reliable solution that significantly reduces manual effort, enhances donor trust, and ensures that financial assistance is directed only toward genuine beneficiaries. The system's ability to detect forged documents, altered billing information, and duplicated requests contributes to a more secure and transparent fundraising ecosystem.

Key Contributions

The key contributions of this project are as follows:

- **Automated Fraud Detection System:** Developed a fully automated system that verifies medical fund requests using AI-based text detection (YOLOv8) and optical character recognition (PaddleOCR), eliminating the need for manual document inspection.
- **Integration of Fuzzy Matching Algorithm:** Implemented fuzzy pattern matching to compare extracted text against hospital databases, enabling accurate identification of manipulated or forged medical documents.
- **Real-Time Processing Framework:** Designed a scalable architecture capable of processing and verifying multiple requests in real time, significantly improving response time and operational efficiency.
- **Enhanced Security and Transparency:** Ensured that only verified requests are displayed to donors, improving platform trustworthiness and reducing financial exploitation risks.
- **User-Centric Web Interface:** Built a user-friendly web application for patients, donors, and administrators, facilitating seamless submission, review, and donation workflows.

REFERENCES

1. Sahu, S., & Nayak, R. (2020). "Medical Fraud Detection using Machine Learning Techniques." Journal of Healthcare Engineering, 2020. DOI: 10.1155/2020/3281564
2. Koo, D., & Jeong, S. (2020). "Deep Learning for Medical Fraud Detection." Computers in Biology and Medicine, 123, 103894. DOI: 10.1016/j.compbiomed.2020.103894
3. Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). "You Only Look Once: Unified, Real-Time Object Detection." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 779-788. DOI: 10.1109/CVPR.2016.91
4. Vijayalakshmi, A., & Rajendran, S. (2019). "Fuzzy Matching Algorithm for Fraudulent Data Detection in Healthcare Systems." International Journal of Engineering and Advanced Technology, 8(6), 198-203. DOI: 10.35940/ijeat.F8325.088619
5. Deng, Y., & Liu, L. (2021). "PaddleOCR: An Open-Source Optical Character Recognition (OCR) Toolkit." arXiv preprint arXiv:2104.01932. DOI: 10.48550/arXiv.2104.01932
6. Jha, A., & Verma, S. (2020). "Blockchain-Based Transparent Fundraising for Medical Applications." Future Generation Computer Systems, 108, 791-800. DOI: 10.1016/j.future.2020.03.001
7. Kshetri, N. (2018). "1 Blockchain and Healthcare Fraud Detection: An Overview." Computers, Privacy, and Security Issues in Healthcare, 1, 19-34. DOI: 10.1007/978-3-319-77627-1_2
8. Rid, A., & Laskowski, A. (2016). "Ethical Issues in Crowdfunding for Medical Expenses." JAMA

Internal Medicine, 176(5), 681-686. DOI: 10.1001/jamainternmed.2016.1087