International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 06 | June – 2025 DOI: 10.55041/ISJEM04277
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

AI-DRIVEN MEDICAL FUNDRAISING VERIFICATION SYSTEM TO DETECT AND PREVENT FRAUDULENT TREATMENT REQUESTS

M. Vasuki[1], Dr. T. Amalraj Victoire[2], I. Charanya[3]
[1]Associate Professor, Department of Computer Applications, Sri Manakula Vinayagar Engineering College (Autonomous), Puducherry 605107, India,
dheshna@gmail.com
[2]Professor, Department of Computer Applications, Sri Manakula Vinayagar Engineering College (Autonomous), Puducherry 605107, India,
amalrajvictoire@gmail.com
[3]Post Graduate student, Department of Computer Applications, Sri Manakula Vinayagar Engineering College (Autonomous), Puducherry 605107, India
charanyairisappin@gmail.com

ABSTRACT

Medical fundraising is a crucial source of financial aid for individuals requiring support for treatments, surgeries, or emergencies. Crowdfunding platforms and charitable organizations play a vital role in raising funds, but fraudulent activities such as fabricating treatment bills have become a significant challenge. Fraudulent requests erode donor trust and divert resources from genuine beneficiaries. Existing fraud detection methods, often manual or semi-automated, are inefficient, error-prone, and struggle to identify sophisticated fraud attempts.

This project introduces an AI-driven system to detect and prevent fraudulent medical fund requests. The system integrates YOLOv8 for detecting text in uploaded treatment bills and PaddleOCR for text extraction. Extracted details such as hospital names, patient data, and treatment costs—are validated against a trusted hospital dataset using the Fuzzy Matching Algorithm, ensuring accurate verification by identifying discrepancies between records. By automating text detection, recognition, and pattern matching, this system enhances verification accuracy, safeguards donor contributions, and fosters trust in medical crowdfunding.

Keywords:

Fraud Detection, Medical Fundraising, YOLOv8, PaddleOCR, Fuzzy Matching Algorithm.

INTRODUCTION

The AI-Driven Medical Fundraising Verification System is a ground-breaking initiative designed to safeguard the ethics of philanthropic contributions in the medical field. In the current digital era, medical crowdfunding has grown in popularity as a means for people to look for funds for necessary treatments. But with its growth comes the possibility of fraudulent campaigns, in which dishonest people fabricate treatment records to exploit supporters' generosity. This not only undermines public trust but also diverts funds from genuine patients in need.

The project is divided into three main stages: fraud detection, verification, and data extraction. In the data extraction stage, YOLOv8 scans submitted medical records to locate relevant text sections. PaddleOCR then extracts key data such as hospital names, treatment prices, and patient details. During the verification stage, the extracted data is compared to a trusted hospital database using fuzzy matching to assess authenticity.

This artificial intelligence (AI) technology provides an effective fraud protection during emergencies when prompt medical assistance is essential. It improves platform confidence while guaranteeing donor monies reach legitimate recipients. It improves accuracy and lessens manual labor, which makes it more effective in combatting false medical requests.

RELATED WORK

Recent developments in artificial intelligence (AI) and machine learning have demonstrated encouraging uses in fraud detection systems, especially in the financial and medical sectors. Conventional medical funding verification methods mostly rely on manual review, which is laborious and prone to mistakes. Using OCR and pattern recognition, some researchers have investigated automated fraud detection; nevertheless, these systems frequently fall short when confronted with complex document modifications.

Research on document verification has employed rule-based anomaly detection and typical OCR engines, however these methods have trouble detecting low-quality pictures or modified formats. Though they often focus on organized papers with neat layouts, efforts like those found on data science sites like Kaggle have investigated receipt and document parsing, which contrasts with the varied and frequently irregular nature of medical bills.

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 06 | June – 2025 DOI: 10.55041/ISJEM04277
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

In order to overcome these constraints, more recent systems have started combining sophisticated OCR frameworks like PaddleOCR with deep learning-based object recognition (such YOLO models). Combining these technologies can increase the accuracy of text extraction, particularly from manipulated or noisy photos, according to research. Fuzzy string matching has also been successful in comparing extracted text with validated databases, allowing for the detection of large fraud indications while tolerating small differences. Despite these developments, a number of obstacles still exist. First, large-scale deployment is limited by the absence of a consolidated database of hospital records. Second, counterfeit signatures and stamps, which are frequently used in fraudulent medical fundraising, are not taken into consideration by many current systems. In conclusion, the majority of academic implementations do not address scalability and donor trust.

The following are some current suggestions to bridge these gaps:

1. Including real-time AI models for anomaly detection and document preparation.

2. Verification via secure APIs with authentic hospital records.

3. The use of automated processes that permit manual examination of situations that have been highlighted but minimize human interaction.

By integrating fuzzy matching approaches, PaddleOCR for identification, and YOLOv8 for text detection, the suggested system expands on these foundations to produce a scalable, donor-trust-enhancing framework that more accurately and efficiently detects and stops fraudulent medical fund requests.

RESEARCH METHODOLOGY

This project verifies submitted treatment papers using artificial intelligence (AI) to identify and stop fraudulent medical fund claims. By combining machine learning and computer vision technology, the technique focuses on automating the fraud detection process. Developed using Python (Flask framework), the system links to medical records for validation. It integrates an organized pipeline that consists of pattern matching, text detection, text recognition, and document preparation. Real-time text region identification in treatment-related documents, including prescriptions and medical invoices, is accomplished by the system using YOLOv8. PaddleOCR is used for high-accuracy text recognition, extracting important data such as patient information, hospital names, and treatment expenses. To confirm validity, the retrieved data is verified with reliable hospital databases using the Fuzzy Matching Algorithm. Requests that include inconsistencies are either refused or reported. The system uses these technologies to provide medical fundraising platforms with an automated, transparent, and scalable alternative to manual verification approaches.

ALGORITHM DETAILS

Text Region Detection (YOLOv8)

Steps:

1. Start by preprocessing the uploaded document.
2. Use YOLOv8 to identify text areas within the image.
3. Detect stamps, signatures, printed content, and layout structure in the bill or prescription.
4. Pass these detected regions to the OCR module.

Text Recognition (PaddleOCR)

Steps:

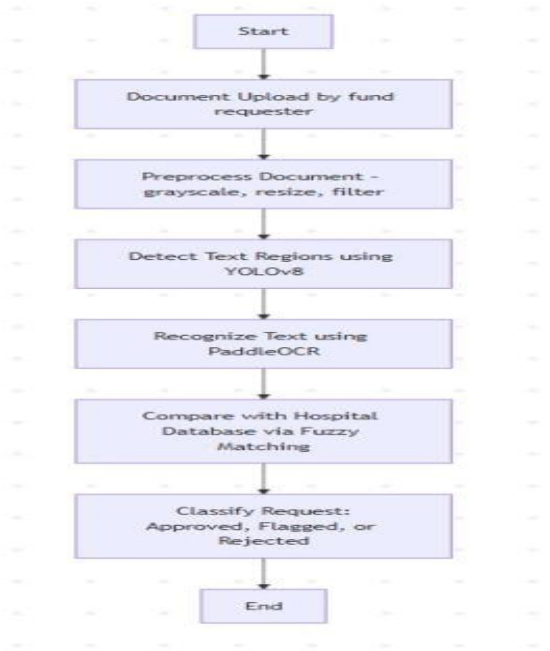
1. Extract text from the identified regions using OCR.
2. Recognize hospital names, patient IDs, billing dates, costs, and other structured data.
3. Convert image-based information into machine-readable text.
4. Organize extracted fields for pattern matching.

Pattern Matching (Fuzzy Matching Algorithm)

Steps:

1. Retrieve corresponding records from the hospital database.
2. Apply fuzzy string comparison between OCR-extracted text and database values.
3. If similarity scores fall below a defined threshold, flag the request as suspicious.
4. If scores are high and consistent across fields, mark the request as genuine.

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 06 | June – 2025 DOI: 10.55041/ISJEM04277
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata



WORKING PROCESS

The steps listed below can be used to explain the working process:

Step-1: Start the document verification process.

Step-2: Upload the document by the fund requester.

Step-3: Preprocess the uploaded document by converting it to grayscale, resizing it, and applying filters.

Step-4: Detect text regions in the document using the YOLOv8 object detection model.

Step-5: Recognize the detected text using the PaddleOCR engine.

Step-6: Compare the recognized text with entries in the hospital database using fuzzy matching.

Step-7: Classify the fund request as Approved, Flagged, or Rejected based on the comparison.

Step-8: End.

Basic Data Exploration

Prescriptions, treatment invoices, and patient records are among the scanned medical papers in the collection. Hospital stamps, patient names, treatment expenses, and signatures are among the textual and visual information included in these records. While some of these components are obviously printed, others could be handwritten or altered digitally. To find areas with pertinent data for fraud verification, the first step in the investigation process is to look at the format and content of these papers.

Feature Engineering

AI algorithms are used to extract important information from medical records. Important details including the hospital name, patient details, treatment type, and billing amounts are extracted from the identified text sections. Additional characteristics, such as the existence of official hospital signatures or stamps, are designed to aid in determining authenticity. This step lays the groundwork for comparing the retrieved data with validated medical records.

Explanatory Data Analysis (EDA)

Using visual analysis tools like tables and bar graphs (via libraries like Matplotlib and Seaborn), prevalent fraud tendencies are investigated. For example, patterns based on hospital names or bill amounts can be found in the distribution of flagged vs legitimate money requests. These insights aid in the identification of traits, such as extremely high treatment expenses or hospital names that don't match, that are commonly seen in fraudulent instances.

Data Preprocessing

Preprocessing includes applying noise-reduction filters such as Gabor or Median filters, scaling, and transforming image-based medical records to grayscale. Next, for clarity, text sections are binarized. Text is identified and transformed into machine-readable forms using OCR (Tesseract or EasyOCR). In order to facilitate additional analysis and model training, these retrieved characteristics are encoded into structured representations.

Modelling

Machine learning models are used after preprocessing to identify fraudulent patterns. Text areas are detected using the YOLOv8 algorithm, and textual data is extracted via OCR. Authentic hospital records are then compared with this text using a fuzzy matching method. The model uses this comparison to categorize money requests into groups like approved, flagged, and rejected. This method assists in identifying medical document falsification and automates the verification process.

Website

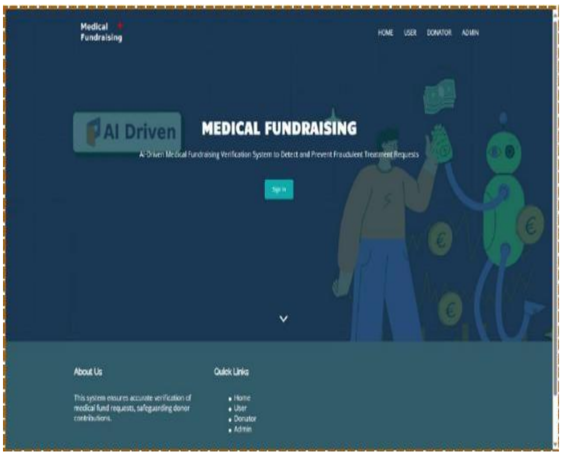
Bootstrap, HTML, CSS, JavaScript, and Python (Flask) are used in the development of the online application. For administrators, donors, and patients, it offers an easy-to-use interface. Patients may submit supporting documentation and check the status of their petitions, while donors can peruse and approve verified money requests. Additionally, to expedite verification, the technology interfaces with hospital databases.

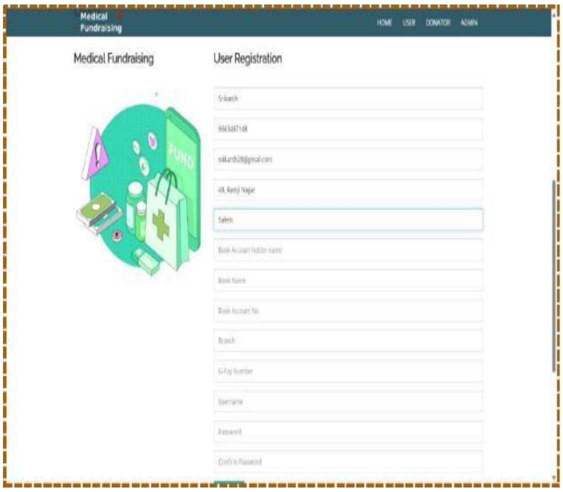
International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 06 | June – 2025 DOI: 10.55041/ISJEM04277
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

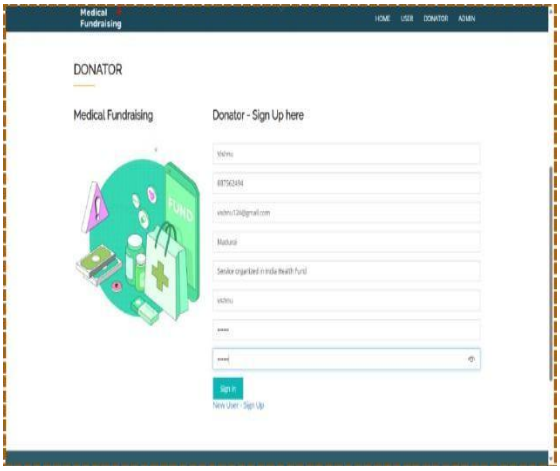
Deployment

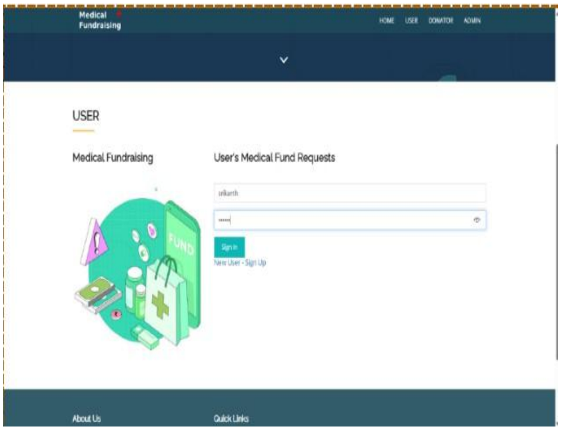
With capabilities like automatic document verification and fraud detection, the system is presently up and running. Multilingual OCR, blockchain integration for tamper-proof data, and donor trust ranking to promote safer contributions are some of the upcoming improvements. The purpose of these modifications is to increase medical crowdfunding's dependability and openness.

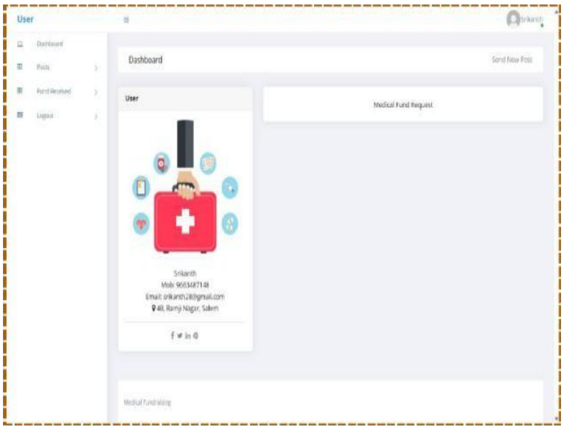
The results are displayed in the figures below.

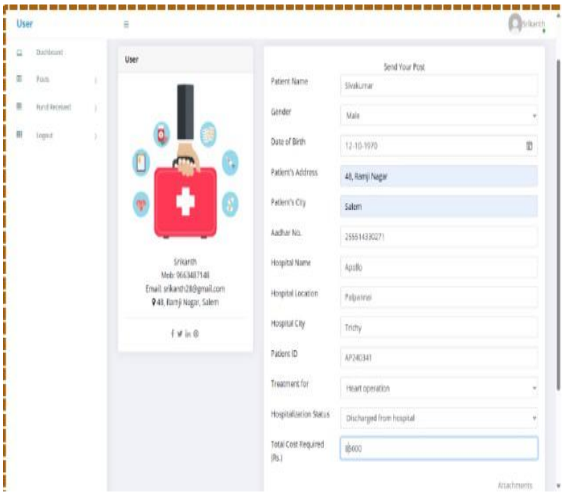


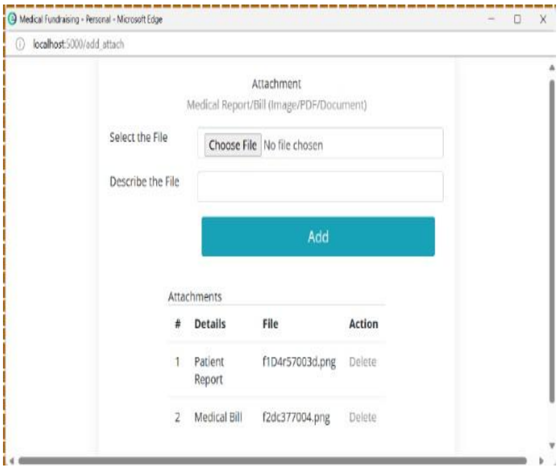




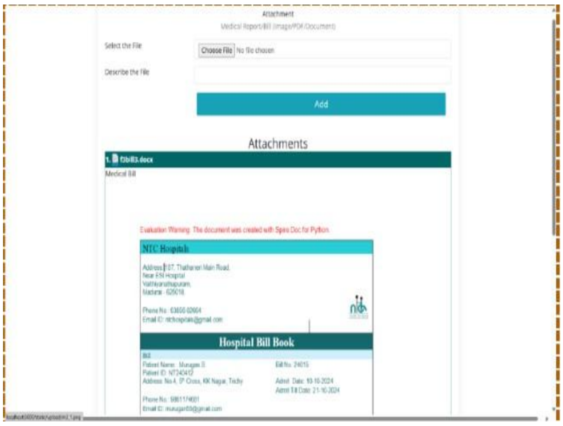


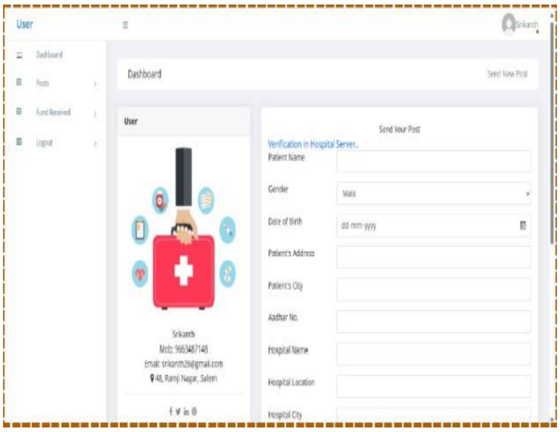


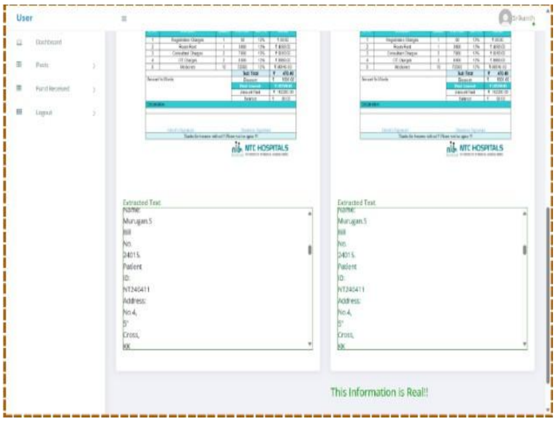


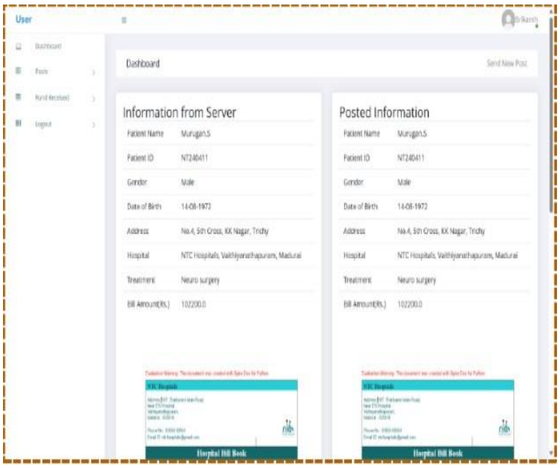


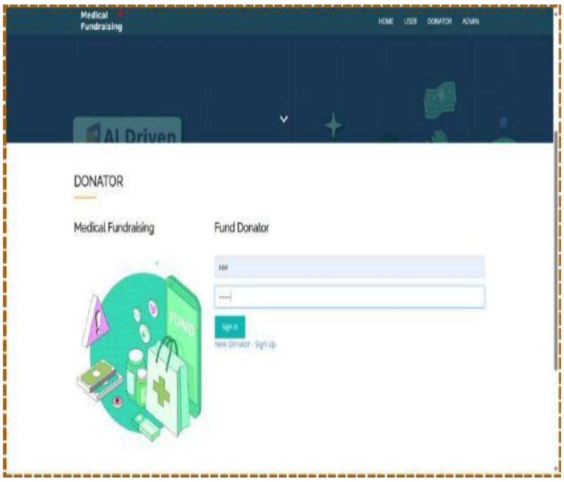
International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 06 | June – 2025 DOI: 10.55041/ISJEM04277
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

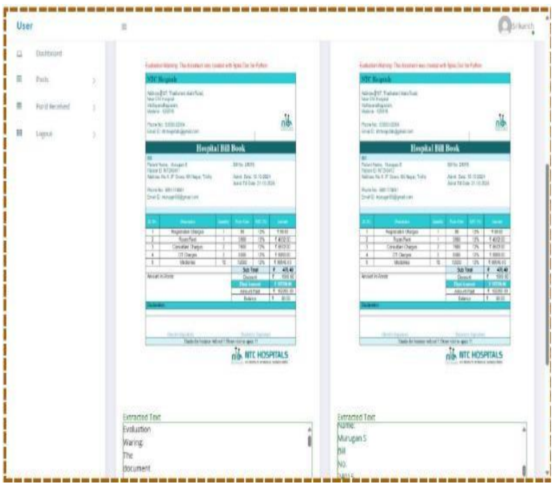


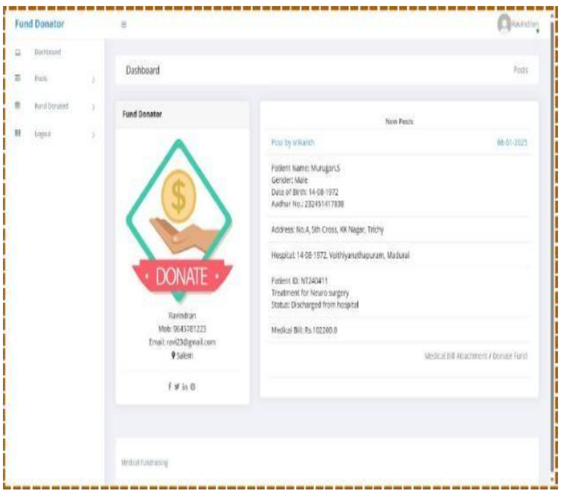




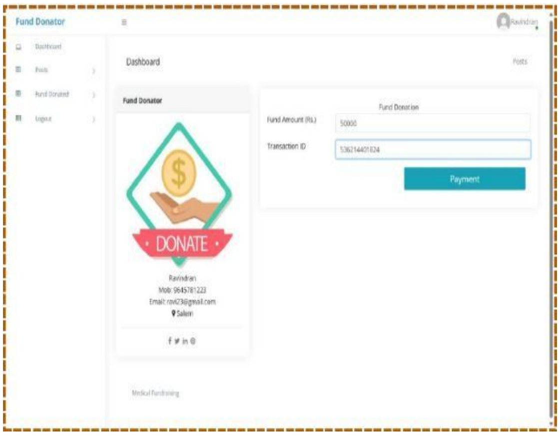


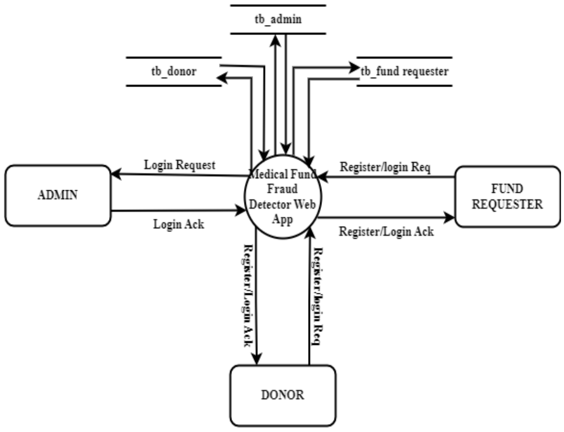






International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 06 | June – 2025 DOI: 10.55041/ISJEM04277
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

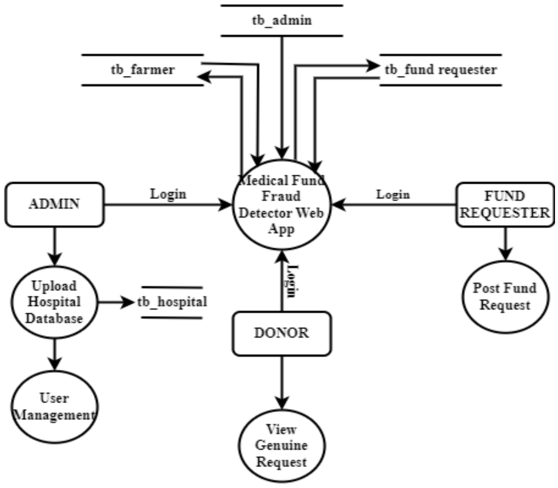




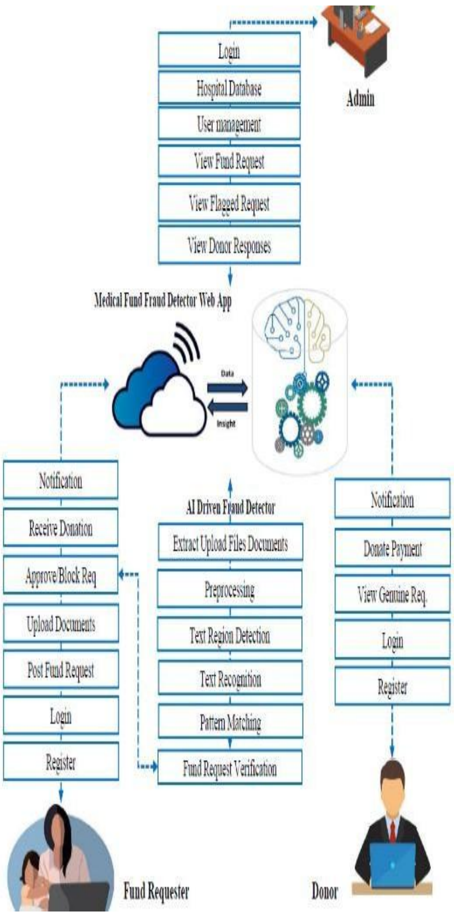
LEVEL 0 – DATA FLOW DIAGRAM

Dataset Selection

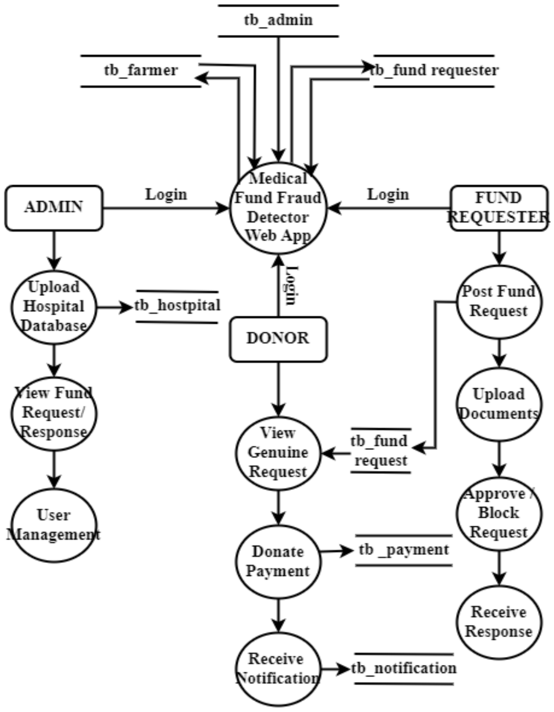
This project's dataset includes real medical billing information as well as hospital records. The purpose of integrating these data into the system is to cross-check information such hospital names, treatment charges, patient names, and stamps. It is necessary to compare this reliable dataset with the content taken from uploaded fund request documents.



LEVEL 1 – DATA FLOW DIAGRAM



SYSTEM ARCHITECTURE



LEVEL 2 – DATA FLOW DIAGRAM

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 06 | June – 2025 DOI: 10.55041/ISJEM04277
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

Data Pre-processing

Before applying machine learning models, all uploaded medical documents undergo preprocessing. This involves using filters like Gabor or Median to eliminate noise, scaling for uniformity, binarizing text for clarity, and turning photos to grayscale. The data is ready for precise text region identification and recognition thanks to these procedures.

Feature Selection

Important characteristics chosen for fraud detection include patient information, hospital names, treatment expenses, signatures, and hospital stamps. Through fuzzy comparison and pattern matching, these characteristics—which are taken from documents using OCR tools—are essential for confirming the legitimacy of medical fund claims.

Prediction Model

PaddleOCR is used for text recognition and YOLOv8 is used for text region identification in the prediction model. Together, these models are able to extract crucial information from the submitted papers. In order to identify discrepancies and categorize the fund request as authentic, flagged, or fraudulent, the retrieved data is further examined using a fuzzy matching algorithm that compares it with reliable hospital records.

Comparison Model

The comparison model is based on the fuzzy matching approach. It looks for differences like fake stamps, inaccurate billing information, or changed signatures by comparing the retrieved text with the reliable hospital dataset. This methodology guarantees that requests are only authorized if they have high similarity scores.

Medical News / Updates

The approach helps raise donor awareness by showing only verifiable medical money requests, even though it doesn't directly convey medical news. This inadvertently keeps donors updated on valid treatment instances that require financial support.

Result Analysis and Discussion

Medical fund requests are classified as either approved, flagged, or rejected by the system. Donors can see approved requests that pass authenticity checks. The rejected ones are verified as fake, and the flagged ones need to be manually reviewed because of irregularities. The accuracy of fraud identification is increased, donor hesitancy is decreased, and monies are guaranteed to reach legitimate patients thanks to this real-time automated verification.

REFERENCES

1. Fong, A., Jain, M., Sacks, W., Ho, A., & Chen, Y. (2020). *Crowdfunding campaigns and thyroid surgery: Who what where and how much?* Journal of Surgical Research, 253, 63–68.
2. Ba, Z., Zhao, Y., Song, S., & Zhu, Q. (2021). *Understanding the determinants of online medical crowdfunding project success in China.* Information Processing & Management, 58.
3. Mejia, J., Urrea, G., & Pedraza-Martinez, A. J. (2019). *Operational transparency on crowdfunding platforms: Effect on donations for emergency response.* Production and Operations Management, 28, 1773–1791.
4. Zhao, K., Zhou, L., & Zhao, X. (2022). *Multi-modal emotion expression and online charity crowdfunding success.* Decision Support Systems, 163.
5. Hou, J.-R., Zhang, J., & Zhang, K. (2023). *Pictures that are worth a thousand donations: How emotions in project images drive the success of online charity fundraising campaigns?* MIS Quarterly, 47, 535–584.
6. Iweriebor, L. E. (2023). *Approaches to Medicare provider fraud detection and prevention.* Doctoral dissertation, Capitol Technology University.
7. Bauder, R. A., & Khoshgoftaar, T. M. (2017, December). *Machine learning models for detecting Medicare fraud.* Presented at IEEE conference.
8. Johnson, J. M., & Khoshgoftaar, T. M. (2019). *Applying neural networks for Medicare fraud detection.* Journal of Big Data.
9. Bauder, R., & Khoshgoftaar, T. (2018, July). *Medicare data processing and integration techniques for fraud detection.* IEEE IRI Conference.
10. Herland, M., Khoshgoftaar, T. M., & Bauder, R. A. (2018). *Multiple Medicare data sources for large-scale fraud detection.* Journal of Big Data.
11. McGee, J., Sandridge, L., Treadway, C., Vance, K., & Coustasse, A. (2018). *Strategic approaches to fighting Medicare fraud.* The Health Care Manager.
12. Mohammed, M. A., Boujelben, M., & Abid, M. (2023). *A novel machine learning framework for fraud detection in blockchain-based healthcare networks.* Future Internet.
13. Castaneda, G., Morris, P., & Khoshgoftaar, T. M. (2019, April). *Using Maxout neural networks for detecting medical fraud in large datasets.* IEEE BigDataService.
14. Nabrawi, E., & Alanazi, A. (2023). *Applying machine learning to healthcare insurance fraud detection.* Risks.
15. Kumaraswamy, N., Markey, M. K., Ekin, T., Barner, J. C., & Rascati, K. (2022). *Retrospective and prospective view on healthcare fraud data mining methods.* Perspectives in Health Information Management.