AADHAAR IDENTITY VERIFICATION SYSTEM

GE19612 - PROFESSIONAL READINESS FOR INNOVATION, EMPLOYABILITY AND ENTREPRENEURSHIP PROJECT REPORT

Submitted by

SHANTHOSH S	**(2116220701263)**
SHARAN KUMAR D	**(2116220701264)**
JAYASURIYA KS	**(2116220701332)**

in partial fulfillment for the award of the degree

of

BACHELOR OF ENGINEERING

in

COMPUTER SCIENCE AND ENGINEERING

RAJALAKSHMI ENGINEERING COLLEGE

ANNA UNIVERSITY, CHENNAI

MAY 2025

RAJALAKSHMI ENGINEERING COLLEGE, CHENNAI

BONAFIDE CERTIFICATE

Certified that this Project titled **“ Aadhaar Card Identity Verification System using Machine Learning”** is the bonafide work of **“SHANTHOSH S (2116220701263), SHARAN KUMAR D(2116210701264), JAYASURIYA KS (2116220701332)”** who carried out the work under my supervision. Certified further that to the best of my knowledge the work reported herein does not form part of any other thesis or dissertation on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.

SIGNATURE
Dr. P. Kumar., M.E., Ph.D.,
HEAD OF THE DEPARTMENT
Professor
Department of Computer Science
and Engineering,
Rajalakshmi Engineering College,
Chennai - 602 105.

SIGNATURE
Dr. S. Senthil Pandi., M.E., Ph.D.,
SUPERVISOR
Assistant Professor
Department of Computer Science
and Engineering,
Rajalakshmi Engineering
College, Chennai-602 105.

Submitted to Mini Project Viva-Voce Examination held on ______________

Internal Examiner

External Examiner

ABSTRACT

"Aadhaar Card Verification System Using Machine learning" is a robust and intelligent solution designed to automate the verification of Aadhaar card authenticity in real time. Leveraging Optical Character Recognition (OCR) powered by Tesseract, the system extracts textual information such as the Aadhaar number and other relevant details from uploaded Aadhaar card images. A validation pipeline checks the extracted data against predefined structural and logical rules to confirm its legitimacy.

The platform is developed as a web application, with a React and Tailwind CSS frontend, a Node.js (Express) backend, and a PostgreSQL database managed via Prisma ORM. The OCR functionality is integrated as a Python microservice, enabling efficient and scalable data extraction. This modular architecture ensures maintainability, scalability, and smooth communication between components.

By automating Aadhaar verification, this system significantly reduces the risk of identity fraud and manual error, while improving processing time for onboarding and verification services. It presents a practical, secure, and efficient solution for government agencies, fintech platforms, and institutions requiring reliable identity validation.

ACKNOWLEDGMENT

Initially we thank the Almighty for being with us through every walk of our life and showering his blessings through the endeavor to put forth this report. Our sincere thanks to our Chairman **Mr. S. MEGANATHAN, B.E, F.I.E.**, our Vice Chairman **Mr. ABHAY SHANKAR MEGANATHAN, B.E., M.S.,** and our respected Chairperson **Dr. (Mrs.) THANGAM MEGANATHAN**, **Ph.D.,** for providing us with the requisite infrastructure and sincere endeavoring in educating us in their premier institution.

Our sincere thanks to **Dr. S.N. MURUGESAN, M.E., Ph.D.,** our beloved Principal for his kind support and facilities provided to complete our work in time. We express our sincere thanks to **Dr. P. KUMAR, M.E., Ph.D.**, Professor and Head of the Department of Computer Science and Engineering for his guidance and encouragement throughout the project work. We convey our sincere and deepest gratitude to our internal guides **Dr. SENTHIL PANDI., M.E., Ph.D.**, We are very glad to thank our Project Coordinator, **Ms. U. FARJANA M.Tech.,** Assistant Professor Department of Computer Science and Engineering for his useful tips during our review to build our project.

SHANTHOSH S	**2116220701263**
SHARAN KUMAR D	**2116220701264**
JAYASURIYA K S	**2116220701332**

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

S. No	ABBR	Expansion
1	AI	Artificial Intelligence
2`	API	Application Programming Interface
3	AJAX	Asynchronous JavaScript and XML
4	ASGI	Asynchronous Server Gateway Interface
5	AWT	Abstract Window Toolkit
6	BC	Block Chain
7	CSS	Cascading Style Sheet
8	DFD	Data Flow Diagram
9	DSS	Digital Signature Scheme
10	GB	Gradient Boosting
11	JSON	JavaScript Object Notation
12	ML	Machine Learning
13	RF	Random Forest
14	SQL	Structure Query Language
15	SVM	Support Vector Machine

CHAPTER 1
INTRODUCTION

1.1 GENERAL

“Aadhaar Card Verification System Using OCR and Tesseract” is an innovative solution aimed at automating the process of Aadhaar card authentication, a crucial step in identity verification across government and private sectors in India. This project addresses the growing need for fast, reliable, and secure verification of identity documents to prevent identity fraud, reduce manual errors, and streamline onboarding processes.

The system utilizes **Optical Character Recognition (OCR)** powered by the **Tesseract engine** to extract textual data from uploaded Aadhaar card images. Once the data is extracted, it is validated using a set of predefined rules, including Aadhaar number format verification, length constraints, and structure matching. This ensures that only genuine Aadhaar cards pass the verification process.

The platform is implemented as a **web-based application** using a modern tech stack: **React and Tailwind CSS** for the frontend interface, **Express.js** for backend API development, **PostgreSQL** as the database, and **Prisma ORM** for database interaction. The OCR component is implemented as a separate Python microservice to maintain modularity and scalability.

This architecture enables seamless communication between components, making the system highly extensible and efficient. The application provides a secure, user-friendly interface for uploading Aadhaar card images, while the backend handles data extraction, verification, and response generation. By integrating OCR and real-time verification logic, the system offers a practical solution for identity validation in domains such as e-KYC, fintech services, educational institutions, and government programs.

By enhancing verification reliability, reducing fraudulent usage of Aadhaar, and ensuring data privacy, this system represents a significant step forward in automated identity authentication.

1.2 OBJECTIVE

The objective of the "Aadhaar Card Verification System Using OCR and Tesseract" is to develop an intelligent, web-based application capable of verifying the authenticity of Aadhaar cards through automated image processing and text recognition. The system aims to streamline identity verification by integrating Optical Character Recognition (OCR) using Tesseract, which accurately extracts textual information from uploaded Aadhaar card images.

This project eliminates the need for manual data entry and validation by implementing backend logic that evaluates the structure and validity of the extracted Aadhaar number. The platform ensures data consistency and enhances security by leveraging a robust backend developed with Express.js, a PostgreSQL database managed via Prisma ORM, and a Python-based OCR service for processing.

Emphasizing reliability, scalability, and user privacy, the system is designed to provide a seamless and secure interface for both users and administrators. It is particularly applicable in areas such as e-KYC (Know Your Customer) processes, digital onboarding, and government welfare schemes. Ultimately, the project aims to reduce identity fraud, increase verification accuracy, and provide a reliable tool for real-time

Aadhaar validation.

1.3 EXISTING SYSTEM

Traditional Aadhaar card verification processes are primarily manual, requiring human intervention to cross-check details against official documents. These methods are time-consuming, error-prone, and often inefficient when handling large volumes of data during tasks such as customer onboarding, government service access, or institutional verification. Manual checks are susceptible to forgery detection failure, especially when faced with high-quality image manipulations or counterfeit Aadhaar cards.

Existing digital solutions, where available, typically rely on basic form validations or rudimentary pattern matching, which lack the sophistication to perform accurate optical character recognition or detect subtle inconsistencies in document formatting. Furthermore, many legacy systems operate in siloed, centralized environments, creating potential security risks and limiting the ability to scale efficiently or integrate modern AI techniques.

These shortcomings necessitate the development of a more robust, automated, and scalable solution that leverages AI-driven OCR technologies to ensure faster, more reliable Aadhaar card validation while minimizing manual involvement and protecting sensitive personal data.

CHAPTER 2

LITERATURE SURVEY

Old document images contains various information but the very difficult to identity but with the advanced technologies we can extract that. They are Optical Character Recognition(OCR) and Document Layout Analysis(DLA) these tools plays the important role in extraction of information. Older version of DLA focus on structural analysis whereas semantic oriented approaches gained a better capture of content. Semantic Document Layout Analysis enables the higher level feature extraction with the clean document retrieval and authentication tasks. In this project, recent research tells the methodologies such as pairwise annotation, comparative feature extraction, and document ranking, which collectively improve recognition performance. These advancements are particularly relevant for real-time verification systems, such as OCR-based Aadhaar card authentication, where precise extraction, comparison, and validation of textual and layout information are essential for reliability and accuracy.

Document image binarization is an important preprocessing technique which is used to separate the foreground data from the noisy degraded background mainly in poor quality documents. This process plays an major role inn ensuring the accuracy of the Optical Character Recognition(OCR) for identifying the documents like Aadhaar card image processing. In recent studies, various binarization methods have been tested and compared using established performance metrics commonly employed has one of the most effective text-background separation in poor conditions. The main mathematical model and operational block diagram shows its robustness and sustainability for real world applications. Such binarization advancements is significantly relying on OCR, where the clean and precise text is extracted from the identity for data verification.

In this modern era, the official documents such as government issued ID's like Aadhaar card, PAN card etc. certificates and other important records are paper based or image based. Many organization face significant challenges in manually verifying, extraction and searching for information from these IDs. Automated Identity Recognition and Classification (AIDRAC) system have emerged as a solution to reduce the work load. These models automatically classify the documents like Aadhaar card into predefined categories and extract the important textual details using OCR and machine learning techniques. This also enables the feature like auto filling the forms, minimizing the user effort. Extracted information are securely stored and accessed by authorized users, ensuring the data protection. Future research focus on developing a more robust, scalable solution that can be implemented across various organisation sectors like education, backing etc.

Identity verification process increasingly rely on automated image analysis and recognition technologies. While steganography, the practice of embedding the data with in the media files has been explored for securing and validating identity documents, this project didn't use the OCR method instead of that steganography. This method improves the accessibility, minimize the data corruption and provide good solution for various sector like education, backing etc. The use OCR simplify the integration, avoids complexity and gives a straight approaches for extracting, validating and securely stores the information.

Increase in the use of digital documents in administration and economic activities, the document forgery has simultaneously increased which leads to financial losses for governments and organizations. In response, many researchers have explored various techniques for automated forgery document detection system using machine learning and image processing approaches. Whereas some studies have convolutional neural networks(CNN) with accuracy rate of 73.95% and recall rate of 97.3%. This project focuses on a more accessible and efficient solution using Optical Character recognition(OCR) methods.

Detection and recognition of multilingual text in document images have evolved as significant challenges, their importance for applications such as tourist assistance and real-time translation. To advance research in this area, large-scale multilingual datasets, like the ICDAR dataset covering six scripts including Arabic, have been introduced. A novel Urdu-Text dataset has been developed, targeting Urdu, a cursive language derived from Arabic. Containing 1,400 scene images and 8,200 segmented words, this dataset provides all information for word-level detection and script identification. These resources improve the quality of OCR-based multilingual text recognition and helping the development of robust real-world applications.

Identity verification technique is the most important in this modern digital transactions, banking, finance, insurance, and education sectors. Traditional manual verification processes are very slow and inefficient. Optical Character Recognition (OCR) works as a key technique for extracting and verifying information in the identity documents like Aadhaar card etc.. and it will reduce authentication time and improves the accuracy. Previous research has explored AI-based solutions, including text detection and recognition methods, for extracting the data from the scanned image identity cards. Whereas some studies propose neural encoder decoder architectures for text recognition, recent comparisons indicates that the OCR engines like Tesseract is most efficient when optimized with preprocessing techniques, remains effective for structured document recognition.

One critical procedure in OCR is to detect text characters from a document image. However, some documents might come with embedded background images which often mislead the algorithms of character detection. For example, small dots or sharp edges from the background image are look like characters and passed to the next stage of the OCR pipeline, which may cause an error in extraction. Motivated by this observation, we present a novel and cost-effective image preprocessing method to accomplish the task. We first enhance the document images before OCR by utilizing the brightness and contrast parameters. Then we convert color images to gray and threshold it. This way, background images can be removed effectively without losing the quality of text characters. The method was tested using Tesseract (an open source OCR engine) and compared with two commercial OCR software. This experimental results show that the recognition accuracies are improved significantly after removing background images.

The analysis of identity document images captured by mobile phone cameras has significant challenges due to the poor quality of image. These images include multiple objects and various backgrounds, complicating the task of automatic text recognition and reducing their usability. Traditional image processing techniques struggle to accurately extract the text from the noisy images. In response to this issue, a novel method was developed for extracting the text within complex scene images to facilitate subsequent OCR processing. The approach employs a text kernel operator to roughly identify text regions, which are transformed as an active contour method. This initialization strategy not only accelerates the active contour convergence but also enhances the distinction between the inner and outer parts of the text contours. Experimental results gives a significant improvement in text localization and preprocessing accuracy.

Identity card recognition is a important and difficult application in the domain of image recognition. However, recognizing ID cards, particularly Chinese ID cards, presents several challenges, including interference from shadow grid lines, noise caused by hardware conditions and lighting, and the complexity of extracting meaningful information from bad backgrounds. To note this issue, researchers are proposed a solution that helps for face detection and national emblem detection to accurately segregate the foreground area of the ID card. And, a rotation correction technique based on the Hough transform is also used to rectify tilts upto 45 degrees in the image. For text identification, morphological image processing techniques is used and the characters are recognized using deep convolutional neural networks (CNNs). This proposed solution significantly improves recognizing accuracy and it achieves a success rate of 99.7% for Chinese and numerical characters, even in the critical conditions involving rotation and complex backgrounds the extraction of data is very smooth.

CHAPTER 3

PROPOSED SYSTEM

3.1 GENERAL

The Real-Time Aadhaar Card Image Verification System is a cutting-edge solution aimed at preventing fraudulent activities related to identity verification. By leveraging machine learning techniques like Optical Character Recognition (OCR) with Tesseract, and advanced classification models such as Random Forest and Support Vector Machine (SVM), the system ensures accurate identification of legitimate Aadhaar card images. The system intelligently processes and analyzes image attributes, including card number, name, photo, and barcode, to validate authenticity.

By integrating Python-based AI algorithms with a Flask web application, the system offers a seamless, efficient, and secure process for Aadhaar verification. This solution significantly reduces the risk of identity theft, fraud, and unauthorized access, creating a trusted and transparent verification environment. Through high accuracy and real-time functionality, it ensures a safer digital experience for users and institutions alike.

3.2 SYSTEM ARCHITECTURE DIAGRAM

The system architecture for the Real-Time Aadhaar Card Image Verification solution integrates machine learning techniques to provide a secure and reliable method for verifying Aadhaar card images. The system includes key roles such as data providers, analysts, and administrators, ensuring seamless data flow and accurate verification. The process begins with data collection and labeling of Aadhaar card images, followed by preprocessing steps like image cleaning, handling missing data, and removing distortions. Feature extraction is performed by analyzing the card's attributes such as the number, name, photograph, and QR code, followed by dimensionality reduction to enhance performance. The system employs machine learning models like Support Vector Machines (SVM), Random Forest, and Gradient Boosting, with a special focus on Gradient Boosting due to its precision in anomaly detection. The performance of the models is assessed using accuracy metrics and confusion matrices. The final, optimized model is deployed through a Flask-based web application, which facilitates the interaction with the frontend and backend components for real-time Aadhaar card verification. A centralized database stores all relevant data, including training results, model predictions, and evaluation metrics, ensuring transparency and accountability. The secure server infrastructure ensures the safe processing of Aadhaar card images, with a focus on providing real-time verification results for users.

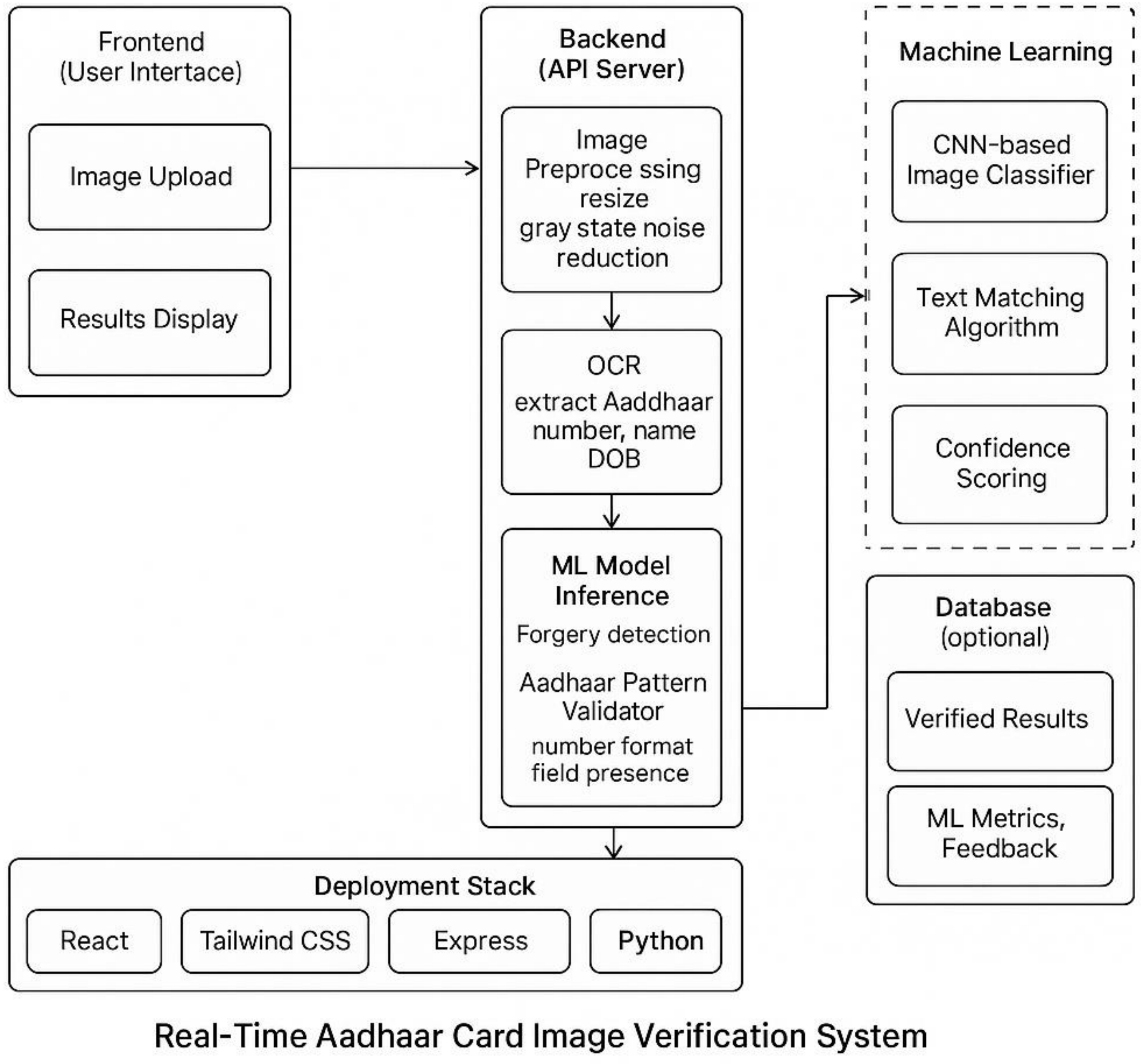

Fig 3.1: System Architecture

3.3 DEVELOPMENTAL ENVIRONMENT

3.3.1 HARDWARE REQUIREMENTS

The hardware specifications provide the foundation for the successful deployment and operation of the Real-Time Aadhaar Card Image Verification System. The system requires moderate computing resources for tasks such as image preprocessing, OCR (Optical Character Recognition) using Tesseract, and machine learning-based validation. These requirements ensure smooth performance for both the backend (Python + Flask + ML processing) and frontend (React + Tailwind CSS) components

Table 3.1 Hardware Requirements

COMPONENTS	SPECIFICATION
GRAPHICS	Integrated GPU(optional)
DISPLAY	Minimum 1080p resolution
OPERAING SYSTEM	Windows 10 or above
INTERNET	Stable broadband connection
PROCESSOR	Intel Core i5 or above
RAM	8 GB DDR4 RAM or higher
POWER SUPPLY	+5V power supply

3.3.2 SOFTWARE REQUIREMENTS

The software specifications define the essential system components and technologies required to develop and operate the Aadhaar Card Verification System. This specification outlines what the system should do, including frontend rendering, backend processing, OCR integration, and machine learning-based validation. It serves as a foundation for estimating project cost, scheduling development phases, assigning tasks, and tracking team progress throughout the system's lifecycle.

Table 3.2 Software Requirements

COMPONENTS	**SPECIFICATION**
Operating System	Windows 10 or higher
Frontend	ReactJS,TailwindCSS
Backend	ExpressJS
Database	PostgresSQL
Machine Learning	Python (OCR Tesseract)
API testing tool	Postman

3.4 DESIGN OF THE ENTIRE SYSTEM

3.4.1 ACTIVITY DIAGRAM

The activity diagram Fig 3.2 represents the workflow for detecting fake Aadhaar profiles using a React-based frontend and an Express.js backend integrated with a Python-based machine learning system. The process begins with the user uploading an Aadhaar card image through the web interface. This image is sent to the Express.js server, which forwards it to a Python service that applies Tesseract OCR to extract text from the image. The extracted data undergoes preprocessing steps such as cleaning, normalization, and validation. It is then passed to a trained machine learning model—such as Random Forest or Support Vector Machine—which classifies the Aadhaar as either "valid" or "fake" based on learned features. The prediction result is sent back through the Express.js backend to the user, ensuring a seamless and accurate verification process.

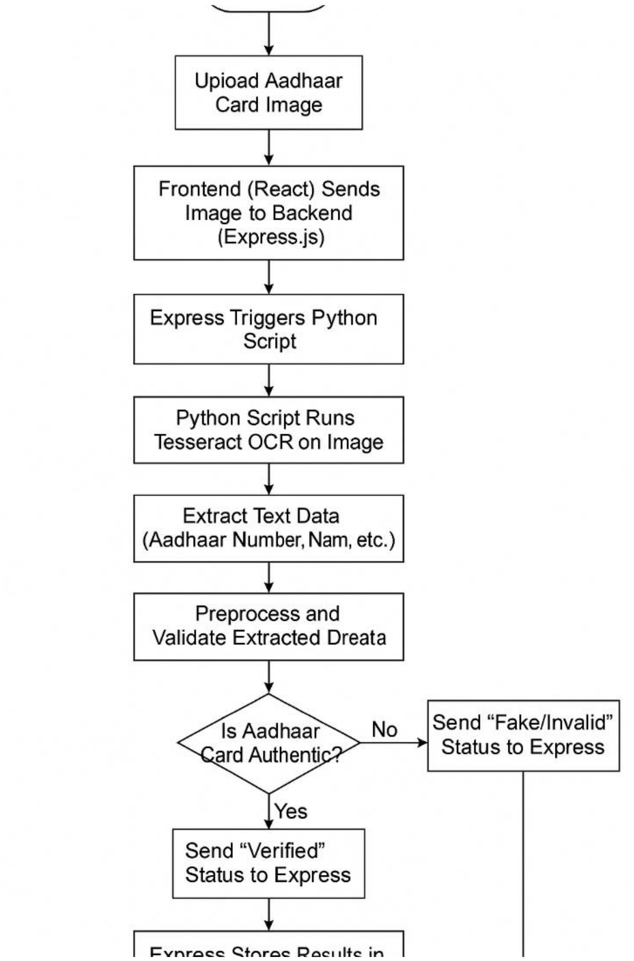
Upioad Aadhaar Card Image
Frontend (React) Sends Image to Backend (Express.js)
Express Triggers Python Script
Python Script Runs Tesseract OCR on Image
Extract Text Data (Aadhaar Number, Nam, etc.)
Preprocess and Validate Extracted Dreata
Is Aadhaar Card Authentic?
No
Send “Fake/Invalid” Status to Express
Yes
Send “Verified” Status to Express
Express Stores Results in

Fig 3.2: Activity Diagram

3.4.2 DATA FLOW DIAGRAM

The data flow diagram Fig 3.3 outlines the process of verifying Aadhaar card authenticity using a machine learning-based system built with React, Express.js, and Python. The workflow begins with users uploading Aadhaar card images through the React frontend. These images are routed to the Express.js backend, which sends them to a Python service where Tesseract OCR extracts textual information. The extracted data is then preprocessed—cleaned, normalized, and filtered—to ensure it meets the model’s input criteria. The preprocessed data is divided into 80% training data and 20% testing data to develop and evaluate machine learning models like Support Vector Machines, Random Forest, or Gradient Boosting. The trained model, deployed in the backend, classifies the Aadhaar as either real or fake. Results are returned to the user through the frontend, providing a secure and accurate system for real-time Aadhaar verification.

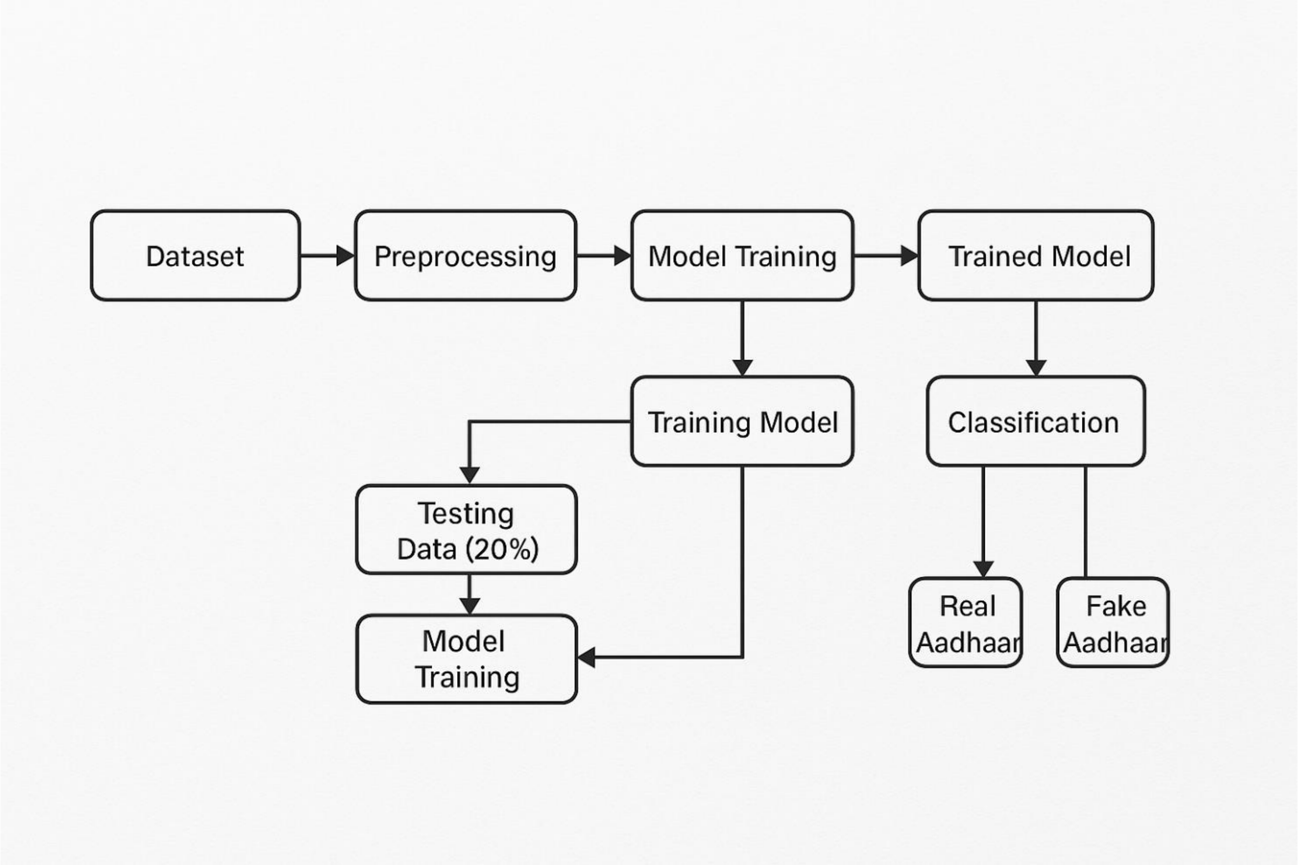

Fig 3.3:Data Flow Diagram

3.5 STATISTICAL ANALYSIS

The feature comparison table highlights the key differences between the Aadhaar Card Verification System using OCR (Tesseract) and traditional manual identity verification methods. The proposed system integrates machine learning-powered text recognition, real-time data extraction, and automated verification logic, ensuring a faster, scalable, and more reliable verification process. While some steps overlap with existing ID verification practices, the integration of Tesseract OCR and a robust web architecture (React frontend, Express backend, and PostgreSQL database) significantly improves accuracy, reduces manual errors, and streamlines the entire workflow of identity validation.

The Aadhaar Card Verification System stands out through its modern and automated approach to identity verification, distinguishing it from traditional manual verification methods. At its core, the system utilizes machine learning-powered Optical Character Recognition (OCR) through Tesseract to accurately extract data from Aadhaar card images. This automation significantly improves verification accuracy, reduces processing time, and minimizes human errors. Built on a robust full-stack architecture—featuring a React-based frontend, Express backend, Prisma ORM, and PostgreSQL database—the system ensures seamless performance, scalability, and ease of use. Real-time verification logic is integrated to validate extracted data against predefined formats and database records, enhancing reliability. The platform also features a clean and intuitive web interface, making it accessible to users with minimal technical knowledge. By automating data extraction and verification, reducing manual workload, and improving accuracy, the system provides a comprehensive solution for secure and efficient identity validation. Figure 3.4 presents a comparative analysis of existing manual verification methods and the proposed system, highlighting its enhanced speed, reliability, and user experience.

CHAPTER 4
MODULE DESCRIPTION

The workflow for the proposed Aadhaar Card Verification System is designed to ensure a structured, efficient, and automated process for verifying user identity using Aadhaar credentials. It consists of the following sequential steps:

4.1 SYSTEM ARCHITECTURE

4.1.1 USER INTERFACE DESIGN

The sequence diagram in Fig 4.1 illustrates the step-by-step process of verifying an Aadhaar card. The process begins with the user uploading an image of their Aadhaar card through the frontend interface. The image is then sent to the server where it undergoes preprocessing to enhance clarity and optimize it for text extraction. Tesseract OCR is applied to extract key details such as Name, Aadhaar Number, Date of Birth, and Gender. These extracted details are passed to the backend system, where validation checks are performed to ensure the data conforms to standard Aadhaar formats. The Express server then interacts with the PostgreSQL database via Prisma to log the verification attempt and optionally compare the data with existing records for identity confirmation. Once verified, the result—Valid, Invalid, or Error—is returned to the user and displayed on the frontend. This sequence ensures a secure, accurate, and streamlined process for Aadhaar identity verification.

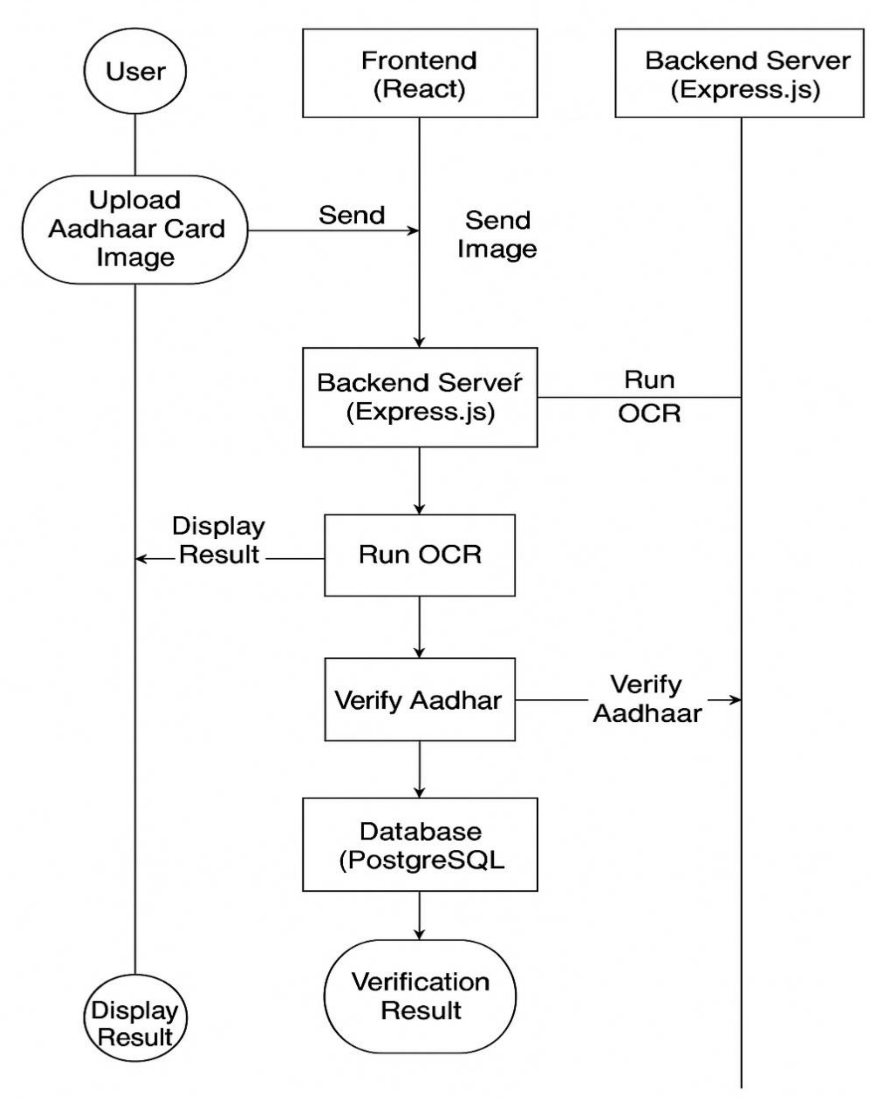

Fig 4.1: SEQUENCE DIAGRAM

4.1.2 BACK END INFRASTRUCTURE

The backend infrastructure of the Aadhaar Card Verification System is designed to ensure efficient processing, secure data handling, and seamless integration between the frontend and backend services. The backend is built using Express.js, which handles API requests and application logic. Prisma ORM is used for database interactions, managing structured user data and storing verification logs in a PostgreSQL database. The backend also handles the image preprocessing pipeline and routes the uploaded Aadhaar card images to the Tesseract OCR engine, which extracts key identity information from the image. Extracted data is then validated and logged accordingly. The system ensures data integrity and reliability through backend validation layers, and API responses are structured to deliver clear verification results to the frontend. This infrastructure enables scalable, real-time Aadhaar verification with a focus on speed, security, and accuracy.

4.2 DATA COLLECTION AND PREPROCESSING

4.2.1 Dataset and Data Labelling

Labeled datasets are collected, including images of Aadhaar cards and corresponding ground truth data such as names, Aadhaar numbers, dates of birth, and other identity attributes. Accurate labeling ensures that each image is correctly paired with its verified textual information, allowing the OCR system to be trained and tested effectively. This labeled data is essential for evaluating the performance of the text extraction and validation pipeline, helping to distinguish between correctly extracted information and OCR-related errors or fraudulent manipulations.

4.2.2. Data Preprocessing

The raw dataset undergoes extensive preprocessing, which includes:

Data Cleaning: Elimination of inconsistent or redundant data. Missing Value Replacement: Imputation techniques to handle incomplete entries.

Outlier Detection : Managing extreme or abnormal values for consistency.

4.2.3 Feature Selection

Attribute Evaluation: The most influential attributes for identity verification, such as the Aadhaar number, name, date of birth, and gender, are identified to ensure accurate extraction and validation. These features are prioritized to enhance the system's reliability in detecting tampered or incorrect data.

Data Normalization and Simplification: Data complexity is minimized by preprocessing steps that standardize the extracted text, ensuring uniformity across different Aadhaar card images, while retaining the critical features necessary for accurate verification.

4.2.4 Classification and Model Selection

Tesseract OCR: Used for extracting text from Aadhaar card images, ensuring accurate recognition of critical attributes such as the Aadhaar number, name, and date of birth.

Validation Algorithms: Various algorithms are employed for data validation, including pattern recognition techniques to verify the correctness of extracted information, such as matching Aadhaar numbers to standard formats.

Anomaly Detection (Machine Learning models): Models like **Random Forest** and **Gradient Boosting** are used to identify inconsistencies or potential fraud in the extracted data, ensuring that the verified information is accurate and not tampered with.

4.2.5 Performance Evaluation and Optimization

Accuracy is calculated to determine the overall success rate of the system in correctly identifying valid and invalid Aadhaar data.

Confusion Matrices are used to visualize the performance of the model by showing true positives, false positives, true negatives, and false negatives, helping to identify areas for improvement.

4.2.6 Model Deployment

The optimized model is deployed via an **Express.js-based system**, ensuring seamless integration with the web application. Real-time identity verification is conducted by processing live image uploads from users. The system extracts Aadhaar card data, validates it against predefined formats, and provides immediate results, ensuring quick and efficient identity verification. Continuous monitoring of model performance is also implemented, allowing the system to adapt and maintain high levels of accuracy as new data is processed.

4.2.7 Centralized Server and Database

All data, including image uploads, extracted text, validation results, and verification logs, is securely stored in a centralized **PostgreSQL database**. The **Express.js server** manages communication between the OCR model, the backend system, and the database, ensuring secure processing and storage of user data. The system implements data integrity measures, ensuring that all verification attempts and results are logged and can be audited, maintaining transparency and trust. Additionally, encryption is used for sensitive data to ensure privacy and compliance with security standards.

4.3 SYSTEM WORK FLOW

4.3.1 User Interaction:

Users initiate the verification process by submitting their **Aadhaar card images** through the web interface. The system processes these images and extracts key attributes, such as **Aadhaar number**, **name**, **date of birth**, and **gender**. The extracted data is then evaluated for accuracy, ensuring that all fields match the expected format and are free from any discrepancies or tampering. Real-time validation checks are performed against predefined rules to confirm the legitimacy of the Aadhaar card information, ensuring a secure and efficient verification process.

4.3.2 Fake Profile Detection:

Simple machine learning techniques, specifically **Tesseract OCR**, are applied to extract key details from the Aadhaar card image, such as the **Aadhaar number**, **name**, **date of birth**, and **gender**. The system processes the extracted text and performs validation checks to ensure the authenticity of the information. By analyzing these extracted attributes, the system verifies that the submitted Aadhaar card data is genuine, accurately matching the predefined formats and expected patterns.

4.3.3 Fraud Prevention & Reporting:

If an Aadhaar card is flagged as potentially fraudulent, users receive a detailed report explaining the identified discrepancies, such as **invalid Aadhaar number formats**, **mismatched names**, or **missing data**. The system allows users to submit additional verification information or appeal the decision, ensuring a **fair and transparent process**. Additionally, the system can automatically notify administrators of the verification service, prompting them to take further action if necessary.

4.3.4 Continuous Learning & Improvement:

The system continuously updates its OCR model based on new patterns and edge cases encountered during the Aadhaar card verification process. Additionally, user feedback and system logs contribute to refining text extraction accuracy and validation checks, ensuring that new types of fraudulent activity or data discrepancies are effectively addressed.

This structured workflow ensures a secure, transparent, and efficient process for verifying Aadhaar cards, fostering a safer and more reliable digital identity verification system.

CHAPTER 5
IMPLEMENTATION

5.1 IMPLEMENTATION

The project is developed and deployed using a robust technology stack, consisting of **JavaScript (Node.js)** for backend processing with the **Express.js** framework and **PostgreSQL** for database management through **Prisma ORM**. The frontend is built using **React** and **Tailwind CSS**, ensuring a responsive and user-friendly interface. For Aadhaar card verification, the system leverages **Tesseract OCR** to extract textual information from images of Aadhaar cards, which is then validated using predefined checks to ensure authenticity.

The implementation includes an intuitive web interface, allowing users to upload Aadhaar card images for automated verification. To ensure security and data integrity, the platform stores verification logs and user data in an encrypted database, preventing unauthorized access and tampering. This enhances trust in the system's accuracy and reliability.

The backend server efficiently processes user data, extracts relevant details from the uploaded images, and performs validation checks. The system can be updated periodically with new fraudulent detection patterns and model improvements, improving its accuracy over time.

CHAPTER 6

CONCLUSION AND FUTURE ENHANCEMENT

6.1 CONCLUSION

The proposed Aadhaar Card Verification System integrates machine learning and cutting-edge OCR technology to create a robust solution for validating Aadhaar card information, effectively addressing identity fraud with enhanced accuracy, reliability, and security. By utilizing Tesseract OCR for text extraction and implementing real-time validation checks, the system ensures precise verification of Aadhaar card details, adapting to evolving patterns of fraudulent activity.

The system's transparent and secure architecture, powered by Express.js, Prisma ORM, and PostgreSQL, ensures the integrity of the verification process, while the user-friendly web interface allows seamless interaction for both users and administrators. The platform enables real-time Aadhaar verification, providing a trusted tool to eliminate fraudulent identity submissions.

With continuous development, this project holds the potential to revolutionize identity verification in digital systems, offering a comprehensive response to rising identity fraud. By combining AI and secure data management practices, the system strengthens digital identity security, fosters user trust, and enhances the integrity of online transactions, ensuring a safer and more reliable online ecosystem.

6.2 FUTURE ENHANCEMENT

Future enhancements for this system could include integrating **deep learning models**, such as **Convolutional Neural Networks (CNNs)**, for more advanced image-based analysis of Aadhaar card images. This could improve the accuracy of text extraction, especially for low-quality or distorted images. Additionally, **transformers** like **BERT** could be used for enhanced text analysis, helping to detect complex patterns of fraudulent activity in the extracted data.

Implementing **smart contracts** and **decentralized identity verification (DID)** on the blockchain could further strengthen security, ensuring tamper-proof records and more transparent validation processes. This would allow users to control their own identity data while providing greater trust in the verification system.

Real-time detection with **adaptive learning** using **reinforcement learning** could improve the system's ability to evolve alongside new types of fraudulent activities, increasing detection accuracy over time.

Furthermore, expanding the detection capabilities across multiple platforms using **federated learning** could improve privacy by allowing models to be trained on decentralized data without compromising sensitive user information. This approach would maintain the effectiveness of fraud detection while preserving user privacy.

Lastly, incorporating **privacy-preserving techniques** such as **differential privacy** could ensure that individual data is not exposed during the verification process, further enhancing the system's security and user trust.

REFERENCES

[1] Kumar, P., et al. "Towards Sustainable Architecture: ML for Predicting Energy Use in Buildings." In 2024 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI), pp. 1-6. IEEE, 2024.

[2] Kumar, P., et al. "Human Activity Recognitions in Handheld Devices Using Random Forest Algorithm." In 2024 International Conference on Automation and Computation (AUTOCOM), pp. 159-163. IEEE, 2024.

[3] Kumar.P., et al. "Improvement of Classification Accuracy in ML Algorithm by Hyper-Parameter Optimization." In 2023 International Conference on Research Methodologies in Knowledge Management, Artificial Intelligence and Telecommunication Engineering (RMKMATE), pp. 1-5. IEEE, 2023.

[4] Ghani, et al. "Securing synthetic faces: A GAN-blockchain approach to privacy-enhanced facial recognition." Journal of King Saud University-Computer and Information Sciences 36, no. 4 (2024): 102036

[5] Baldimtsi, Foteini, Konstantinos Kryptos Chalkias, Yan Ji, Jonas Lindstrøm, Deepak Maram, Ben Riva, Arnab Roy, Mahdi Sedaghat, and Joy Wang. "zklogin: Privacy-preserving blockchain authentication with existing credentials." In Proceedings of the 2024 on ACM SIGSAC Conference on Computer and Communications Security, pp. 3182-3196. 2024.

[6] Yaga, Dylan, Peter Mell, Nik Roby, and Karen Scarfone. "Blockchain technology overview." arXiv preprint arXiv:1906.11078 (2019).

[7] Paul, Shovon, Jubair Islam Joy, Shaila Sarker, Abdullah-Al-Haris Shakib, Sharif Ahmed, and Amit Kumar Das. "Fake news detection in social media using blockchain." In 2019 7th international conference on smart computing & communications (ICSCC), pp. 1-5. IEEE, 2019.

[8] Farhan, Maruf, Rejwan Bin Sulaiman, and Abdullah Hafez Nur. "Blockchain: A secure solution for identifying counterfeits and improving supply chain reliability." In 2024 International Conference on Advances in Computing, Communication, Electrical, and Smart Systems (iCACCESS), pp. 1-6. IEEE, 2024.

[9] Althero, Zacky, Jazlan Syahreza, and Alvano Ortiz. "Blockchain Technology for Authentication and Validation Social Network Accounts." Blockchain Frontier Technology 3, no. 1 (2023): 32-38.

[10] Yaga, Dylan, Peter Mell, Nik Roby, and Karen Scarfone. "Blockchain technology overview." arXiv preprint arXiv:1906.11078 (2019).

[11] Baldimtsi, Foteini, Konstantinos Kryptos Chalkias, Yan Ji, Jonas Lindstrøm, Deepak Maram, Ben Riva, Arnab Roy, Mahdi Sedaghat, and Joy Wang. "zklogin: Privacy-preserving blockchain authentication with existing credentials." In Proceedings of the 2024 on ACM SIGSAC Conference on Computer and Communications Security, pp. 3182-3196. 2024.

[12] Bharti, Nasib Singh Gill, and Preeti Gulia. "Exploring machine learning techniques for fake profile detection in online social networks." International Journal of Electrical and Computer Engineering (IJECE) 13, no. 3 (2023): 2962-2971.

[13] Chakraborty, Partha, Mahim Musharof Shazan, Mahamudul Nahid, Md Kaysar Ahmed, and Prince Chandra Talukder. "Fake profile detection using machine learning techniques." Journal of Computer and Communications 10, no. 10 (2022): 74-87.

[14] Jestin Johny., et al. "Potential of blockchain technology in supply chain management: a literature review." International Journal of Physical Distribution & Logistics Management 49, no. 9 (2019): 881-900.

[15] Shahbazi, Zeinab, and Yung-Cheol Byun. "Fake media detection based on natural language processing and blockchain approaches." IEEE Access 9 (2021): 128442-128453.

[16] Rani, Poonam, Vibha Jain, Jyoti Shokeen, and Arnav Balyan. "Blockchain-based rumor detection approach for COVID-19." Journal of Ambient Intelligence and Humanized Computing 15, no. 1 (2024): 435-449.

[17] Farooqui, Faisal, and Muhammed Usman Khan. "A literature review on automatic detection of fake profile." International Journal Of Engineering And Management Research 13, no. 2 (2023): 196-200.

[18] Sarmah, Simanta Shekhar. "Understanding blockchain technology." Computer Science and Engineering 8, no. 2 (2018): 23-29.

[19] Alam, Md Jahangir, Ismail Hossain, Sai Puppala, and Sajedul Talukder. "Combating identity attacks in online social networks: A multi-layered framework using zero-knowledge proof and permissioned blockchain." In Proceedings of the International Conference on Advances in Social Networks Analysis and Mining, pp. 636-643. 2023.

[20] Shahbazi., et al. "Fake media detection based on natural language processing and blockchain approaches." IEEE Access 9 (2021): 128442-128453.