A PROJECT REPORT

On

AI-Driven Document Intelligence: A Comprehensive Approach to Classification, Q&A Handling, and Fraud Detection

Submitted in partial fulfillment of the requirements

for the award of the degree of

BACHELOR OF TECHNOLOGY

in

COMPUTER SCIENCE AND ENGINEERING

ARTIFICIAL INTELLIGENCE

Submitted by

Shambhavi Rai	**00901172020**
Sonanshi Goel	**01601172020**
Mehak Aggarwal	**03401172020**
Princy Singhal	**05101172020**
Kanika Kanojia	**06301172020**

Under the guidance of

Ms. Ritika Kumari
Assistant Professor
Department of AI & DS

DEPARTMENT OF ARTIFICIAL INTELLIGENCE AND DATA SCIENCES
INDIRA GANDHI DELHI TECHNICAL UNIVERSITY FOR WOMEN
KASHMERE GATE, DELHI - 110006

MAY 2024

Indira Gandhi Delhi Technical University for Women

STUDENT(s) UNDERTAKING

This is to undertake the work titled “AI-Driven Document Intelligence: A Comprehensive Approach to Classification, Q&A Handling, and Fraud Detection” in the B.Tech. Major Project report during academic year 2023-24 under the guidance of Ms. Ritika Kumari is our original work

The report has been written by us in our own words and not copied from elsewhere. This report was submitted to plagiarism detection software on 28 May 2024 and percentage similarity found was, similarity report attached as Appendix.

Anything that appears in this report which is not our original has been duly and appropriately referred / cited / acknowledged. Any academic misconduct and dishonesty found now or in future in regard to above or any other matter pertaining to this report shall be solely and entirely my responsibility. In such a situation, We understand that a strict disciplinary action can be undertaken against us by the concerned authorities of the University now or in future and we shall abide by it.

Student Signature
Student Name: Sonanshi Goel

Student Signature
Student Name: Princy Singhal

Student Signature
Student Name: Shambhavi Rai

Student Signature
Student Name: Kanika Kanojia

Student Signature
Student Name: Mehak Aggarwal

Date of submission: 28 May 2024
Place: IGDTUW, Delhi

CERTIFICATE

This is to certify that the work titled "AI-Driven Document Intelligence: A Comprehensive Approach to Classification, Q&A Handling, and Fraud Detection" submitted in this B.Tech. Major Project Report during the academic year 2023-24 was done under my guidance and supervision.
This work is original work to the best of our knowledge and has not been submitted anywhere else for the award of any credits / degree whatsoever. The work is satisfactory for the award of B.Tech. Minor Project credits.

Ms. Ritika Kumari
Assistant Professor
Department of Artificial Intelligence and Data Science,
IGDTUW, Kashmere Gate, Delhi-110006

ACKNOWLEDGEMENT

We, Kanika Kanojia, Mehak Aggarwal, Princy Singhal, Shambhavi Rai, Sonanshi Goel would like to acknowledge our mentor Ms. Ritika Kumari for her very helpful comments, support and encouragement.

Finally, We are grateful to Indira Gandhi Delhi Technical University For Women for providing a healthy, supportive and understanding environment. They allowed us the freedom to explore innovative models to simplify a complex business problem. This made our project work possible without any hindrance.

Shambhavi Rai
00901172020
Sonanshi Goel
01601172020
Mehak Aggarwal
03401172020
Princy Singhal
05101172020
Kanika Kanojia
06301172020

Abstract

In the modern era of significant information overload, our unique idea acts as a lighthouse of hope for overcoming the difficulties related to document interaction. We use state-of-the-art technologies to transform how people interact with documents in accordance with the demand to speed up workflows. Our methodology is based on the deliberate categorization and extraction of relevant content pieces, which improves usability and user productivity when going through the large amount of available data.

Users get an effortless experience as they quickly enter PDF documents, opening up a multitude of features intended to meet their various demands. The document interaction is made more sophisticated by a variety of sophisticated features which includes text-to-speech capabilities, automated summarization, document classification, and fraud detection.
This endeavor is a dynamic tool that has multiple uses. It helps students understand difficult research papers while enhancing operations in the financial sector by quickly classifying and arranging huge quantities of records.

Text-to-speech features give users an effective and different method to consume content; this is especially helpful for people who learn best by hearing things out loud or who have visual limitations. By compressing large papers into brief summaries, automated summarizing saves users a great deal of time and assures that important information is kept without being flooded by information. Being able to classify documents offers the system an additional layer of capability and assists in categorizing a variety of content in an easy-to-use manner.From the field of education, our project is essential in helping students understand challenging research papers. By providing resources that speed up comprehension, we improve the effectiveness of learning. Our system's characteristics are not limited to meeting the demands of students; they also have wider uses, demonstrating its ability to adapt to a range of issues in numerous fields.

In simple terms, our concept offers a transformative solution that extends beyond simple document interaction, thereby surpassing the constraints imposed by information overload. Our focus on incorporating modern technology and user-friendly features paves the way for a more streamlined, effective, and convenient method of handling documents in the digital era.
Our work is a driving force driving progress, resulting in enhanced efficiency, comprehending and user experience in both the banking industry and educational settings. The usefulness and attractiveness of technology as a whole are further strengthened by ancillary capabilities like grammar checking and paraphrasing, which enhance document quality and adaptability.

INDEX

INTRODUCTION

Finding duplicates of an individual image, referred to as a "template," in an image, also what location they are, and how they are oriented—is an enduring challenge in image processing and, in particular, pattern recognition. The process of looking for a small template image within the scene image is known as template matching. Generalized names for the resulting algorithms are template matching algorithms [20]. In image processing, robust matching and searching are tough obstacles [21] . Rotation [22], distortion [23], dimension changes [24], partially obstructed [25], and light exposure alterations [26] are some of the issues that arise. Current techniques for matching and searching make an effort to address these issues by utilizing either one characteristic or intricate characteristic groupings [27]. However, it is challenging to distinguish identical items and correlate the backdrop complexity substantially and precisely. Many template matching methods have been presented [28] for lowering the amount of computing power required because the algorithm's effectiveness relies heavily on its processing time.

At heart, AI-Driven Document Intelligence: A Comprehensive Approach to Classification, Q&A Handling, and Fraud Detection employs AI techniques that facilitate smart document categorization thus enabling quick identification and extraction of vital content segments. Classification and categorisation of identity documents like Aadhaar card, PAN card, etc facilitate more effective document management. We can leverage deep learning algorithms that demonstrate good transferability and high performance. Document structure, encompassing text blocks, illustrations, and tables, is characterized by its spatial layout. After that, these descriptions are used for categorization [8] or similarity calculations [17]. Text is used in the second kind of technique. These techniques provide a bag of words or Word2Vec, which is fed into classifiers, describing the text content (extracted with an OCR in the case of scanned documents) [18]. Recurrent neural networks, also known as RNN, are being utilized to document classification lately. [19].

Millions of people utilize large language models (LLMs) in real-world applications like chatbots, search engines, and coding assistants. LLMs have also greatly expedited the field of natural language processing (NLP) [2].

Additionally, document management would be easier to develop using chatbots and scalable AI/LLM apps with the LangChain framework. The LLM Model is a massive language model that may be used to write text, translate across languages, produce other kinds of creative content, and respond to your questions in a useful manner [1] .

This paper focuses on the components of the system and workflow of DIA outlining their practical applications across different areas. The DIA has a transformative potential right from educational perspectives where students can easily identify key points from research papers up to government institutions that make administrative operations efficient through fast handling of documents.

Introducing a document text paraphrasing feature to improve clarity, accessibility, and adaptability in various communication and learning contexts.

1.1 Motivation

Our technology for document interaction emerged from an intense need to address the challenges presented by the information-rich environment of the modern day. In order to successfully navigate a world that is characterized by an overwhelming stream of data from many sources, the necessity for innovative solutions for speeding up document interaction becomes more and more apparent.

Our study is primarily motivated by the realization of the profound impact that technology may have in educational environments. Our dedication is motivated by the widespread issue of overload of information that students face when juggling complex research papers and other academic resources. Our intention is to give them a tool that will allow them to quickly extract important insights while also streamlining the document interaction process.

1.2 Scope of Project

With a wide-ranging and all-inclusive scope, the document engagement system project addresses important facets of managing documents and user engagement. The project offers a broad range of capabilities aimed at significantly enhancing the document interaction experience, and it was created with a diverse user base covering different sectors and scenarios in mind.

Document kinds: A broad variety of document categories, such as educational papers, financial reports, legal documents, and general textual material, are included in the project's scope. This wide range guarantees that our method is relevant to all fields, which makes it a useful resource for scholars, professionals, and students alike.Although identity documents like PAN cards and Aadhar cards were our first emphasis for document classification, current research aims to encompass an even wider variety of document types, significantly increasing the system's utility.

Sector of Education: Our method can be a fantastic ally for students who are having difficulty with complicated research papers, as it enables rapid comprehension of crucial concepts. The entire educational sector is included, in addition to student support. By using the system, educators may increase the material's accessibility and foster an environment that is conducive to improved learning. The features of the system are intended to benefit individual pupils as well as improve comprehension and knowledge retention, which will enhance the instructional environment. By creating a tool for paraphrasing document content, it seeks to improve text comprehension and adaptability while promoting greater accessibility and more understandable communication.

LITERATURE REVIEW

In this study by Siqing Huo, Negar Arabzadeh, and Charles Clarke, the authors investigate methods to automatically validate answers generated by large language models (LLMs) using a corpus. They conduct two experiments using the MS MARCO (V1) test collection and a retrieval pipeline[3]. The method makes use of the advantages of both LLM Model and LangChain to develop a chatbot that can respond to inquiries regarding PDF files .

Xiaoxi Li, Yujia Zhou, and Zhicheng Dou propose UniGen is a Unified Generative system designed for query responding and retrieval. Using the power of big language models, UniGen combines both duties in just one generation model. For question answering and generative retrieval, it uses two separate decoders and a shared encoder.[4].

A substantial amount of modern approaches use the bag-of-words (BOW) approach [5]. Local feature points, such as SIFT [6], are extensively utilized as local features because of their description capabilities. In the second phase, image encoding, BOW were initially employed to encode the feature point distribution in a global image representation [7, 8]. Fisher vectors and VLAD later outperformed the BOW [9, 10]. Pooling has also seen significant advancements; for example, spatial and feature space pooling approaches have been extensively studied [11]. Finally, for the final phase of the pipeline, discriminative classifiers such as linear Support Vector Machines (SVM) are commonly acknowledged as the benchmark in terms of classification performance [12].

Massive classification of image datasets like ImageNet [12] have seen the successful application of deep CNN approaches recently, producing modern outcomes that outperform bag-of-words or Fisher vectors. Compared to standard representations, these networks have a significantly more complex structure, with multiple convolutional layers being followed by fully connected layers, which results in a large number of variables that need to be obtained from the data used for training. An intermediate to high-level hierarchical representation can be extracted by learning these parameters for networks on large image datasets [14]. Additionally, VLAD [15] and Fischer vectors encodings have recently been combined with Deep CNN representations. [16].

The study introduced TM-CNN, a technique for detecting small defects in periodic structures like magnetic labyrinthine patterns. It combined template matching with a convolutional neural network, reducing manual annotation workload, and achieving an impressive F1 score of 0.988 in testing. This approach enhances the understanding of magnetic materials properties and outperforms traditional methods [29].

The research introduced a novel mathematical theory for efficient template matching that simultaneously handled rotations and translations without the need to repeatedly sample the entire space of rotations. This approach integrated rotated versions of the template into a single symmetric tensor template, significantly reducing computational complexity. By correlating the image with

independent tensor components, it accurately recovered template positions and rotations. This method had the potential to greatly accelerate conventional template matching computations, particularly for 3D images [30].

The paper introduces an algorithm for multi-scale and rotated image template matching, addressing the inefficiencies of traditional methods. By computing ring projection vectors and utilizing normalized cross-correlation, it accurately determines optimal matching positions and scale ratios. Experimental results confirm the effectiveness of the proposed approach in locating correct matches for multi-scale and rotated images [31].

Ion Androutsopoulos and Prodromos Malakasiotis, in their paper "A Survey of Paraphrasing and Textual Entailment Methods," cover methods for recognizing, generating, and extracting paraphrases—expressions conveying the same or nearly the same information. They discuss using bilingual parallel corpora for extracting paraphrases, highlighting Bannard and Callison-Burch's (2005) method, which involves using a pivot language to identify paraphrase pairs . [32]

PROBLEM IDENTIFICATION

3.1 Dataset Overview:

Composition and Diversity: The dataset has 340 photos from five different classes, each of which represents a typical mix of identity documents.

Document Type	Aadhar Card	Pan Card	Driving License	Voter Id	Passport
No. of Documents	129	45	64	76	37

Characteristics:The photos show a range of features, such as various lighting settings, positions, and resolutions, which give the model a realistic context to work with.

Img : Images from Training data

Img : Images from Training data

3.2 Data Preprocessing and Augmentation Techniques:

In the field of deep learning, data augmentation is essential, especially for picture classification tasks. It entails creating fresh training samples from preexisting ones in order to artificially increase the dataset. This method is essential for enhancing the robustness and performance of the model, particularly when working with sparse data, as our identification document categorization project does.

For this project, we used TensorFlow's ImageDataGenerator to increase the diversity of our dataset, which includes pictures of different identity documents such as driver's licenses, Aadhar cards, PAN cards, passports, and voter IDs. The following augmentation methods are employed:

- Standardization: To provide uniform input to the model, images were scaled to a uniform size.
- Rescaling: The range is changed between [0, 255] to [0, 1] by dividing each pixel value by 255. During model training, normalization is a common approach that aids in accelerating convergence.
- Rotation: A rotation that went as much as 40 degrees was applied. For document pictures, this transformation is important since it enables the model to learn how to recognize papers in any orientation.
- Shifts in Width and Height: We apply a 20% augmentation to the width and height of the image. These adjustments explain differences in the orientation of papers after scanning or taking pictures.
- Shear Transformation: We train the model to identify documents from diverse viewpoints by simulating the impact of documents being collected from different angles by applying a shear strength of 0.2.
- Zoom: Up to 20% of random zooming is included in the augmentation. This modification replicates the impact of capturing documents at different distances.

- Horizontal Flip: In order to produce mirrored versions of the papers, we incorporated horizontal flips. This adds a kind of variation that can help the model generalize, even though it may not always be realistic for texts with a lot of text.
- Fill Mode: When set to "nearest," this option controls how freshly formed pixels are filled in. freshly created pixels may show up following a shift in width or height or rotation. By selecting the "nearest" option, you may be confident that the values that fill these new pixels closely match those of their surroundings.

Rationale and Benefits:

Enhanced Generalization: We give the model the tools it needs to handle real-world situations where document photos can be taken in a variety of settings and orientations by incorporating these various transformations.

Warfare Overfitting: Overfitting is a typical problem in neural network models learned on small datasets. Augmentation is very helpful in addressing this issue. The model learns to pay attention to pertinent attributes instead than committing particular visual elements to memory by intentionally increasing the variety of the training data.

Dataset Balancing: In light of the disparity in our data set, augmentation contributes to a more balanced process of learning by artificially supplying more instances of underrepresented classes.

3.3 Problem Statement and objectives

The primary challenge addressed by our project is the development of an integrated system for document management that encompasses three core functionalities: document classification, interactive question answering, and template matching for fraud detection. Each of these functionalities poses unique challenges:

Document Classification: Classifying various types of identity documents accurately is crucial for automating verification processes in organizations. Traditional methods, while effective, often require extensive training data and are time-consuming to fine-tune. Moreover, achieving high accuracy across diverse document types with varying formats and content is challenging.

Interactive Question Answering: Extracting relevant information from unstructured PDF documents and providing accurate responses to user queries in real-time requires advanced natural language processing techniques. Existing solutions often struggle with integrating large language models (LLMs) with efficient retrieval systems, leading to inefficiencies in handling large-scale applications.

Template Matching for Fraud Detection: Detecting subtle document forgeries involves comparing a test document against a standard template to identify discrepancies. This process needs to be robust

against variations in text style, logos, and formatting while accurately identifying fraudulent documents. Current methods rely heavily on OCR for textual similarity and lack the ability to fully leverage feature-based template matching, especially for structured documents.

In addition to the core functionalities, our project incorporates several side features to enhance the overall user experience and utility of the system. These features address additional needs in document handling and processing:

- Read Aloud: Utilizing text-to-speech technology, this feature reads documents aloud, improving accessibility for visually impaired users or those who prefer auditory learning.
- Summarization: Automatically generates concise summaries of lengthy documents, allowing users to quickly grasp the main points without needing to read the entire text.
- Grammar Check: Scans the text for grammatical errors and provides suggestions for corrections, ensuring documents are well-written and error-free.
- Paraphrasing: Rewrites text in different ways while preserving the original meaning, which is useful for creating varied versions of a document or avoiding plagiarism.

In summary, the problem statement is to design and implement a comprehensive AI-driven document intelligence system that can:

1. Accurately classify a wide range of identity documents.
2. Provide real-time, accurate responses to user queries based on document content.
3. Detects document fraud by comparing test documents against standard templates using advanced image processing and feature matching techniques.
4. Enhance document accessibility, comprehension, and quality through auxiliary features like read aloud, summarization, grammar checking, and paraphrasing.

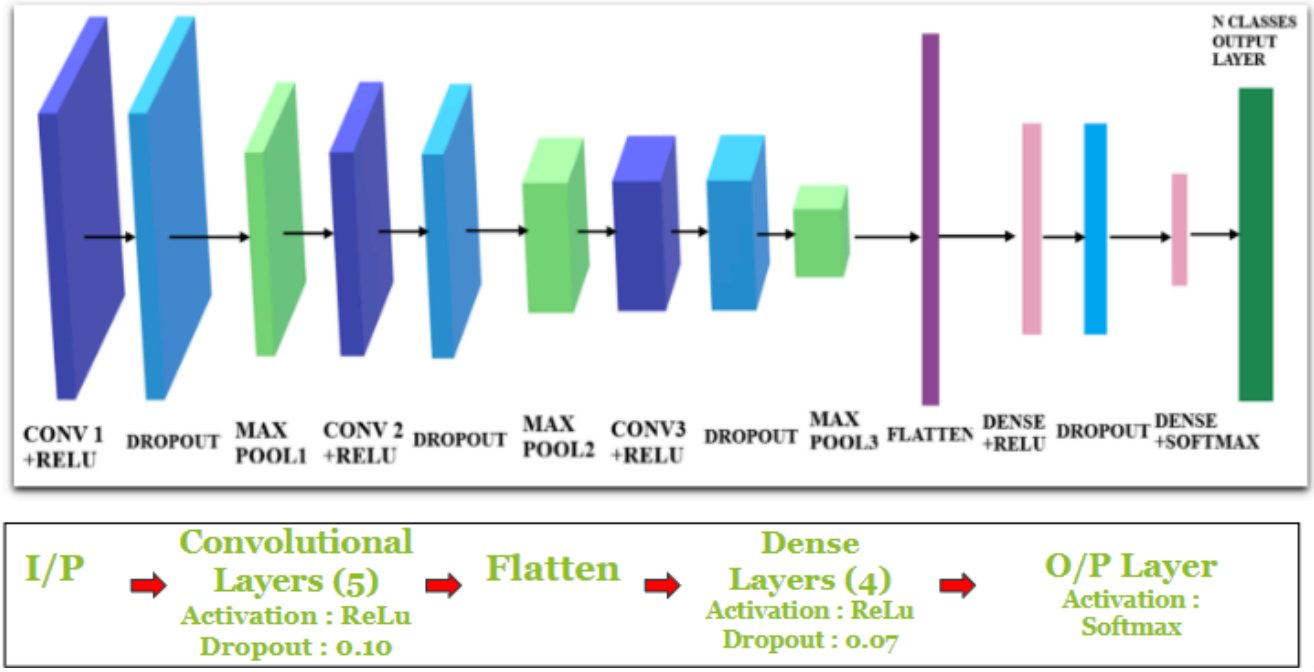
3.4 Models Used

3.4.1 Custom CNN Model:

Our custom Convolutional Neural Network (CNN) architecture is meticulously designed for the task of document classification from PDFs or images containing multiple document types. The model's primary function is to accurately classify individual pages into distinct document categories such as Aadhar, PAN, driving license, passport, and voter ID. This is achieved through a series of carefully considered design choices, including the optimal number of layers, neurons, filters, and kernel sizes. Additionally, extensive hyperparameter tuning is performed to fine-tune these parameters, ensuring that the model strikes a balance between accuracy and computational efficiency.

The architecture consists of five convolutional layers that extract intricate features from the document images. These layers are followed by a flatten layer, dropout layers to prevent overfitting,

and four dense layers that enable the model to learn complex patterns within the data. The dropout layers are particularly essential as they enhance the model's robustness by randomly setting a fraction of input units to zero during training, which helps mitigate the risk of overfitting. Through rigorous training and iterative hyperparameter adjustments, the model proves its effectiveness in accurately identifying and categorizing various document types. This capability significantly advances document processing and information extraction tasks, making the model a valuable tool in automated document classification.



3.4.2 VGG16 Model:

Utilizing the pre-trained VGG16 model, we leverage its deep convolutional neural network architecture to extract features from document images. Originally trained on the massive ImageNet dataset, VGG16 is renowned for its simplicity and efficacy in various computer vision tasks. By employing transfer learning, we harness the knowledge encoded in VGG16's early layers, which capture generic features such as textures and edges. This allows for efficient feature extraction from document images, enhancing the performance of our document classification model. The adoption of VGG16 facilitates robust document classification, enabling accurate identification and extraction of document types across diverse datasets, thereby showcasing its utility in real-world applications. Adding 5 layers to the model The dataset is split into a 1:3 ratio for validation and training

respectively. Early stopping mechanisms are implemented to prevent overfitting during training. Hyperparameters, including the number of layers, neurons, filters, kernel size, dropout rates, epochs, and batch size, are fine-tuned to optimize model performance.

3.4.3 OCR for Information Extraction:

For text extraction and information retrieval from document images, we employ Optical Character Recognition (OCR) using Pytesseract. Prior to text extraction, images undergo a series of preprocessing steps including resizing, conversion to grayscale, and sharpening. These preprocessing steps enhance the image quality, making the text more legible for the OCR process. Once the text is extracted, it is subjected to a cleaning process to remove any extraneous characters or noise.

The cleaned text is then analyzed using document-specific keywords to extract relevant information. This process is tailored to identify and extract key data points pertinent to each document type. The extracted information is subsequently saved in an Excel file for further analysis. This systematic approach to text extraction achieves an accuracy of 88%, underscoring the effectiveness of our preprocessing and OCR techniques in handling a variety of document types and ensuring reliable information retrieval. This combination of advanced image processing, robust feature extraction, and meticulous data handling enables our system to perform efficiently in real-world document classification and fraud detection tasks.

3.4.4 BART, a powerful Transformer Model for Paraphrasing

The primary problem addressed is to generate paraphrased versions of text extracted from PDF documents. This problem can be broken down into several specific challenges:

1. Text Extraction from PDFs: PDF documents are a common format for distributing information, but extracting text from them is often difficult due to their complex structure. Text within PDFs can be formatted in various ways, including embedded fonts, images, and different layouts, making it hard to convert into plain text reliably.
2. Handling Large Text Volumes: Once text is extracted from a PDF, it can be extensive. Processing such large volumes of text in a single block is not feasible due to memory and

computational constraints, especially when using NLP models that have input length limitations.

3. Paraphrasing the Text: The extracted text needs to be paraphrased, which involves rewording the content while preserving the original meaning. This requires sophisticated NLP models capable of understanding and generating natural language.

3.4.5 OpenAI's GPT-based Model for Interactive Question Answering:

The integration of LangChain within OpenAI Language Models represents a significant advancement in the field of natural language processing, particularly in the realm of question-answering systems. LangChain is a framework designed to enhance the capabilities of language models by effectively bridging the tasks of language understanding and generation through a unified approach. This integration is particularly powerful as it leverages various components to deliver precise and contextually appropriate answers to user inquiries, enhancing both the accuracy and reliability of the responses.

A key aspect of LangChain's functionality is its ability to seamlessly combine document-based insights with query matching to generate well-informed answers. This is achieved through a multi-step process involving several specialized tools and techniques. The first crucial component in this process is the CharacterTextSplitter, which plays a vital role in segmenting large blocks of extracted text into smaller, more manageable chunks. This segmentation is essential for ensuring that the subsequent processing stages can handle the text efficiently without losing the contextual coherence necessary for accurate interpretation.

Following text segmentation, the next step involves generating dense vector representations of these text chunks, a task handled by the OpenAIEmbeddings module. This module is instrumental in capturing the semantic meaning of the text, transforming it into a format that the language model can process more effectively. These dense vectors encode the nuanced relationships between different parts of the text, allowing the system to understand and retrieve relevant information based on semantic similarity rather than relying solely on keyword matching.

Once the text chunks have been transformed into their vector representations, the FAISS (Facebook AI Similarity Search) system comes into play. FAISS is a powerful tool for efficient similarity search, enabling the retrieval of the most relevant text chunks from a large corpus based on their proximity to the query in the vector space. This ensures that the information retrieved is highly relevant to the user's question, forming a robust foundation for generating accurate answers.

The final step in this intricate process involves setting up a question-answering chain using LangChain, which leverages OpenAI's advanced language model. This chain orchestrates the entire workflow, from query processing to answer generation. By utilizing the retrieved text chunks, the language model can generate responses that are not only contextually appropriate but also enriched with detailed insights from the source documents. This integrated approach ensures that the answers provided are both precise and informative, addressing the user's inquiry with a high degree of accuracy.

3.5 Performance Metrics

Confidence Score: This technique generates zero-centered log-likelihood ratios. A higher score indicates a high likelihood that a hypothesized word is correct, while a lower score suggests a higher probability of the word being incorrect .

Similarity Score: This score is calculated by evaluating the ratio of high-quality matches (good matches) to the total number of matches (close matches). It ranges from 0 to 1.

SSIM Score (Structural Similarity Index): This index measures the similarity between images based on structural information, brightness, and contrast. Its range is from -1 to 1, with higher scores indicating greater similarity .

Train Accuracy: Train accuracy is the percentage of correctly classified instances in the training dataset. It is used during the training phase to assess a machine learning model's performance .

Train Loss: Train loss measures the discrepancy between the actual targets and the predicted outputs of a machine learning model on its training dataset .

Validation Accuracy: Validation accuracy is the proportion of correctly classified instances in the validation set. It evaluates how well the model generalizes to previously unseen data .

Validation Loss: Validation loss measures the difference between the actual targets and the predicted outputs in the validation dataset. A low validation loss indicates good model generalization, while a high validation loss suggests overfitting .

PROPOSED SOLUTION

4.1 Document Classification

4.1.1 Training the Model

During a crucial phase known as model training, the neural network learns to connect input data—in this example, identity documents—to the anticipated outputs—classification labels. This method includes choosing an optimization tool, the loss function, splitting the dataset, and regularization techniques.

Adam in the Role of Optimizer: Adam has been chosen to be the optimizer. Adam is an adaptive optimization technique that adjusts training rates. This adaptability encourages efficient convergence and improved model performance.

Entropy Cross-Over As the Loss Function, categorical: Categorical cross-entropy is used as the loss function. This is suitable for multi-class classification jobs since it calculates the difference among the actual and anticipated distributions of probability of the classes.

Dividing the Dataset: Training is done at a 3:4 ratio, whereas validation is done at a 1:4 ratio. This ensures that the algorithm is trained on a significant portion of the input information and has a different set for validation in order to track generalization.

Early stopping for Regularization: An early stopping method is employed as a regularization technique to prevent overfitting. By stopping the training process when the model's accuracy on the validation set no longer improves, it avoids superfluous complexity.

4.1.2 Hyperparameter Tuning

The practice of optimizing the performance of a model by changing multiple variables that are not found during training is known as hyperparameter tuning. This step is crucial to achieving the maximum degree of model correctness.

Variable hyperparameters to consider: Parameters like functions for activation, batch size, drop amounts, kernel size, and others can all be changed. These choices have an impact on the model's design and learning procedure.

How to Display Model Performance using Graphics: Graphs that display the model's precision and loss patterns over time can help one comprehend how the model learns. This graphic facilitates the process of hyperparameter tuning by helping to strike a balance among underfitting and overfitting.

4.1.3 Feature Extraction

The fundamental stage of feature extraction, which entails autonomously and hierarchically extracting important information from input images, makes use of convolutional neural networks, or CNNs.

CNNs are chosen on how well they can decipher intricate patterns and characteristics from identity documents. For the model to be able to extract meaningful information from the improved and diverse dataset, this step is crucial. CNN's hierarchy learning approach ensures that the model can adapt to the complex features of identity documents.

4.1.4 Choosing the Final Model

Measures of Accuracy Attained: The chosen VGG16 model achieves 90.51% training accuracy and 87.69% validate accuracy. These measurements show how well the model classified identifying documents during training and on unseen data.

4.1.5 Pre-trained VGG16 Model

Architecture of VGG16: the VGG16 model is a deeply convolutional neural network design that has gained popularity due to its effectiveness and simplicity. Its initial training set was the enormous ImageNet dataset, which included millions of photos in several classes.

Transfer Learning: During ImageNet training, a pre-trained model like VGG16 can impart its knowledge to us. The early layers of these models capture generic features (materials, edges) that are applicable to a range of computer vision applications.

The already trained VGG16 model is used to extract features. This means that instead of beginning from the start while training a deep network, we use the information embodied in the model that has been trained to extract important characteristics from our document images.

4.1.6 Choosing Number of Epochs and Early Stopping

Epochs & Overfitting: The dataset is processed through multiple times, or epochs, during neural network training. When an algorithm fits the data used for training too precisely and performs poorly on new data, this is known as overfitting. Using excessive epochs may be the cause of this.

Early Stopping: The use of early stopping is a regulatory approach that tracks an indicator on a validation set and stops training when performance starts to decline in order to prevent overfitting.

In this case, choosing 15 epochs is a good balance between allowing the model to grow and preventing overfitting. This number could need to be adjusted depending on the size of the information set, the intricacy of the algorithm, and the available processing capacity.

Early Stopping: The early stopping response is crucial to preventing overfitting. If, after a predefined number of epochs (five in this example), the loss during validation stops decreasing, training is terminated and the model's variables are retained at the best validated performance. This protects the model from deteriorating on unidentified data.

4.1.7 Model Fine-tuning Possibilities

Hyperparameter tuning: Variations in learning rate, dropout rate, or layer size are made to improve model performance.

Examining Model Architecture: More complex architectures, including deeper artificial neural networks or different pre-trained models, can be studied for potentially better results.

Hyperparameter Adjustment: More advancement can be possible by modifying hyperparameters like the rate of learning or dropout rate. This strategy may include experiment and validation set monitoring of performance.

Model Architecture: Depending on the characteristics of the document dataset, experimenting with a number of already trained models (such as VGG19 and ResNet) or developing a bespoke architecture may yield better results. The level of complexity of the different types of documents and the available processing power could have an impact on the choice.

4.2 Interactive Query Answering

Embeddings generated from the PDF documents and user queries are used to implement a generative question-answering system. The objective is to find out relevant documents using cosine similarity and extract answers through a question-answering model. This enables users to ask questions and get answers based on the contents of the PDFs in an interactive manner.

4.2.1 PDF Documents Vector Embeddings

The PDF documents are transformed into texts, which are then cut into small chunks. After extracting the text content, it is further broken down into smaller units by use of the CharacterTextSplitter object. This is done so as to ensure that the sizes of these text chunks are

manageable enough for processing, while preventing memory issues as well as enabling efficient computation by not exceeding certain token limits.

4.2.2 Vector Embeddings from LLM

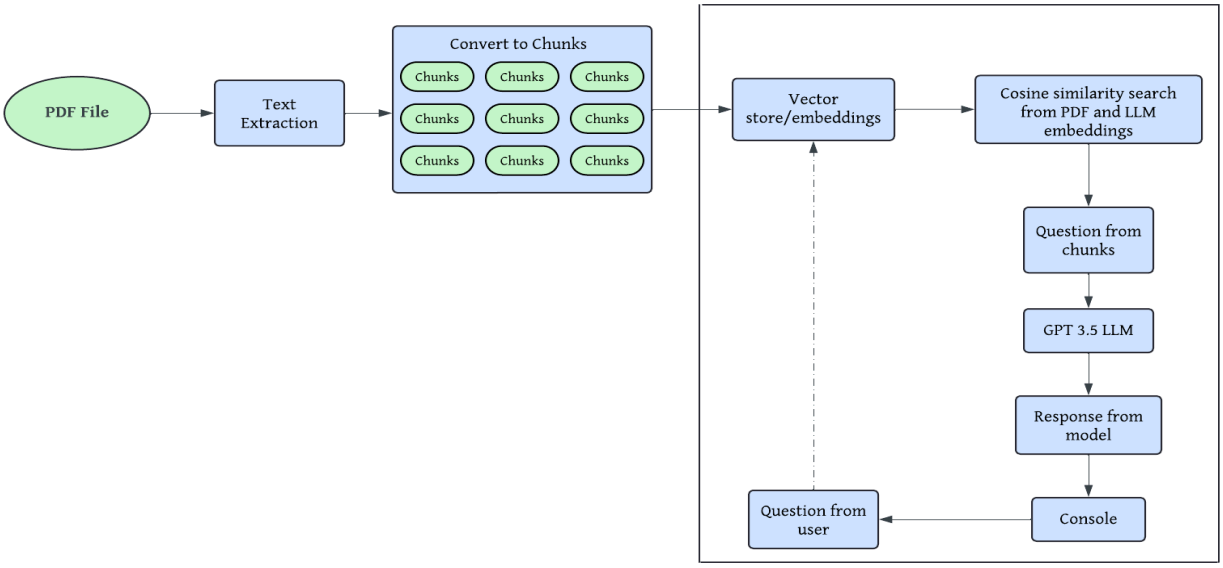
Additionally, OpenAI's GPT-based model has been used via its OpenAIEmbeddings class to produce embeddings for each part of text that have been split before. These embeddings thus represent query's semantic meaning within this same high-dimensional vector space where they dwell alongside those coming from the documents contained in PDFs.

4.2.3 Calculation of Cosine Similarity

The embeddings for both the chunks of PDF documents and the query are obtained, then cosine similarity is calculated by comparing the embeddings of the query to each chunk of text from the PDF documents. It measures the cosine angle between two vectors and is a well-known statistic used for comparing vector similarity in high dimensional spaces.
The range of cosine similarity score is -1 to 1, where a value that closer to 1 indicates vectors are closer and a value that closer to -1 means dissimilarity.

4.2.4 Searching for Similarity and Relevance Ranking

The chunks of text from PDF documents are ordered according to their cosine similarity scores with the query. Documents having higher cosine similarity scores are deemed more relevant to it. These documents are then fed into a question-answering model which further analyzes them. The question-answering model takes into account content in most relevant documents as it tries extracting answers for user's questions from these similar wordings.



4.3 Similar Document Template Matching Algorithm

4.3.1 Template Extraction

ROI Extraction:

Regions of Interest (ROIs) were extracted from document pictures to begin the procedure. To make edge identification a bit simpler, the images were transformed to grayscale using OpenCV. After applying sharp edge detection to draw attention to important details in the pictures, contour detection was used to separate out specific areas.

With the development of dynamic thresholding algorithms, edge fidelity and noise reduction were enhanced by their ability to adjust to differences in quality of images.

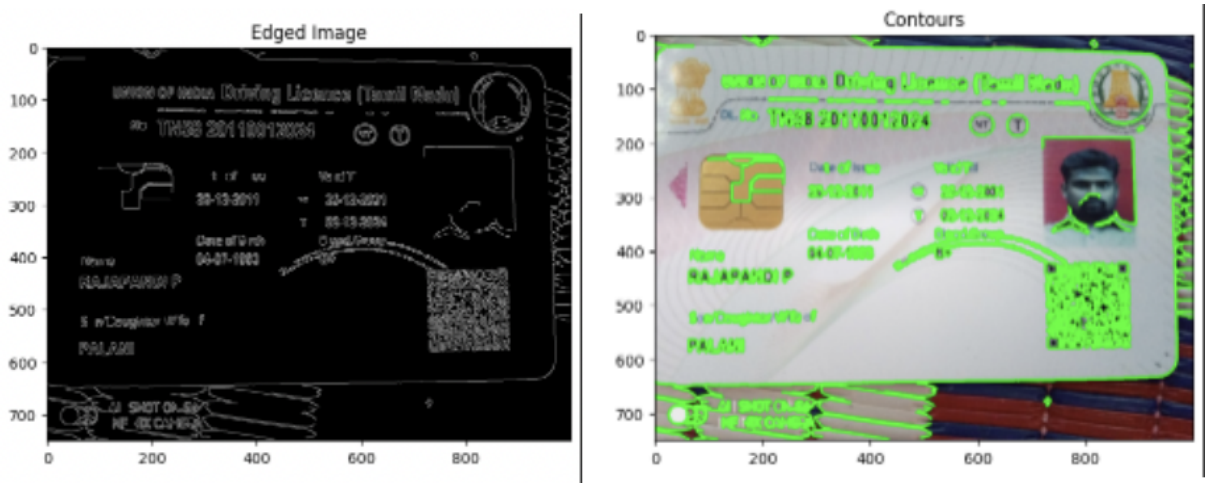


Fig : edge detection and contrast enhancement

Preprocessing of Extracted ROIs:

To improve feature identification, a number of preliminary processing procedures were applied across every ROI. This involved binary thresholding to get the regions ready for extraction of features, reducing noise, and tone enrichment using CLAHE (Contrast Limited Adaptive Histogram Equalization).

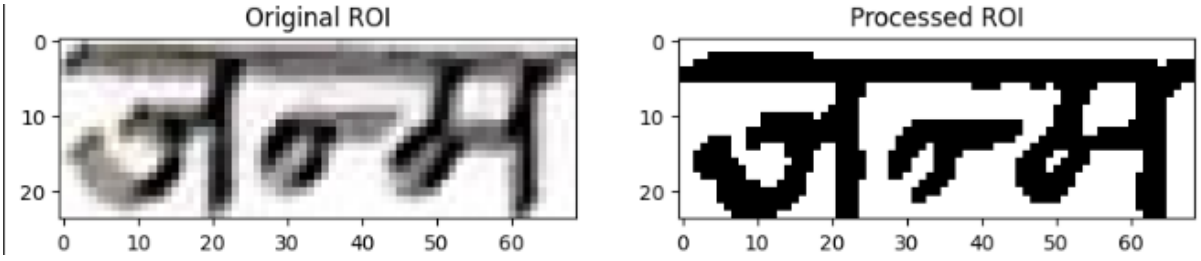


Fig : Processed ROI of input image

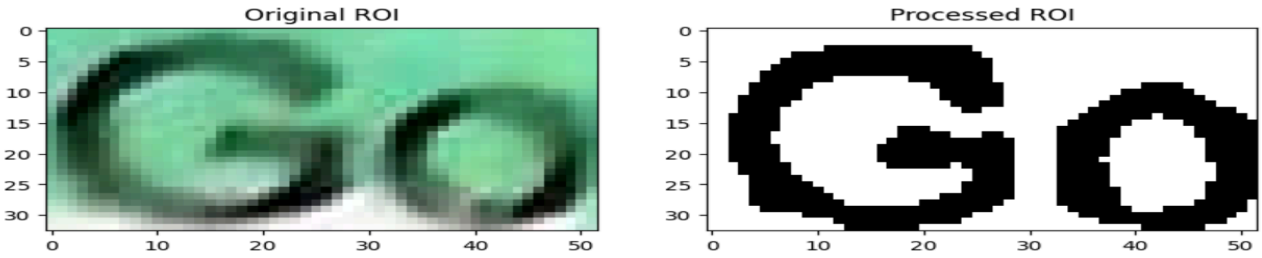


Fig : Feature extraction

Fig : Real and Fake Document

Feature Extraction:

ORB (Oriented FAST and Rotated BRIEF) descriptor was employed due to its performance and efficiency in rotation- invariant comparison.The extracted features were quickly extracted in form of key points(interesting points in an image) and descriptors (description of these points).

4.3.2 Template Matching & Comparison

The features obtained in the previous step in the form of key points and descriptors are passed to the function(match features) which finds two things: Total close matches and Good Matches. The most closely comparable characteristics between the test document and a predetermined authentic document are called Total close matches. This is obtained using a Brute-Force or BF matcher with the Hamming distance.Further, Good matches or

exceptional similarities are identified using the nearest neighbors ratio out of the Total close matches.

4.3.3 Fraud Detection

Similarity scoring

The algorithm's ability to distinguish between authentic and falsified documents is measured by the similarity_score, which is obtained using the percentage of high-quality similarities over the entire set of matches. The value of this score indicates how successful the match is. To evaluate the layout of the document pictures even more, the Structural Similarity Index (SSIM) was incorporated. The resultant statistic offered details about the texture and structural analogies among the document with its template, serving as an addition to feature-based comparison.

Decision Making

In order to maximize the ability to distinguish between genuine and counterfeit documents, decision-making parameters such as match_threshold and similarity threshold were progressively changed, resulting in a binary decision output.

The result of the operation is shown as either "probably genuine" or "potentially fake" based on this binary output.

4.3.4 Iterative Development

Hyperparameter Optimization

High false positive rates were one of the shortcomings discovered through initial analyses on an experimental dataset. These were fixed by improving the preliminary processing phases and modifying feature comparison limits.

- ROI Extraction: To better obtain significant characteristics, the blur kernel size and Canny thresholds were tweaked.
- Preprocessing Steps: Adaptive thresholding and improved Gaussian blur parameters to improve feature definition.
- Feature Matching: To increase matching integrity, the k-value and matching ratio in kNN were adjusted.
- Decision Criteria: To improve the identification of authentic documents, match and similarity thresholds in make_decision were modified.

This solution stands out in the document authentication sector since it uses one standard format for all document types, enhanced by adaptive thresholding and ORB feature recognition. The technique's utility for analyzing documents tasks is demonstrated by the promising early results on a limited set of documents. These early findings point to the possibility of wider usage, but additional testing with a larger number of samples may be required to accurately assess and improve its efficacy.

4.4 Paraphrasing

To tackle the problem of generating paraphrased text from PDF documents, our solution employs a multi-step approach that leverages a state-of-the-art NLP model, specifically the BART (Bidirectional and Auto-Regressive Transformers) model from Hugging Face's Transformers library. This approach ensures the efficient extraction, processing, and paraphrasing of text while maintaining high-quality output.

The solution can be broken down into the following key steps:

1. Text Extraction from PDFs:
    - The first step involves extracting text from PDF documents. This is handled by the upload_pdf_and_convert(menu_option) function, which converts the PDF content into plain text. This step is crucial as PDFs often have complex structures, and reliable extraction ensures that we have a clean, usable text for further processing.
2. Chunking the Text:
    - Given the limitations of NLP models like BART, which can process a maximum of 1024 tokens in a single pass, the extracted text must be split into smaller chunks. The chunk_text function is designed for this purpose. It splits the text into chunks that are within the model's token limit, ensuring that each chunk retains as much context as possible without exceeding the input size constraints. This step is vital for handling large volumes of text efficiently.
3. Paraphrasing the Text:
    - The core of our solution lies in the paraphrasing step, handled by the paraphrase_text_f function. This function utilizes the BART model (facebook/bart-large-cnn), a pre-trained transformer model known for its effectiveness in tasks such as summarization and paraphrasing. The BART model combines the strengths of both bidirectional and auto-regressive transformers, making it exceptionally good at understanding and generating natural language.
    - In this function, each text chunk is prefixed with a paraphrasing prompt ("paraphrase: ") and tokenized using the BART tokenizer. The tokenized input is then fed into the

BART model, which generates paraphrased text using beam search. Beam search is a technique that explores multiple possible output sequences and selects the most likely one, enhancing the quality of the generated text. The model's output is decoded back into human-readable text, and each paraphrased chunk is collected.

4. Reconstruction and Output:
    - After paraphrasing each chunk, the paraphrased chunks are concatenated to form the final paraphrased text. The text is then split into sentences and printed line by line to ensure readability. This step ensures that the paraphrased content is presented in a coherent and user-friendly format.

Key Metrics for Evaluating the Solution

1. Semantic Fidelity: The paraphrased text should retain the original meaning and information content.
2. Fluency and Readability: The paraphrased output should be grammatically correct and easy to read.
3. Variety: The paraphrased text should differ sufficiently from the original text to be considered a true paraphrase.
4. Efficiency: The solution should handle large text volumes efficiently without significant performance degradation.
5. Robustness: The approach should work reliably across different types of PDF documents and content styles.

4.5 Read Aloud

The "Read Aloud" feature in our system provides an alternative and efficient way for users to consume document content through auditory means. This is particularly beneficial for individuals with visual impairments or those who prefer auditory learning. The feature converts the text from a PDF document into spoken words, which can be listened to as an audio file.

Steps of the Read Aloud Feature

1. Prompt for PDF Upload: The system prompts the user to upload a PDF file that they wish to convert to audio.

2. Extract Text from PDF:
    - The upload_pdf_and_convert function is called with a menu option (in this case, 1), which handles the uploading of the PDF and converts its content into text.
    - The extracted text is stored in a dictionary format.
3. Clean the Extracted Text:
    - The extracted text often contains newline characters (\n). These are replaced with spaces to ensure the text flows naturally when read aloud.
    - This cleaned text is stored in a list format.
4. Combine the Text:
    - All pieces of cleaned text are combined into a single string. This ensures that the entire content of the PDF is included in the audio output without segmentation.
5. Convert Text to Speech:
    - The combined text string is passed to the Google Text-to-Speech (gTTS) engine, which converts the text into spoken words.
    - The spoken words are saved as an audio file (output.mp3).
6. Play the Audio:
    - The system provides an audio player that automatically plays the generated audio file, allowing the user to listen to the content of the PDF immediately.

4.6 Grammar Check

Grammar checking is performed using advanced NLP techniques, leveraging transformer-based models for accurate and context-aware corrections. The grammar check feature is designed to process text extracted from a PDF document, identify grammatical errors, and provide corrected text. This is achieved through the following steps:

4.5.1 Text Extraction from PDF:

The first step involves extracting text from PDF documents. This is handled by the upload_pdf_and_convert (menu_option) function, which converts the PDF content into plain text. This step is crucial as PDFs often have complex structures, and reliable extraction ensures that we have a clean, usable text for further processing.

4.5.2 Grammar Correction Using Transformers:

The grammar_checker function takes the concatenated text and processes it using a transformer-based model from Hugging Face's Transformers library for grammar correction.

Model Selection: The model used is pszemraj/flan-t5-large-grammar-synthesis, a variant of the T5 model fine-tuned for grammar correction tasks.

Pipeline Initialization: The pipeline function from the Hugging Face Transformers library is used to

initialize the model for text generation tasks. In this case, it generates corrected text based on the input.

Model Initialization: The Hugging Face pipeline function is used to initialize a text generation model specifically fine-tuned for grammar correction (pszemraj/flan-t5-large-grammar-synthesis).

Grammar Correction: The input text is passed through the model, which processes it and returns a grammatically corrected version. The results[0]['generated_text'] retrieves the corrected text from the model's output.

DESIGN OF PROPOSED SYSTEM

Our system seamlessly integrates three core functionalities to provide a comprehensive solution for document handling and security.

1. **Document Classification** : For identifying ID documents such as passports or Aadhaar cards, we use a Convolutional Neural Network (CNN) model, specifically VGG16 with custom layers. This model classifies the documents accurately, ensuring they are correctly identified and categorized.

2. **Question and Answer System** : For general documents, our system enables users to ask specific questions about the content. Using Pytesseract for Optical Character Recognition (OCR), the text from the documents is extracted. Then, Langchain processes these texts to provide immediate, relevant answers. Users can continue asking follow-up questions until they have all the information they need, creating an interactive and user-friendly experience.

3. **Fraud Detection** : For detecting fraudulent documents, users can upload documents which are then compared against a standard template using Oriented FAST and Rotated BRIEF (ORB) descriptors. The similarity between the uploaded document and the template is evaluated using similarity scores and Structural Similarity Index (SSIM). This process ensures the authenticity of the documents by detecting any discrepancies.

Each of these processes operates independently, yet they collectively contribute to a robust and secure framework. This integration enhances user interaction by providing precise document classification, a dynamic Q&A system, and reliable fraud detection, ensuring both efficiency and security.
Our project also includes additional features implemented using Python libraries to further enhance user experience:

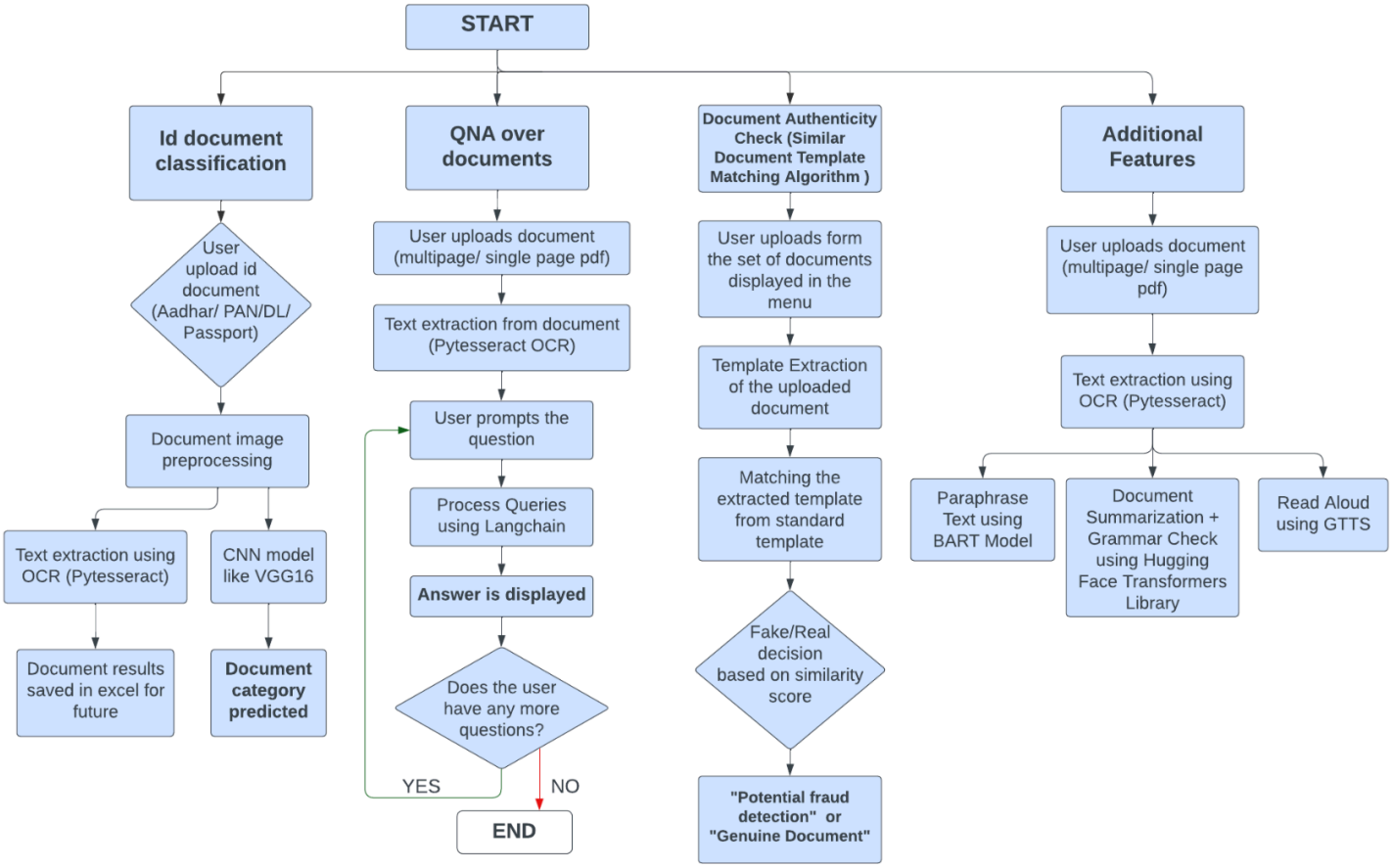
- **Read Aloud**: This feature uses text-to-speech technology to read documents aloud to users, improving accessibility for visually impaired users or those who prefer auditory learning.

- **Document Summarization**: Automatically generates concise summaries of long documents, helping users quickly grasp the main points without reading the entire text.

- **Grammar Check**: Scans the text for grammatical errors and provides suggestions for corrections, ensuring that documents are well-written and error-free.

- **Paraphrasing**: Rewrites text in different ways while preserving the original meaning, which can be useful for creating varied versions of a document or avoiding plagiarism.

These side features further enhance the system's utility, making it a comprehensive tool for document management and security.



Flow of our Document Intelligent System

IMPLEMENTATION AND RESULTS

6.1 Requirements and Specifications

The system requires Python, TensorFlow, OpenCV, Langchain, FAISS, and Pytesseract. Hardware requirements include a GPU for training models and adequate storage for datasets and model outputs.

All the code implementation and experiments have been conducted on Google Colab version 1.0.0.

6.2 Pipeline Outcomes

The CNN achieved a training accuracy of 64.44% and a validation accuracy of 83.33%. Refining VGG16 led to significant improvements, with training and validation accuracies reaching 90.51% and 87.69%, respectively. This demonstrates our model's enhanced ability to generalize from limited data through optimized hyperparameter tuning

Hyperparameter Tuning Results for CNN model

Hyperparameters Tuned	Train Loss	Train Accuracy	Validation Loss	Validation Accuracy
Kernel size (3,3)	68.01	69.44	67.23	83.33
Activation function (Sigmoid)	59.14	69.44	53.38	83.33
Dropout (0.05)	1.09	51.90	96.05	64.15
Dropout (0.07)	57.31	73.42	74.99	77.36
Batch size (64)	1.34	40.00	1.42	25.00
Batch size (128)	1.43	33.85	1.44	34.09

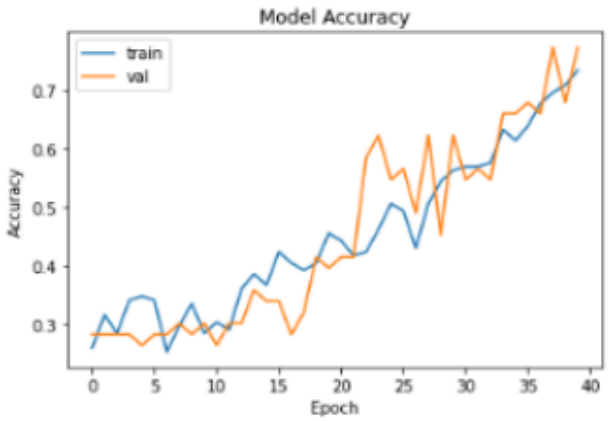


Fig. 2 : Train and Validate Accuracy at each Epoch

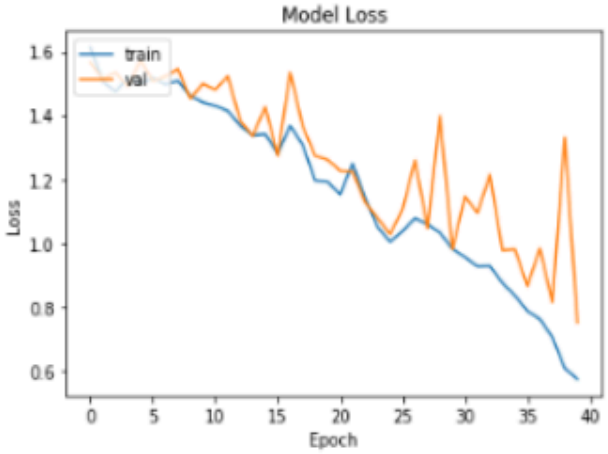


Fig. 3: Train and validation loss at each epoch

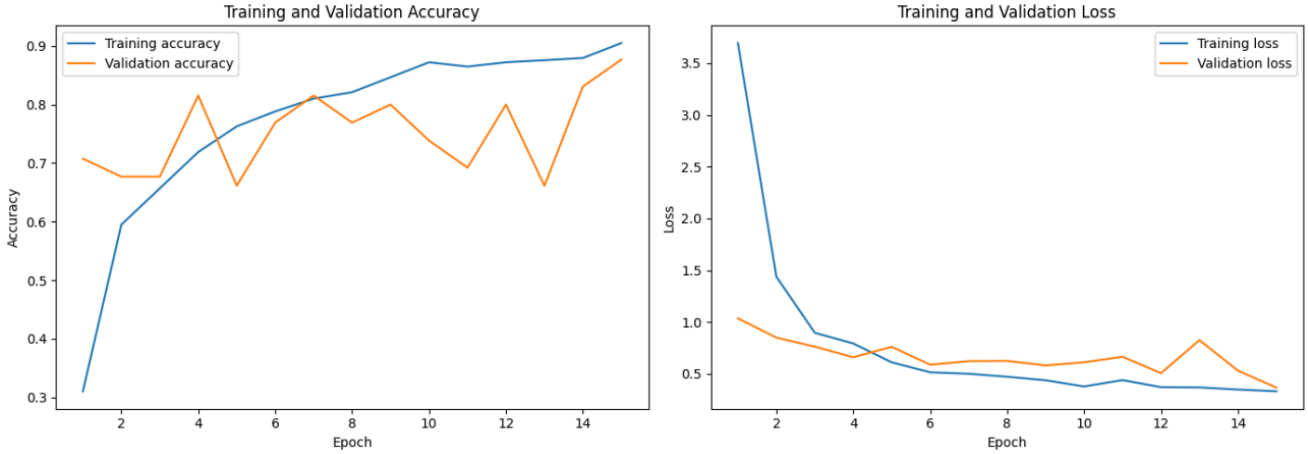


Fig : Accuracy and loss graph for finalized VGG16 model

Comparative Analysis of CNN and VGG16 model

Model	Number of layers	Epochs	Parameter	Train Accuracy (in %)	Validation Accuracy (in %)	Observation
CNN	9	40	Dropout = 0.07	69.44	83.33	Limited data augmentation and a smaller network architecture
VGG16	Additional 5	40	Learning rate = 1e-5	90.51	87.69	Better generalisation leading to improved accuracy

In assessing our Q&A system, LLM models consistently outperformed BERT, maintaining confidence scores above 90% for documents up to 12 pages.

Comparison between LLM and BERT

Length of pdf (in pages)	Confidence Score (LLM Model)	Confidence Score (BERT Transformer)
1	98.43	80.04
4	94.46	56.51
12	92.85	55.46

Turning to our document fraud detection, we evaluated several documents for authenticity. For instance, Aadhar_img1, showing significant alterations, was rightly flagged as potentially fraudulent with a low SSIM score of 0.29. In contrast, Driver_img]2, which is color invariant and only had format changes, was correctly identified as genuine. This showcases our system's precision in handling variations and ensuring reliable fraud detection.

Analysis of Document Authenticity through Feature Matching and SSIM Evaluation

Image	Key features	SSIM	Decision Obtained	Expected Decision	Observation
aadhar_img1	With display picture variation, no qr code, missing name in hindi	0.29	Potential fraud detected	Fraud	Significant changes detected very well
driver_Img2	Black and white, rotated	0.54	Document is likely genuine	Real	Color invariant
passport_img3	Face hidden	0.20	Fraud	Fraud	Recognises inconsistencies with Display image in the id
aadhar_img4	Date font size changes and logo missing	0.20	Potential fraud detected	Fraud	Worked on subtle variation Like text font and size
aadhar_img5	Colored, well aligned and illuminated	0.21	Potential fraud detected	Real	False positives are high. Here additional checks can be employed like re-uploading etc
pan_Img6	Colored, well aligned and illuminated	0.51	Document is likely genuine	Real	Ideal document

Comparative Analysis of Grammar Check Libraries

Library	**LanguageTool**	**Hugging Face's Transformers Pipeline**
Accuracy	0.93	0.96

CONCLUSION

In conclusion, our research has culminated in the development of a comprehensive Document Interaction Assistant equipped with multiple features to streamline document management processes. It is an assistant that responds interactively to questions, classifies documents and matches templates to detect fraud. This way, it significantly minimizes time wastage and enhances accuracy across various sectors.

The interactive question answering system allows users to access relevant information from documents fast without going through all the necessary manual searches. Also, the feature of document classification helps in organizing and sorting out different types of documents for easier handling.

One of its most remarkable qualities is its ability to match templates, which it uses to identify fraudulent cases. Our system compares documents against a database of templates which allows it to identify disparities and irregularities taking place hence minimizing the risk of fraudulent activities. Overall, our work emphasizes how intelligent document interaction systems can transform industry-wide practices in managing documents.

FUTURE SCOPE

In its development, the Document Interaction Assistant (DIA) presents several directions for future research and innovation that promise greater functionality, ease of use as well as improved security.

The global linguistic needs of users can be addressed by extending the reach of DIA through multi-language support. To allow seamless interaction and document processing across different languages to cater for a diverse user base DIA needs to incorporate natural language processing capabilities for multiple languages.

To provide voice-controlled document processing it would be worth exploring integration with popular AI assistants such as Siri, Alexa or Google Assistant. This will make DIA more convenient to end-users enabling them to communicate with it using voice commands more easily.

Document security, integrity, and traceability can be improved by considering blockchain technology integration. The DIA will also employ blockchain technology in ensuring tamper proof documents processed through it; therefore, enhancing trust and reliability in handling of documents.

Customization options should be implemented together with personalized recommendations based on user preferences, document history and usage patterns so that the platform experience can be tailored around individual user needs.

With the inclusion of data analytics and reporting features, users will therefore be able to gain an understanding of the way their documents are used, metrics on productivity as well as aspects that need improvement. When users study the data generated from this platform they can make their document workflow more effective thus increasing efficiency and effectiveness.

It is possible for DIA to become a universal and highly useful instrument for managing papers in different areas by moving in these directions.

REFERENCES

[1] Pesaru, Arjun, Taranveer Singh Gill, and Archit Reddy Tangella. "AI assistant for document management Using Lang Chain and Pinecone." *International Research Journal of Modernization in Engineering Technology and Science* (2023).

[2] Kim, Jaehyung, Jaehyun Nam, Sangwoo Mo, Jongjin Park, Sang-Woo Lee, Minjoon Seo, Jung-Woo Ha, and Jinwoo Shin. "SuRe: Improving Open-domain Question Answering of LLMs via Summarized Retrieval." In *The Twelfth International Conference on Learning Representations*. 2023.

[3] Huo, Siqing, Negar Arabzadeh, and Charles Clarke. "Retrieving supporting evidence for generative question answering." In *Proceedings of the Annual International ACM SIGIR Conference on Research and Development in Information Retrieval in the Asia Pacific Region*, pp. 11-20. 2023.

[4] Li, Xiaoxi, Yujia Zhou, and Zhicheng Dou. "UniGen: A Unified Generative Framework for Retrieval and Question Answering with Large Language Models." In *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, no. 8, pp. 8688-8696. 2024.

[5] Csurka, Gabriella, Christopher Dance, Lixin Fan, Jutta Willamowski, and Cédric Bray. "Visual categorization with bags of keypoints." In Workshop on statistical learning in computer vision, ECCV, vol. 1, no. 1-22, pp. 1-2. 2004.

[6] Lowe, David G. "Object recognition from local scale-invariant features." In Proceedings of the seventh IEEE international conference on computer vision, vol. 2, pp. 1150-1157. Ieee, 1999.

[7] De Las Heras, Lluis-Pere, Oriol Ramos Terrades, Josep Llados, David Fernandez-Mota, and Cristina Canero. "Use case visual bag-of-words techniques for camera based identity document classification." In 2015 13th International Conference on Document Analysis and Recognition (ICDAR), pp. 721-725. IEEE, 2015.

[8] Kumar, Jayant, and David Doermann. "Unsupervised classification of structurally similar document images." In 2013 12th International Conference on Document Analysis and Recognition, pp. 1225-1229. IEEE, 2013.

[9] Jégou, Hervé, Florent Perronnin, Matthijs Douze, Jorge Sánchez, Patrick Pérez, and Cordelia Schmid. "Aggregating local image descriptors into compact codes." IEEE transactions on pattern analysis and machine intelligence 34, no. 9 (2011): 1704-1716.

[10] Perronnin, Florent, Jorge Sánchez, and Thomas Mensink. "Improving the fisher kernel for large-scale image classification." In Computer Vision–ECCV 2010: 11th European Conference on Computer Vision, Heraklion, Crete, Greece, September 5-11, 2010, Proceedings, Part IV 11, pp. 143-156. Springer Berlin Heidelberg, 2010.

[11] Lazebnik, Svetlana, Cordelia Schmid, and Jean Ponce. "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories." In 2006 IEEE computer society conference on computer vision and pattern recognition (CVPR'06), vol. 2, pp. 2169-2178. IEEE, 2006.

[12] Chen, Siyuan, Yuan He, Jun Sun, and Satoshi Naoi. "Structured document classification by matching local salient features." In Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012), pp. 653-656. IEEE, 2012.

[13] Deng, Jia, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. "Imagenet: A large-scale hierarchical image database." In 2009 IEEE conference on computer vision and pattern recognition, pp. 248-255. Ieee, 2009.

[14]Oquab, Maxime, Leon Bottou, Ivan Laptev, and Josef Sivic. "Learning and transferring mid-level image representations using convolutional neural networks." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1717-1724. 2014.

[15] Arandjelovic, Relja, Petr Gronat, Akihiko Torii, Tomas Pajdla, and Josef Sivic. "NetVLAD: CNN architecture for weakly supervised place recognition." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 5297-5307. 2016.

[16] Cimpoi, Mircea, Subhransu Maji, and Andrea Vedaldi. "Deep filter banks for texture recognition and segmentation." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 3828-3836. 2015.

[17] Eglin, Véronique, and Stephane Bres. "Document page similarity based on layout visual saliency: application to query by example and document classification." In Seventh International Conference on Document Analysis and Recognition, 2003. Proceedings., pp. 1208-1212. IEEE, 2003.

[18] Xing, Chao, Dong Wang, Xuewei Zhang, and Chao Liu. "Document classification with distributions of word vectors." In Signal and information processing association annual summit and conference (APSIPA), 2014 asia-pacific, pp. 1-5. IEEE, 2014.

[19] Lai, Siwei, Liheng Xu, Kang Liu, and Jun Zhao. "Recurrent convolutional neural networks for text classification." In Proceedings of the AAAI conference on artificial intelligence, vol. 29, no. 1. 2015.

[20] Brunelli, Roberto. *Template matching techniques in computer vision: theory and practice*. John Wiley & Sons, 2009.

[21] Jiang, Xingyu, Jiayi Ma, Guobao Xiao, Zhenfeng Shao, and Xiaojie Guo. "A review of multimodal image matching: Methods and applications." *Information Fusion* 73 (2021): 22-71.

[22] Liu, B.; Shu, X.; Wu, X. Fast screening algorithm for rotation invariant template matching. In Proceedings of the 2018 25th IEEE International Conference on Image Processing (ICIP), Athens, Greece, 7–10 October 2018; pp. 3708–3712. [**Google Scholar**]

[23] Lee, H.; Kwon, H.; Robinson, R.M.; Nothwang, W.D. DTM: Deformable template matching. In Proceedings of the 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Shanghai, China, 20–25 March 2016; pp. 1966–1970. [**Google Scholar**]

[24] Lan, X.; Zhu, X.; Gong, S. Person search by multi-scale matching. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 536–552. [**Google Scholar**]

[25] McLaughlin, N.; Ming, J.; Crookes, D. Largest matching areas for illumination and occlusion robust face recognition. *IEEE Trans. Cybern.* **2016**, *47*, 796–808. [**Google Scholar**] [**CrossRef**] [**PubMed**]

[26] Mudunuri, S.P.; Biswas, S. Low resolution face recognition across variations in pose and illumination. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *38*, 1034–1040. [**Google Scholar**] [**CrossRef**] [**PubMed**]

[27] Ma, J.; Jiang, X.; Fan, A.; Jiang, J.; Yan, J. Image matching from handcrafted to deep features: A survey. *Int. J. Comput. Vis.* **2021**, *129*, 23–79. [**Google Scholar**] [**CrossRef**]

[28] T. Kawanishi, T. Kurozumi, K. Kashino, S. Takagi. “A fast template matching algorithm with adaptive skipping using inner subtemplates’ distances,” In Proc of ICPR. March, 2004, pp. 1051-4651.

[29] Okubo, Vinícius Yu, Kotaro Shimizu, B. S. Shivaram, and Hae Yong Kim. "Characterization of Magnetic Labyrinthine Structures through Junctions and Terminals Detection using Template Matching and CNN." arXiv preprint arXiv:2401.16688 (2024).

[30] Almira, José María, Harold Phelippeau, and Antonio Martinez-Sanchez. "Fast Normalized Cross-Correlation for Template Matching with Rotations." arXiv preprint arXiv:2311.07561 (2023).

[31] Qi, Xinwei, and Ligang Miao. "A template matching method for multi-scale and rotated images using ring projection vector conversion." In *2018 IEEE 3rd International Conference on Image, Vision and Computing (ICIVC)*, pp. 45-49. IEEE, 2018.

[32] Androutsopoulos, Ion, and Prodromos Malakasiotis. "A survey of paraphrasing and textual entailment methods." *Journal of Artificial Intelligence Research* 38 (2010): 135-187.

APPENDIX

1. Plagiarism Report
2. Research Paper - AI-Driven Document Intelligence: A Comprehensive Approach to Classification, Q&A Handling, and Fraud Detection
3. Acceptance letter - ICAMC 2024
4. Registration Copy
5. Award for Best paper presentation

Indira Gandhi Delhi Technical University for Women

PLAGIARISM REPORT

ORIGINALITY REPORT

11%	6%	4%	6%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Indira Gandhi Delhi Technical University for Women Student Paper	6%
2	link.springer.com Internet Source	1%
3	hal.archives-ouvertes.fr Internet Source	<1%
4	hal.inria.fr Internet Source	<1%
5	arxiv.org Internet Source	<1%
6	Submitted to University of Queensland Student Paper	<1%
7	Xinwei Qi, Ligang Miao. "A Template Matching Method for Multi-Scale and Rotated Images Using Ring Projection Vector Conversion", 2018 IEEE 3rd International Conference on Image, Vision and Computing (ICIVC), 2018 Publication	<1%

Indira Gandhi Delhi Technical University for Women

RESEARCH PAPER

AI-Driven Document Intelligence: A Comprehensive Approach to Classification, Q&A Handling, and Fraud Detection

Sonanshi Goel[1], Shambhavi Rai[1], Princy Singhal[1], Mehak Aggarwal[1], Kanika Kanojia[1], *Ritika Kumari[1,2], Poonam Bansal[1]

[1]Department Of Artificial Intelligence and Data Sciences, IGDTUW, Delhi, 110006, India.
[2]USICT, Guru Gobind Singh Indraprastha University, Dwarka, New Delhi, India.

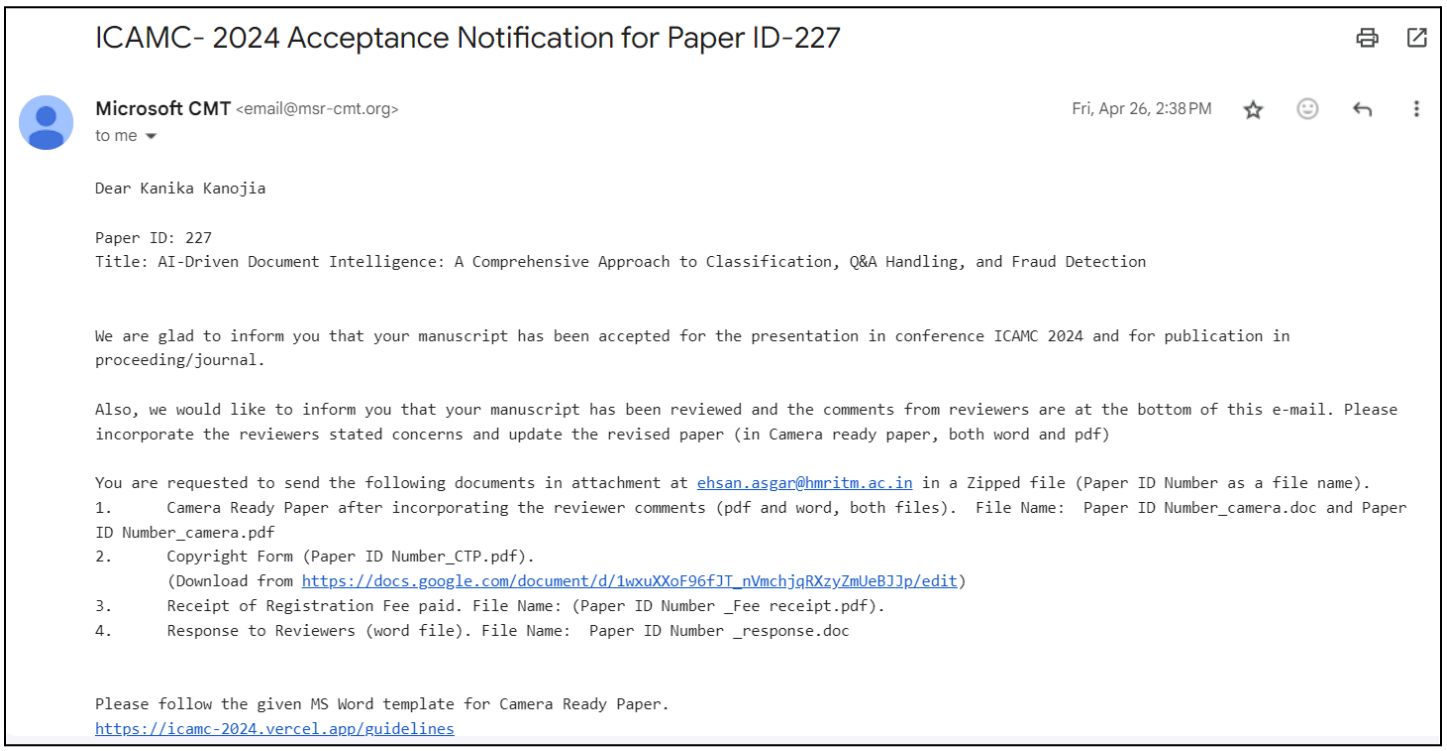
Contributing authors: sonanshig02@gmail.com; shambhavi.rai1604@gmail.com; princysingla11@gmail.com; mehakagg1313@gmail.com; kanikakj07@gmail.com; *ritikakumari@igdtuw.ac.in; poonambansal@igdtuw.ac.in;

Abstract

With the growth in dependency of electronic verification systems used both by government and private organizations, there has been an increase in usage of digitized manuscripts. This demands for user friendly and efficient document interaction tools. Optimisation of the document management process has become AI driven with features like identity document classification, authenticity checking algorithms and interactive question answering. If we effectively identify the document type, we can streamline workflows and enable secure verification processes.The project employs Deep Learning techniques to classify documents like PAN cards, Aadhar cards etc. Also we employ a Similar Document Template Matching Algorithm which can seamlessly extract features, match the template with a real document and thereby detect fraud. Additionally LLMs have been leveraged to enable interactive question answering over documents. Experimental findings in our study showed the effectiveness of our work in improving efficiency, accuracy and usability in document management.

Keywords: CNNs for Document classification, LLMs for Question Answering, Langchain, Similar Document Template Matching Algorithm, VGG16, Fraud Detection,ORB

ACCEPTANCE LETTER



ICAMC- 2024 Acceptance Notification for Paper ID-227

Microsoft CMT <email@msr-cmt.org> Fri, Apr 26, 2:38 PM

to me

Dear Kanika Kanojia

Paper ID: 227
Title: AI-Driven Document Intelligence: A Comprehensive Approach to Classification, Q&A Handling, and Fraud Detection

We are glad to inform you that your manuscript has been accepted for the presentation in conference ICAMC 2024 and for publication in proceeding/journal.

Also, we would like to inform you that your manuscript has been reviewed and the comments from reviewers are at the bottom of this e-mail. Please incorporate the reviewers stated concerns and update the revised paper (in Camera ready paper, both word and pdf)

You are requested to send the following documents in attachment at ehsan.asgar@hmritm.ac.in in a Zipped file (Paper ID Number as a file name).

1. Camera Ready Paper after incorporating the reviewer comments (pdf and word, both files). File Name: Paper ID Number_camera.doc and Paper ID Number_camera.pdf
2. Copyright Form (Paper ID Number_CTP.pdf).
   (Download from https://docs.google.com/document/d/1wxuXXoF96fJT_nVmchjqRXzyZmUeBJJp/edit)
3. Receipt of Registration Fee paid. File Name: (Paper ID Number _Fee receipt.pdf).
4. Response to Reviewers (word file). File Name: Paper ID Number _response.doc

Please follow the given MS Word template for Camera Ready Paper.
https://icamc-2024.vercel.app/guidelines

Indira Gandhi Delhi Technical University for Women

REGISTRATION COPY

HMR INSTITUTE OF TECHNOLOGY AND MANAGEMENT

ICAMC 2024

Registration Page

Name	Kanika kanojia
Email	kanikakj07@gmail.com
MobileNo	9310660967
Name_of_Institute_or_Organisation	Indira Gandhi Delhi Technical University for Women
Place	Delhi
Delegate_Type	Research_Scholar_or_Student
No_of_papers	1
Paper_id	227
Amount	2500
PayMode	UP
Payment Date	28-04-2024 11:41:42
Transaction_number	eshf_662de87d43c651714284669
Status	Success

Powered by:
eShiksa Technologies

AWARD FOR BEST PAPER PRESENTATION

ICAMC

CADD CENTRE

IES MASTER
Institute of Engineers (IES/GATE/PSUs)

Springer

HMR Institute Of Technology and Management

Affiliated to Guru Gobind Singh Indraprastha University, New Delhi, India

ICAMC - 2024

(16-17 May, 2024)

CERTIFICATE OF BEST PAPER PRESENTATION

This is to certify that Prof./Dr./Mr./Ms. PRINCY SINGHAL of IGDTUW awarded **Best Paper Presentation** titled To AI DRIVEN DOC. INTELLIGENCE : A COMPREHENSIVE APPROACH CLASSIFICATION, QA HANDLING & FRAUD DETECTION in the INTERNATIONAL CONFERENCE ON EMERGING APPLICATIONS OF ARTIFICIAL INTELLIGENCE, MACHINE LEARNING AND CYBERSECURITY organized by HMRITM, New Delhi, India.

DR. MD EHSAN ASGAR
Convener

DR. V. C. PANDEY
Director