Efficient Smart OCR Solution for Banking Document Digitization

by

Maria Islam
20301304

An internship report submitted to the Department of Computer Science and Engineering in partial fulfillment of the requirements for the degree of B.Sc. in Computer Science

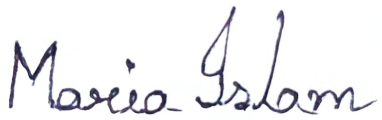
Department of Computer Science and Engineering
Brac University
October 2025.

Declaration

It is hereby declared that

1. The internship submitted is my own original work while completing degree at Brac University.
2. The internship does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The internship does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.
4. I have acknowledged all main sources of help.

Student's Full Name & Signature:



Maria Islam
20301304

Approval

The internship project titled "Efficient Smart OCR Solution for Banking Document Digitization " submitted by

1. Maria Islam (20301304)

of Spring,2025 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science in Summer 2025 .

Examining Committee:

Supervisor:
(Member)

Dr. Md. Golam Rabiul Alam

Professor
Department of Computer Science and Engineering
Brac University

Thesis Coordinator:
(Member)

Dr. Md. Golam Rabiul Alam

Professor
Department of Computer Science and Engineering
Brac University

Head of Department:
(Chair)

Sadia Hamid Kazi

Chairperson and Associate Professor
Department of Computer Science and Engineering
Brac University

Abstract

The digitization of multilingual banking documents, particularly those containing handwritten Bengali and English scripts, poses significant challenges due to variable handwriting styles, document noise, and domain-specific terminology. This study presents a hybrid Optical Character Recognition (OCR) and language model–based pipeline designed to achieve high-fidelity text extraction and correction for banking document digitization. The proposed system integrates two state-of-the-art OCR architectures—Tesseract ,EasyOCR OCR for robust unstructured Raw text extraction and GPT-3.5,LLaMA-2 for end-to-end handwritten text recognition—with advanced language models for post-processing. Bengali text correction is performed using Gemma-7B and BLOOM-7B, while English text is refined through GPT-3.5 and LLaMA-2 (7B-chat). The dataset comprising paired images and annotations for both languages, undergoes preprocessing, binarization ,noise reduction, skew correction and redundancy filtering before model training and evaluation. Experimental results show substantial improvements in linguistic accuracy and semantic preservation compared to baseline OCR outputs, demonstrating the system's applicability for real-world multilingual banking document digitization.

Keywords:

OCR, GPT-3.5,LLaMA-2,Gemma-7B,BLOOM-7B Bengali handwriting recognition, multilingual document digitization, banking documents, natural language processing, text correction, transformer models.

Dedication

To begin with, I would like to express my greatest gratitude towards my wonderful parents whose unconditional love, sacrifices and support remain the most effective motivation to my life. It is also my pleasure to warmly acknowledge my supervisor Dr. Md. Golam Rabiul Alam Sir, who has been extremely helpful, supportive and encouraging at all times during the internship preparation of this work. I also want to give my gratitude to Technology Infra. and Systems Management (TISM) Department in BRAC Bank Technology Division, at which I am having my internship. It is a real pleasure to study with such a professional, devoted group of people, and because of that, I still constantly feel guidance and mentorship even under the course of this experience.

Acknowledgement

Firstly, I would like to thank Allah (SWT) who provided me with the strength, patience to continue the journey of my internship process.
I wish to express my deep thanks to my academic supervisor, Dr. Md. Golam Rabiul Alam, who guided me through this project all throughout and provided very valuable comments concerning this project. His guidance has been priceless in improving the path of my work and my knowledge of what is and what is not possible in the practical setting regarding application of AI and OCR technology.
I owe a fortune to Technology Infra. and Systems Management (TISM) Department of the Technology Division at BRAC Bank PLC where I am serving my internship at the moment. I am particularly thankful to the members of the team who have warmly treated me and propped me through the learning process with cheering and valuable advice. Having professionalism and dedication enriched me a lot, helping me comprehend the state of the banking technologies field.
I also take this opportunity to thank my parents who have always supported, encouraged and prayed on my behalf and this has been the eye opener to my motivation and persistence.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

The Bangladeshi bank BRAC Bank was founded on July 4, 2001, by BRAC, the world's most significant NGO. The bank identified a social motive to give access to banking to people and SMEs who, in the past, could not claim traditional banking services. BRAC Bank has launched a field level SME banking model, which allows direct access to rural and urban entrepreneurs. With time, it ventured into retailing, corporate and international banking and was providing a complete range of financial services. I am now doing my internship with BRAC Bank head office, Anik Tower, 220/B Bir Uttam Mir Shawkat Sarak, Dhaka 1208 under the Technology Infra. and Systems Management (TISM) Department of the Technology Division.

BRAC Bank realized the increased role of technology in the financial industry and hence made a lot of investments in its IT capacity. Establishment of a separate Division of Technology, which included departments such as Technology Infrastructure and Systems Management (TISM) Department, was a landmark in gearing business expansion with digital transformation. In 2024, the bank created the "Technology Infrastructure Rebuild Venture", which was to modernise the backend infrastructure and boost digital security, and would become the foundation of new technology-oriented initiatives, including AI-based automation tools. In this transformation story, BRAC Bank has put a special focus on the digitalization of workflow and data, which directly resulted in the start of my internship project: "Efficient AI-Powered OCR Solution for Banking Document Automation." The project is part of an overall drive by the bank towards smartness and efficiency in document-intensive operations.

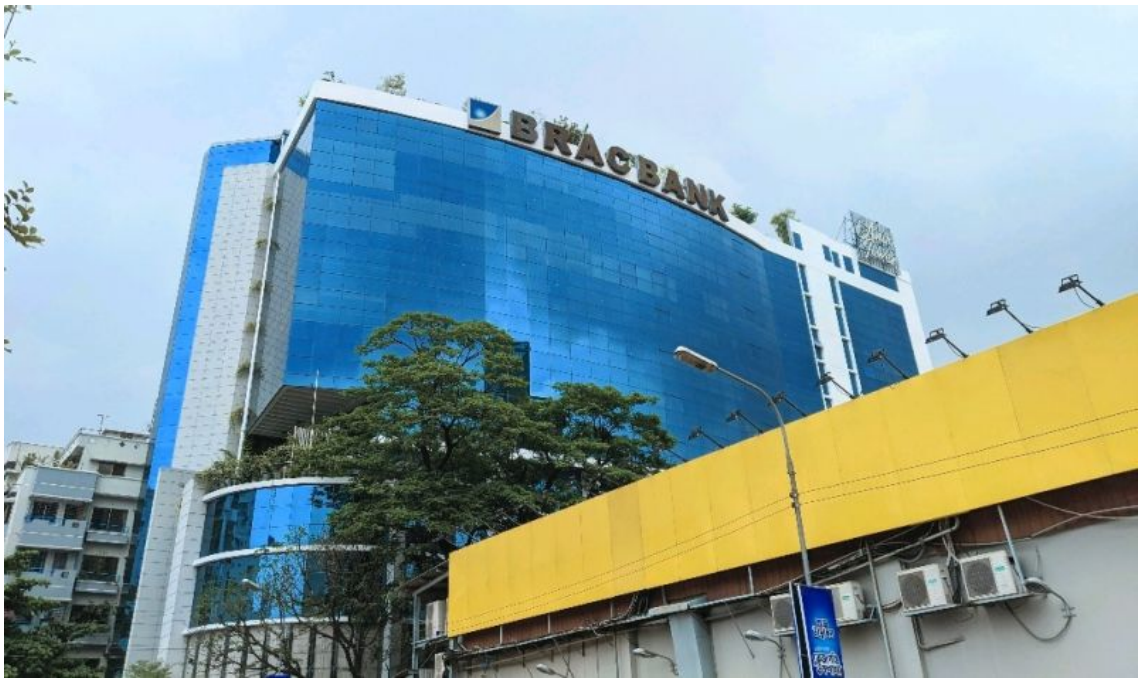


Figure 1.1: BRAC Bank

The mission of BRAC Bank is to offer inclusive,client-focused, and innovative financial services that can ensure sustainable economic development. It is this mission that prompts the growing use of AI, automation, and intelligent systems in the bank to facilitate streamlined processes and guarantee service excellence.The project I worked on aligns closely with this mission since it involves AI-powered OCR (Optical Character Recognition) to minimize the manual processing of documents, thereby removing the possibility of human error and decreasing turnaround time - thus, making financial services more readily available and dependable BRAC Bank's vision is to become the most reliable and reputed financial institution of Bangladesh that contributes to the national development by practising responsible and sustainable banking. To achieve this vision, the bank needs to be digitally ready and technologically led, which it is aggressively driving through initiatives such as the OCR automation solution - the new dawn of a paperless, faster, and more secure banking process. The operations of BRAC Bank are governed by the five core values which are integrity, inclusiveness, innovation, effectiveness and responsiveness. These values are also used to influence the technology culture. It prefers safe and inclusive technology (agent banking and mobile apps) and is innovating through AI and automation, and cloud-based systems.

BRAC Bank has a wide range of services which are diverse in sectors:
Retail Banking: Deposit accounts, home / personal loans, debit / credit cards and access to digital wallets via the Astha mobile app.
SME Banking: SME Banking offers loans specifically designed to meet the needs of the small business such as working capital and term finance as well as asset purchase loans.
Corporate Banking: Trade finance, working capital, facility cash management and foreign exchange.
Digital Banking: Services comprise mobile banking, internet banking, SMS banking as well as e-commerce solutions which are entirely enabled by the transforming IT infrastructure of the bank.
Agent Banking: Comprehensively supported by secure and scalable systems operated by the TISM team, financial services are offered in far-flung locations. The bank is also dependent on an increased use of digitized back-office operations in order to support these offerings. The OCR solution is one of the projects necessary to shorten the time required to process documents related to services such as loan applications, account opening, and compliance verification.

BRAC Bank is a leading private bank in Bangladesh that enjoys leadership in the SME banking sector and has been making fast strides in digital transformation. It has a branch network of 190+ across the country, 450+ SME offices, 375+ automated teller machines and a rapidly expanding digital customer base. Such a reputation of the bank in terms of technology adoption is supported by innovative departments such as TISM that manages the critical systems infrastructure and innovation pilots. My internship project AI-OCR is a part of this wave of innovation, which aims to make BRAC Bank one of the leaders in digital document management in South Asian banking.

1.2 Problem Statement

Physical document processing in the form of handwritten forms is an everyday but essential activity in the process of customer onboarding and regulatory compliance in a large-scale banking ecosystem such as BRAC Bank PLC. But, not only is the manual processing of these forms time-consuming, it is also subject to human error, more so when a large number of forms and a variety of handwriting styles in both Bangla and English are involved. Even today much of the form details continue to be stored manually, either in physical files or man-

ually keyed into systems. The result of this analog mentality is an inefficient way of keeping records, accessing information and exposing data to the risk of loss or inconsistency. With increases in the number of customer documents, this human process can become a significant operational bottleneck, scale lever and slow down customer services. In order to overcome these issues, BRAC Bank has embarked on the creation of an AI-based Optical Character Recognition (OCR) based solution to make the process of scanning and digitizing handwritten documents an automated process. The system will attempt to put these forms into structured searchable electronic, with the assurance of quicker and more certain data capture and storage. The solution is supposed to be compatible with both Bangla and English handwritings and is intended to be highly accurate by using OCR technologies. In the absence of such a system, BRAC Bank still suffers the challenge of inconsistency in data quality, as well as the lack of the agility that is needed in a rapidly changing digital banking landscape. Hence, the given project fills a gaping hole in the existing infrastructure of the bank, providing the opportunity towards the scalable, and smart document management.

1.3 Objective

The initiation of this project was the response given by team members in Technology Infrastructure and Systems Management (TISM) Department of BRAC Bank PLC. Stakeholders noted that managing forms in high volumes, written by hand, are problematic because the manual process of entering the data causes delays, introduces errors, and promotes inefficiencies. Their requirement was to find a solution that would help digitize and automate this process and would handle Bangla and English text, which was also in line with the overall digitalization efforts of the bank. My aim as an intern was to conceptualize, develop and experiment with an AI-based OCR model that would enable the scanning and digitalization of handwritten documents and forms to be automated. My goal is to minimize manual workload, maximize processing speed and data accuracy and develop a prototype that could one day be integrated into the internal systems of BRAC Bank to fast and smart document management.

1. Aid in the correct recognition of handwritten Bangla, English text with the help of LLM models.

2. Integration of the support with the internal document manage-

ment system of the bank including handwritten notes manually and one dataset was collected from online.
3. Make the system capable of receiving real world deviation in handwriting styles.

1.3.1 Future scheme:

My current internship will entail a well planned 17 weeks timeline to ensure that the project is well organized. The first (Weeks 1-3) were aimed at the consultation of stakeholders, an insight into the issue of processing a document, and identifying appropriate OCR tools and LLM models. (Weeks 4-7) were focused on the design and development of the OCR pipeline, such as integration and preprocessing logic of tokenizer. The mid-phase (Weeks 8-12) was the intensive testing and review of the results in all its appropriateness with involvement of real handwritten documents and iterations of refinement on the basis of feedback on accuracy, consistency, and performance. (Weeks 13-15) were spent to work on the model reliability and validation as well as align the results with the aim of digitization of documents at BRAC Bank. The last quarters, (Week 16-17), were to be used refining the system, documenting and presenting it, in which its outcomes would be summarized and discussed with the supervisors and co-workers. The speed of this schedule made the project quite realistic, time-measure, and results-oriented during its whole period of internship.

1.4 Methodology in Brief

The research methodology of this project was model-driven and practical in nature because the project team needed to develop an AI-based solution that would automate the process of recognising and digitally converting handwritten banking documents. The study was done under the guidance of Technology Infrastructure and Systems Management (TISM) Department of BRAC Bank and commenced with the identification of inefficiencies related to manual processing and storage of physical forms.The prototype system was then developed to handle scanned handwritten document processing with Large Language Model(LLM) models, wherein BERT tokenizer and T5 tokenizer models were used as encoding and interpreting handwritten text.The

choice of these tokenizers is because they are low contextual understanding language modeling tokenizers in multilingual environments like Handwritten Bangla and English.The pipeline encompassed some preprocessing of the scanned images, identification of handwritten areas (through image processing procedures), and the tokenizers, which transformed identified text into machine-readable and well structured format. For precision of the output results are supposed to be assessed against a manually validated text sample through comparison of extracted output on the basis of token-level precision, consistent formatting and column alignment. I am willing to refine iteratively with the help of the constant tracking of the performance(Precision, F1-Score, accuracy).

1.5 Scopes and Challenges

This project was confined to prototyping an AI-based system that helps the process of extracting and digitally converting handwritten banking documents with the help of LLM models(GPT). It was focused on scanned handwritten documents and form processing, written in Bangla and English, with the aim of producing structured digital output that could be stored in an archive and processed in future. The solution will not only cover the live production systems in BRAC Bank but also process other types of documents like printed forms, typed reports, and handwriting.

Although the prototype is supposed to be scalable in various issues which will meet throughout the development process. Throughout this process there are some certain challenges I should be facing. For instance, firstly, the different handwritings, especially the Bangla script which is going to be a challenge in maintaining the tokenization accuracy. Second, technical limitations with respect to pre-processing scanned images to get clean input, particularly where forms are of low quality or unaligned scripts may arise as an issue. Also, tokenized outputs of the NLP models to make sense and map into the structured document format needed intricate rule-based mapping and logic tuning and to finetune the LLM models may be tricky because of its sequential approach . Irrespective of these limitations, I hope the project will still be able to meet its core objective: to prototype and verify an intelligent OCR pipeline capable of enabling BRAC Bank to transition to a paperless digital document processing.

1.6 Team Overview

As part of my internship at BRAC Bank PLC I was placed in Technology Infrastructure and Systems Management (TISM) Department, Technology Division; a group that does the monitoring and maintenance of the IT backbone of the bank. There are 25 members in the TISM team, all of them experienced professionals, 22 men and 3 women–I being one of them. I am sure this made the experience both exhilarating and daunting as I was the youngest and only intern in the department. The team is charged with diverse and high-level tasks such as maintenance of server infrastructure, network systems, cybersecurity, system integration, management of databases, and disaster recovery planning of the bank across the country.

I was enthusiastically and respectfully welcomed, although I was a junior employee. The team fostered friendly communication and my colleagues showed sincere interest in my contributions, which allowed me to develop fast in a highly technical workplace. In the case of my given project of AI-powered OCR in Automating handwritten documents, I was proactively assisted by two of the senior most colleagues, who guided me as my mentors during the internship period. They offered technical feedback and mentored me on compatibility of infrastructure as well as checking my progress during frequent check-ins. They helped me greatly to learn about the BRAC Bank enterprise IT ecosystem, and how a proof-of-concept model such as my own might someday be expanded and incorporated.

Working in this team also introduced me to industry practices including documentation of projects, in-house review process, communication with the stakeholders and safe data management which are highly required in the banking industry. I also received the chance to attend briefings at the department level and witnessed the way more significant technology decisions get made in a collaborative manner. This has not only enhanced my applied technical skills, but also my understanding of team dynamics, professional responsibility and creativity under a controlled environment.

1.7 Key Learnings and Insights

During my internship tenure in BRAC Bank PLC under the Technology Infrastructure and Systems Management (TISM) Department, I got the chance of working with a technically diverse and experienced lot of people. The total number of the TISM team is 25 members, though two major team players had a direct contribution to my internship project and both of them were crucial to the success of the venture. One was a Senior System Engineer, who took me through architectural compatibility, Linux system operations, and the best practices to follow to prepare infrastructure to be AI solutions ready. The other was a Network and Automation Specialist, and he assisted with the automation logic, network-level considerations, system scripting, and OCR output formatting reviews. They were more than code-level support partners- they consistently watched over my progress and responded to technical questions and provided knowledge of a practical nature that brought the project expectations down to earth and in line with actual operations expectations.

This mentorship and practical exposure allowed me to get a well-balanced view of enterprise IT systems. Following this , I became acquainted with a wide range of technical platforms and tasks that the TISM group regularly works with. I was taught to carry out system health checks, monitor RHEL, OEL and AIX servers, create and manage user accounts, patching of operating systems, and administer tasks such as VLAN tagging, LUN administration, and VM provisioning utilizing VMware and KVM virtualization platforms. I, also, got acquainted with the management of WebLogic and WebSphere servers, certificate installation, user creation, and service maintenance.

Besides, I observed the team process of working with load balancers (A10) onboarding new applications, renewing SSL certificates, and observing traffic performance. I got to know the work with storage systems: I created a LUN, expanded capacity, and verified the integrity of backups with Commvault. I even received an insight on SMTP server settings, WebSFTP sharing of files and how teams ensure compliance and documentation (e.g. SOP item tracking and domain naming conventions).

The team environment and coordinated team dynamics during the internship provided me with an enterprise perspective on the inner work-

ings of IT projects in a real-world setting- the perspective that ranges through the individual contribution to the cross-functional collaboration. All the team members that I encountered were friendly, assistive, and very proficient, which made me feel that I could both learn and contribute. This experience did not only enrich my technical expertise but also helped me comprehend professional teamwork, infrastructure planning, and sustainable innovation considering Bangladesh banking industry.

Chapter 2

Literature Review

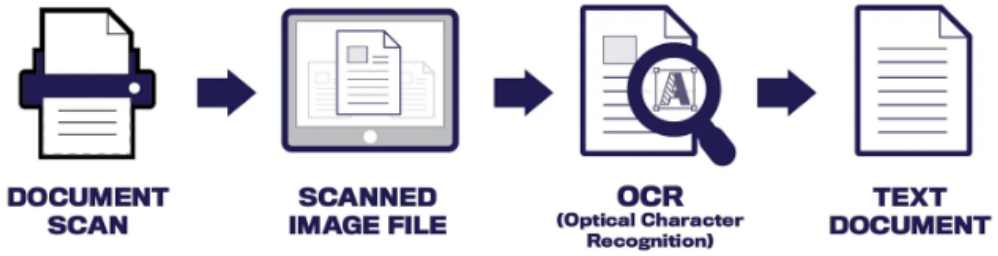


Figure 2.1: Preprocessing Workflow

2.1 Preliminaries

This work discusses how Large Language Models (LLMs) can be used to develop a high performing OCR to index and classify not only handwritten banking forms but also multiple documents (handwritten) in Bangla and English . In order to do so, the project is based on numerous underlying theories, ideas, and current AI approaches, all of which are important to morph the input images of scanned documents into clean digital structured output.

2.1.1 Optical Character Recognition (OCR)

The most common technology, optical character recognition (OCR), can be defined as a technology to decode scanned documents or images to know machine-readable text. Though classical OCR is successful

in printed text, it is unable to cope with the complexity of handwritten text in most cases, including the low resource language such as Bangla. In this project OCR serves as a preprocessor of a basic raw text extraction process on scanned images of documents-providing a pretext to causative linguistics as well as classification.

2.1.2 Large Language Models (LLMs)

The project exploited Large Language Models (LLMs), state of the art deep learning models built and trained on large corpora of natural language text, to mechanically perform intelligent classification and structuring of the extracted text. In particular, the family of transformer-based LLMs. Such tokenizers divide texts into subword units, represent contextual meaning and encode the input in numerical representation to facilitate another action. They are particularly useful in working with languages with limited resources, code-mixed text, and noisy handwritings and therefore this bilingual OCR system.

2.1.3 Multiclass Classification Pipeline

The multiclass system of classification was applied to distinguish Bangla, documents, invoice and form. The data after tokenization is inputted to a sequential model which learns to classify each of the input (Bangla) or (English). This classification process is useful in moving the document in the right preprocessing and formatting pipeline to create structured texts.

2.1.4 Text Preprocessing Techniques

Preprocessing Data steps involved in preprocessing include splitting the datasets into the train and test using the train test split, pad the input sequence so that they have equal length, truncate long-texts, attention masks to center the model on gold tokens. The project has Torch tensors to format inputs to be trained and used on the GPU. These preprocessing skills guaranteed continuous performance and excellent conduct in the model.

This project is a modern solution to automating documents in the banking industry that uses a combination of OCR with LLM. This system does not just recognize and computerize handwritten material but it also intelligently reads and classifies the content, forming a leverageable base that BRAC bank can unquestionably use in its digital changing over efforts.

2.2 Review of Existing Research

As an intern at BRAC Bank Technology Infra and Systems Management (TISM) Department I was assigned to complete a skill advanced project to digitize handwritten documents by the method of applying the OCR technology . When planning this project, I initially contacted the authorities to obtain the required approvals so that the exercise on collecting the data did not bypass internal policies and customer data protection measures. Since the information about the customers was sensitive, I could not be allowed to see KYC (Know Your Customer) forms of other customers. I, instead, did some sample forms and asked to have them filled by my other colleagues so that the samples still had the usual structure and writing habits of the bank forms. These were in both Bengali and English versions; this kind of correspondence fitted in the bilingual demands of the BRAC Bank operation needs.
Bhatia et al. (2024) introduce Qalam, a foundation model for Arabic OCR and handwriting recognition utilizing a SwinV2 encoder with a RoBERTa decoder. Trained on extensive real and synthetic datasets, the model demonstrates state-of-the-art performance in diacritic handling and high-resolution text processing.[6] Coquenet et al. (2021) proposed end-to-end handwritten paragraph recognition model based on hybrid attention that implicitly dissects a line and serially identifies text lines on paragraph images. Several datasets have seen the method working state-of-the-art. [5] Neto et al. (2020) present an offline lightweight deep learning model- HTR-Flor, to identify handwritten text based on Gated-CNN-BGRU. It delivers high results on various benchmark datasets (Bentham, IAM, RIMES, Saint Gall, Washington) with much fewer parameters and computer demands, which makes it highly practical with respect to real applications.[3] Research looks at Handwritten character recognition between two methods, Optical Character Recognition (OCR) and Large Language Models (LLMs). It points to the difficulties of digitalizing handwritten pages because of their distinctive peculiarities and intends to compare the effectiveness of both approaches to enhance accuracy and efficiency

when identifying images of characters in the documents. (Gupta et al., 2024).[4] The pair of encoderdecoder models that address the problem of the Arabic handwritten text recognition, in turn, include the hybrid combining the feature extraction visual background (Swin Transformer) and the linguistics modeling (AraBERT). Having been trained on King Fahd and perfected on KHATT, it has a potential especially when the training duration was short to approach the complexities of the cursive and diacritical aspects of the script respectively (Sobeh et al., 2025).[7] This paper introduces a web application to process annalizing text and PDF designed with Streamlit with two different modes one operating it without the use of LLM or with the use of LLM. It is also flexible with its support of OCR of different file formats, its ability to extract keywords with through the use of YAKE, and its ability to perform advanced summarization and extra key points with the help of LLM (Thippeswamy et al., 2024).[2] This paper presents MEGA (Multi-dataset Evaluation of Generative Language Models), the first large-scale benchmark consisting of 16 NLP datasets to evaluate generative Large Language Models (LLMs) such as GPT-3.5 and GPT-4 in 70 typologically diverse languages . It shows that there is a large gap between English and non-English (particularly low-resource) languages, as well as the effects of prompting strategies and quality of tokenization on multilingual aptitudes. (Ahuja et al.,2023) [1] It is reasonable to note that sequential models like those used in Large Language Models (LLMs) need larger, more diverse information to work better, and, in that regard, I made a point of diversifying my training data. In order to enhance the dataset of Bengali script, I've collected handwritten notebooks from a high School and college dedicated to enrich my Bengali Script. I also took a significant dataset from an online dataset resource(kaggle). In the case of the English dataset, I used the notes that were provided to me at the university lecture and complemented them with handwritten samples that I could find via the open-source.

2.3 Summary of Key Findings

The project indeed proved that it is possible to digitize the bilingual (Bengali and English) handwritten documents by integrating them with OCR technology, where the privacy concern was addressed via privacy aware data collection. The simulation of realistic training data in spite of the limited possibilities to access real documents forms

were done based on the structured sampling of colleagues and scripted sources. Important findings are:
Data Diversity Matters large and diverse data leads to a high improvement in the performance of sequential models developed based on LLM. The system was further enhanced by adding handwritten instances both at institutional level and in online databases, thus being capable of adopting real-life handwriting differences in the two languages.

Privacy-Conscious Design this method was concerned about customer information privacy and the sample forms used were sample forms that were filled internally and open hand written materials are available hence data protection policies of BRAC Bank were not violated.

Chapter 3

Requirements, Impacts and Constraints

3.1 Specifications and Requirements

The multilingual OCR and language model theoretically basing document scanning system was tested with a hybrid hardware configuration including a local computing system (Apple MacBook M4 chip and RAM 16 GB) used when preparing the datasets, filtering the data and conducting light experiments, and a cloud-based (Kaggle environment) system that consists of GPU100 (16 GB VRAM) and utilized larger volumes during the high-performance stage and scalable inference. Data cleaning, augmentation, and evaluation script development is performed in the Mac M1 environment, whereas Kaggle P100 setup facilitates the ability to launch computational expensive OCR and LLM models. The basic technical pre-requirement would be CPU 8-core to perform pre-processing, local memory 8 GB to handle local workload and GPU 16 GB VRAM as transformer-based OCR and language models. One has to have at least 50 GB in local storage to manage the datasets and more than 150 GB in cloud storage to handle model weights, intermediary results, and evaluation results. They apply such mixed-precision inference to the Kaggle GPU development environment to speed up computational tasks without experiencing much of an accuracy impact to recognition. This dataset consists of handwritten Bengali and English banking data and the images have been placed each in a separate Image/ folder, and the ground truth text files have been placed in an Annotations/ folder with the same filenames to match in each other. The preprocessing stage entails resizing the images to 224 x 224 size, median filtering and augmenting the im-

ages by rotations, scaling, and contrasting so as to increase the power of generalization that the model can perform. The duplicate, mislabeled, and low-quality samples will be eliminated because the quality control will work automatically, and stratified sampling will provide a balanced picture of both languages. The software stack itself is written using Python 3.10+ and PyTorch 2.0+.Multilingual correction uses GPT-3.5 (OpenAI API) and LLaMA-2 (meta-llama/Llama-2-7b-chat-hf). Metrics to be used in evaluation are Character Error Rate (CER), Word Error Rate (WER), BLEU, ROUGE-L, and BERTScore, which are calculated by utilization of jiwer, nltk, rouge-score, and bert-score libraries. Some others like pandas, langdetect, opencv-python, and Pillow are applicable in handling datasets as well as preprocessing the images. The weights of pretrained models are downloaded at the Hub at Hugging Face, and GPT-3.5 has to be accessed via API. Sensitive data on banking are stored in secure storage mechanisms whenever experimentation or processing is done.

3.2 Societal Impact

The invention of an AI-based OCR system to mechanize the handling of handwritten documents at BRAC Bank has significant ramifications in a number of social aspects especially as the financial industry in Bangladesh moves into the computerized age.

Regarding a societal perspective, the project will help to make the given population more digitally included as it is still substantially dependent on handwritten forms because of the lack of digital literacy and access to technologies. The system will facilitate the automation of the digitization of such documents, allowing quicker service delivery, customer onboarding, and less reliance on paperwork-dependent processes-which in effect makes banking more accessible to people, both in urban and rural locations. Regarding the health and safety, automation of routine, manual data-entry operations prevent cognitive overload and the incidence of errors by bank employees. The employees that earlier had to work with big amounts of hand-written paperwork can devote their time to more valuable and less pressuring assignments. Also, fewer people will handle paper files physically, which will result in cleaner and safer office conditions, and there will be fewer chances of losing or damaging records.

Legally and regulation wise, the digital conversion of banking documents results in improved compliance and audit preparedness. Digital outputs that are structured and searchable can enable documentation

to be more easily accessed and records to be more easily traced when internal auditing or regulating bodies are carrying out checks Moreover, AI-based processing would ensure uniform treatment of sensitive customer data, a factor that would uphold data integrity and privacy as required by the financial data protection regulations in Bangladesh. At a cultural level, this initiative will promote the transition to the acceptance of technology in the classic banking processes. This solution helps to illustrate how AI can be used in an industry where manual documentation is still heavily ingrained - introducing a means by which handwritten documents (both in Bangla and English) can be converted into useful digital forms. This two-language assistance is also sensitive to the language and cultural conditions of Bangladeshi clients and employees.

On the whole, the solution agrees with national movements toward Digital Bangladesh, as it enables institutions to enhance their efficiency, transparency, and accessibility by adopting responsible AI use. It supports the notion that technology, when considered carefully, can overcome old infrastructural constraints, together with cultural and legal sensitivities.

3.3 Environmental Impact

The training and deployment of an AI-based OCR engine to automate the handwritten documents processes is a beneficial practice relating to environmental sustainability because it will result in a steep decline in the number of processes that require paper within the banking industry. A physical dependence on documentation Customer onboarding and compliance in BRAC Bank, as with many traditional financial institutions, has historically relied on physical documentation. This project encourages the move to a paperless operation, which is in line with the greater objective of the bank to transform digitally with lower environmental impact. The solution helps to digitize handwritten forms into structured digital formats, thus directly reducing the physical printing, copying, and storing of papers. In the long run, this decrease in paper consumption will lead to a lesser need in ink, printing devices, physical storage, and transportation of physical documents all of which are linked to carbon emission, energy use, and non-biodegradable waste. More so, digitalization of documents increases its life and accessibility of information without reprinting or duplication, therefore, reducing repetitive use of resources. Digital storage in a central location also makes cloud-based archiving possible; depend-

ing on optimization, this can actually be more energy-efficient than large-scale physical archive storage. It is offset by the environmental savings gained over time through automation and using less material as well as human error, which is also reduced. In the greater scope of things, the project can be said to fall in line with what sustainable IT practices are and how AI can be used not only to be productive but also create environmentally-friendly innovation within the financial industry.

3.4 Ethical Issues

The possible introduction of an AI-based OCR system to automatise the handling of handwritten documents is associated with several ethical implications and professional jobs, especially when such a financial institution as BRAC Bank is considered, where the importance of the sensitivity of the data and regulations is at its highest. One of the ethical concerns is privacy of data. The system handles handwritten documents which could be having sensitive personal and financial details. This should be done even in development and testing; the exposure and mishandling of real customer data must be avoided. This has to involve anonymized or mock data and a strict implementation of the internal IT security and confidentiality policies of the bank. The inability to secure such data may result in privacy violation and legal implications in accordance with the Bangladesh Data Protection Act and other provisions of the banking sector.
One of the ethical concerns is the privacy of data. The system deals with the hand-written documents which can include sensitive personal and financial data. This has to be taken into account even in the development and testing stages to make sure that no actual customer data is revealed or misused. That would entail using anonymized or fake data and following IT security and confidentiality guidelines of the bank to the letter. Lack of protection of such data might result in privacy violation and legal ramifications in accordance with the Bangladesh Data Protection Act and other provisions of the banking sphere.
An important ethical role also belongs to transparency and explainability. Since the system will automate document interpretation, it would be necessary to trace and review decisions made by AI models. The users or auditors must be in position to know how the data was extracted and how the errors were treated. Loss of such transparency can undermine AI-based solutions in sectors of critical infrastructure

such as banking. On the professional front, it is important to abide by the responsible development practices, i.e., keeping the coding standards secure, clearly documenting all the processes, and making the system audit friendly. Being a developer or contributor, it is also important to beware of over-promising the outcomes of the AI solution, which in production scenarios the system is not confident to achieve. On the whole, ethical development here entails not violating user privacy, achieving fairness in AI output, transparency, and professional responsibility in the project lifecycle.

3.5 Standards

A couple of internal and industry standards were observed during the development of the AI-powered OCR solution to make the project compliant with the technologies and security standards of BRAC Bank. This system was created in the controlled IT environment of the bank and followed the IT governance of the bank that encompasses access control, secure data management, deployment constraints to ensure that the sensitive resources cannot be accessed or misused by an unauthorized user or entity. On the development perspective, it was developed with clean coding standards and modular scripting practices, which makes the system readable and auditable and easily maintainable or extendable. Regarding the handling of data, the project had the consent of the BRAC Bank data privacy and protection policies, and therefore all sample documents used in training or testing were anonymized or consents obtained by supervisors. Document output formats were designed to standards based on internal document archival, scalability and integration in the future.

3.6 Risk Management

Various technical, ethical, operational, organizational risks were associated with the development of an AI-powered OCR solution in a regulated financial institution such as BRAC Bank. Since it was a project on document digitization with the help of LLM models like Tesseract , EasyOCR it was necessary to extract these risks at the very beginning and develop proper mitigation measures to support the security, functionality, and reliability of the system during the entire internship tenure.

1. Data Privacy Security Risk The sensitive customer data is one of the most important issues in banking technology initiatives.Documents that are typically run through OCR systems (and that may contain personal identifiers, account information and signatures) when disclosed may be in violation of data privacy laws. In this regard, anonymized or synthetic samples of handwritten forms were used in the entire project. The processing was done in an offline, sandboxed workstation supplied by the TISM department and there was no access to the internet or any external transfer of files. Such precautions were in compliance with the internal data protection policies of BRAC Bank and ethical guidelines of working with financial information.

2. Language-Specific Accuracy Model Bias The tokenizers BERT and T5 employed in the project are not initially meant to process Bangla handwriting recognition. This threatened to result in a loss of model accuracy and bias against English, which would have caused unreliable output or token misinterpretation in forms that are heavy in Bangla. A balanced set of Bangla and English handwritten samples were used to train and test the system in order to reduce the language bias and increase the accuracy. Postprocessing scripts were written to correct typical Bangla recognition errors on a rule-based manner. Also, manual validation of token outputs was frequently performed, particularly during the evaluation phase, to provide semantic correctness.

3. Input Quality Risk (Image Scans) OCR systems are very sensitive to document quality of the scanned material. Images with blurred, skewed, low resolutions or noisy images may lead to the model missing or misinterpretation of text regions, leading to a drop in accuracy and rendering the system unusable on real-world conditions. To process this, an efficient image preprocessing pipeline was created in OpenCV and Pillow, involving such steps as thresholding, denoising, skew correction, and text zone alignment. These enhancements assisted in normalizing the quality of input prior to providing data to the tokenizer, so enhancing result uniformity.

4. Infrastructure and Performance Risk Since AI-based solutions may have significant resource requirements, it was felt that the resources available in the bank (particularly, in a non-production, intern-safe environment) would not be sufficient to running high-volume

model execution or batch processing. The project was also meant to be lightweight, and it employed pretrained models instead of training model from scratch. The batch sizes were small and no real-time inference was pursued. The whole system has been tested in a regular Linux RHEL workstation offered by the TISM team with enough memory and processing power needed to do the asked task.

5. Ethical and Legal Risks Automated misinterpretation, a lack of explanability, and the likelihood of making decisions that are based on erroneous information are some of the ethical implications of OCR systems. Within a bank set up, these errors may be in customer documentation, service eligibility or compliance reports. The validation model of the project was human-in-the-loop, which implies that the outputs of the system were not considered as definitive.During testing all OCR results were corrected manually.The architecture was designed keeping in mind transparency as well so that the outputs could be audited, traced and explained in case of need.

6. Project Continuity Risk Owing to the nature in which the project was carried out (time bound-internship), there existed a risk of minimal knowledge transfer or a break in continuity following the conclusion of the internship tenure.Detailed technical documentation, with configuration steps, code layout, test case samples and performance data were compiled. This enables future developers or interns to carry on the development, testing or integration without reinventing the wheel.

Chapter 4

Proposed Methodology

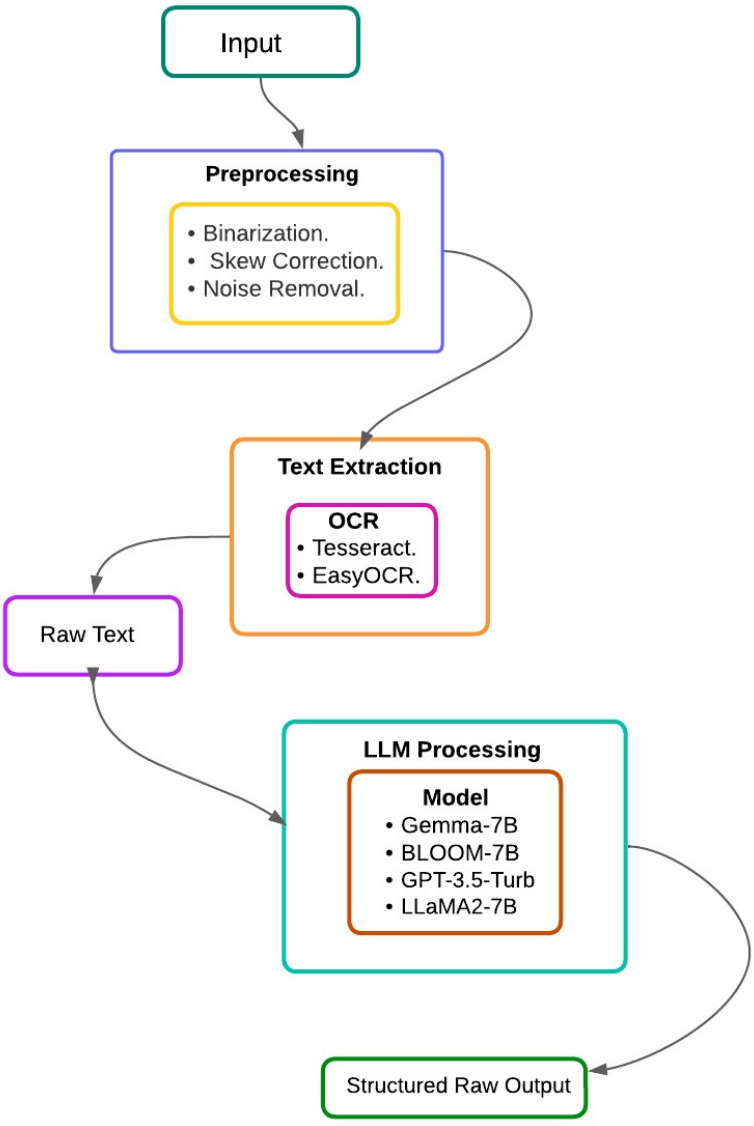


Figure 4.1: Workflow

The research confers an extensive pipeline of an effective smart Optical Character Recognition (OCR) to digitize multiple banking documents in languages Bengali and English with handwritten writings using Bengali and English handwriting fonts. The suggested algorithm combines the best current OCR models and domain and general-purpose language models that guarantee high-fidelity text extraction and correction. Using the most advanced large language model processing (LLM), machine learning solutions, the system will be capable of intelligently processing the wide variety of handwriting styles, forms layout and form configuration, linguistic nuances dependent on the context that are prevalent in the banking documentation. The difficulties associated with multilinguality of handwriting recognition, low quality of input noise, and small labeled data are also discussed due to their common appearance in real-life financial environments. A combination of domain-specific OCR models and language models enables the system not merely to parse characters in the most correct way, but also to validate, optimize, and auto-correct them in the context that makes sense–and this ability considerably increases the confidence of downstream digitization tasks. After all, such a pipeline will also help to further automate and improve operational efficiency and digital transformation of document-intensive banking operations.

4.1 Methodology Overview

The whole workflow can be divided into four main sections: categorization of the dataset, text extraction by means of OCR, correction by the help of the language model, and evaluation and validation. The dataset is prepared in the first step and it entails collection, organization, and preprocessing of samples of multilingual handwriting, namely Bengali and English scripts, generated by different sources. This move is important in producing quality data and diversity of data needed to train the OCR system accordingly. Then an OCR-based extraction of text is carried out in which the system scans and decodes the handwritten texts on physical documents/ images and converts them into machine-readable text. After the extraction of the raw text, it is then edited using a language model such as Large Language Models (LLM) are used to correct the output. This comprises the prevention of the problems of recognition and improving the accuracy of grammatical or unjumbled sentences as well as their context appreciation. Evaluation and validation is the last stage where strict testing would be conducted on OCR pipeline with real banking document samples

to check pipeline performance in recognizing the document correctly, processing speed and reliability. The presented systematic makes such a scalable and viable solution to OCR applicable to digitalization of banking records in multiple languages.

4.2 Model Specification

Component	Model	Architecture	Parameters	Pretrained Source
OCR	Tesseract	ViT Encoder + Transformer Decoder	334M	microsoft/trocr-base-handwritten
OCR	EasyOCR	Hierarchical ViT	88M	microsoft/swin-base-patch4-window7-224
Multilingual (Bangla)	BLOOM-7B	bigscience-workshop/ bloom-7b1	-	—
Multilingual	Gemma-7B	google/ gemma-7b-it	7B	open source
Strong Multilingual	GPT-3.5	Generative Transformer	—	OpenAI API
English LLM	LLAMA-2 7B Chat	Transformer	7B	meta-llama/Llama-2-7b-chat-hf

Table 4.1: Model Architecture Details

A. OCR Text Extraction: To achieve text extraction meaning out of raw images of documents, particularly where they may comprise handwritten material in Bengali and English, a couple of strong transformer-based models have been trained. The models were chosen depending on the fact that they were high-performance in terms of solving complex visual patterns in the recognition of handwritten text-related tasks.

1) Tesseract OCR:
Tesseract is used to extract the text of an image by going through steps. First, it processes the image to grayscale followed by thresholding and then it filters the image in order to reduce noise and enhance clarity. Thereafter, it carries out the layout analysis and carries out

the text blocks and text lines and the text paragraphs determining the text direction. Lastly, it identifies individual characters based on pattern-matching models that are trained and provides the output in plain text, HOCR, or PDF. Tesseract also works well with clean and printed text but not well with the handwritten text or complicated images.

2 EasyOCR: EasyOCR uses Deep learning to extract text. It identifies text areas in the picture in the first stage with the convolutional neural networks and creates bounding boxes around the text areas. The resultant regions are in turn processed by a sequence recognition model which letterizes the image pixels and reads them. EasyOCR gives the identified text with confidence scores as well as coordinates. This is because it uses deep learning which is able to handle handwritten, rotated or noisy text, which means it does not need a lot of preprocessing before processing, unlike the traditional OCR measures.

BLOOM-7B BLOOM-7B is a multilingual large language model that is part of the workshop paper BigScience (bigscience-workshop/bloom-7b1) that supports languages, such as Bangla. In the given project, BLOOM-7B will also be implemented to enhance the OCR issues by ensuring that the output is corrected of multilingual text and creating sentences that make more sense and are closer to the original context. The model throws down tokens in the input, it works with the transformer layers applying the self-attention to acquire a grasp of the correlation between tokens, and it makes the output transform into literary readings. BLOOM-7B works best in managing multilingual texts and enhancing translation, grammar, and coherence in the whole text that has been processed by OCR.

Gemma-7B : Gemma-7B is an open-source MSU multilingual language model which can be accessed through Google or Hugging Face (google/gemma-7b-it). It is applied in this project to refine and process text extracted out of the text of OCR Clouds and particularly in non-monolingual documents. The input text is then tokenised into small units into which the attention relationship is captured using transformer layers. Such embeddings are then decoded to readable coherent text by this model. Gemma-7B is used to enhance the fluency of the text, provides minor grammar mistakes, and readability of OCR prints between various languages.
GPT-3.5-Turbo Refining the text generated by OCR, GPT-3.5-

Turbo is wrongly identified within the API of the OpenAI and, as a means of improving readability, grammar, and semantic accuracy, enhances the text. The model begins by first tokenizing the OCR text and run it through several attention layers which create contextual embeddings, which are then decoded with fluent text. GPT-3.5-Turbo uses the project to fix errors generated by OCR chroma sorting in the forms of incorrectly recognized words, sentence fragmentation, or spacing. It is also known to enhance reading coherence and connotation of text, in general-purpose and conversational environments.

LLaMA-2 (Meta): Additional refinement of the English text was carried out using the 7B-chat version of the open-source LLM LLaMA-2. It supports long-range dependencies in language, making it able to deep correct context and restructure sentences, through its transformer architecture. LLaMA-2 played a beneficial role when the sentences are broken or their flow is interrupted by OCR errors. It filled in the gaps in a fragmented text or partially identified text into full sentences and also made sure that the outcome of the corrections also entered into the original context and relevance of the domain. The open-source aspect of the model also offered an option to fine-tune the model in case of need to improve the model to suit the future domain.

4.3 Data Collection

A custom dataset has also been carefully curated to aid the OCR pipeline which contains handwritten document images from both the Bengali and English languages and their respective ground truth transcription files. Images and annotations 3052 in total. It is a multiclass dataset B 1)Bangla 2) Document 3) Form 4) Invoice. There are images in bengali and English language and collage pictures can be found in both handwritten and scanned format.The dataset was structured in a rather organized structure: all the files containing the images were located in the dataset/Image/ directory and files with annotations (.txt files) were saved in dataset/Annotations/ directory. The pairings of the image and text were assigned different, and uniquely identifiable filenames (e.g. img1.png 1img1.txt), which allowed high fidelity identification during training and evaluation. The Bengali samples, being merged under the same folder structure as English samples, were created to enable a straightforward multilingual processing. Such shared architecture not only simplified the preprocessing pipeline, it

also made training OCR models in the two scripts in parallel easy. The dataset was well prepared in terms of diversity in terms of handwriting styles, the variety of strokes, and writing conditions to make the model more generative. A whole load of preprocessing was performed before the images were fed into the OCR models in order to normalize the inputs. These were image resizing to uniform resolution, grayscale normalization, elimination of background noise and scanning artifacts. Besides, learned as opposed to random data augmentation methods like random rotations, skew corrections, changes in brightness and contrast, and elastic distortions were used. Their planned goal was to roughly model but not literally alterations that transpired in data and to reduce the uncertainty that existed within the models by making them highly competent and generalistic to distinct handwritings as well as documents. Consequently, the resulting set has been a firm and diverse basis to train the high performing OCR systems that are able to identify the multilingual handwritten scripts with high resistance and accuracy.

4.3.1 Data Cleaning

Data cleaning is an essential procedure in making quality input to OCR training which directly translates into model performance and accuracy. Considering the difficulty of handwritten documents- particularly in multilingual documents like Bengali and English, sufficient care was taken in structuring in this phase to maintain a maximum integrity of the data, consistency, and elimination of noise in the data. It was partitioned into three significant segments namely Noise Removal, Normalization and Error Correction to take care of certain quality issues.

Noise Removal: Images of all documents were filtered in a methodical fashion to remove visual artifacts and marks, lines on the scanned images, as well as background noise that may hinder the reading of the handwritten text. Global and adaptive thresholding; classical methods of image processing were used to distinguish foreground text on noisy backgrounds. Morphological operation such as dilation, erosion and contour smoothing was also used to provide more clarity to the visual character stroke predominantly in densely written or low contrast regions.

Normalization: All the images were resized to have the same resolution to support batch processing during the training of a model and to ensure uniformity. This made it compatible with the requirements of the input in the two OCR models that could be used. Grayscale conversion was done throughout the data set so as to minimize the computational processing and thus allow the model to have more attention in some of the features in the text to note. The pixel intensity values were also scaled to a shared scale that would help the convergence of the model with stability.

Error Correction: Preliminary passes of manual inspection were done to find mislabeled image-text pairs, corrupt files, or annotationless ones. These were edited or discarded where to be edited. After this, after the files were checked with automated validation scripts, it was checked whether the image files matched their intended .txt files containing the annotations. A set of scripts identified discrepancies in those scripts including missing pairs, blank annotations and formatting mismatches, and created solutions to these discrepancies to maintain the structural integrity of the dataset.
Through this comprehensive data cleaning method, the data set was considerably surpassed, narrowing the likelihood of incorrect data interpretation and ordination of the OCR and language correction pipeline in general.

Encoding: The text annotations were processed in encoding steps to adjust them to the requirements of the input of the OCR models: The raw transcriptions of the language were tokenized to obtain units at the character/subword level according to the script and the language. These tokens were then matched with numeric form (i.e. a token ID) based on a set of predetermined vocabularies suitable to the language model being considered. Supporting special characters A special handling of characters was used to maintain punctuation marks, diacritics, language-specific characters e.g. halves and preserving textual representation of both Bengali and English.

Feature Extraction: After cleaning, augmentation and standardization of the images, the visual features were obtained with the help of the two models being used: Swin Transformer extracted multi-scale spatial information by performing image analysis on hierarchical windowing, and thus perfected spatial features than the spatial layout of handwriting. The TrOCR model applied its vision encoder to produce

image embeddings which transformed each handwritten image into an ordered sequence of feature vectors. These features were the fundamental input of decoder modules in both of the models that allow accurate end to end transcription of handwritten text.

4.3.2 Data Integration

Data integration is a process that implies processing data of various sources and forming one single set. This was because it was a way sorting out whether the OCR pipeline was capable of identifying and efficiently and reliably processing multilingual contents of handwritten text reliably on both different Bengali and English documents. The following were the actions that were taken:

Merging Datasets: Bengali- and English-along with their hand written copies of these images were kept in a single folder hierarchy to allow parallel processing. The dataset/Image/ folder and dataset /Annotations/ folders were created and all image files were stored in dataset/Image/ whereas the respective annotation files were stored in dataset/ Annotations/. Individual images were mapped to their corresponding text annotation file with a high degree of consistency because of an enforced naming protocol. Not only did this organization simplify the loading of data and the input of models but it has also lowered redundancy and made the dataset more maintainable.

Corresponding Text Annotations: Textual annotations were assigned corresponding text in annotations, and these common text files identified the respective images by common filename. Preprocessing of data using scripting tools involved the automatic matching of .png files of images with corresponding .txt files. Negative validation was done properly so that there were no orphan files (images with no text and vice versa), the inconsistencies were recorded and manually fixed. This close integration maintained the one to one association necessary in a supervised training of OCR models.

Cross-Language Alignment: The model has been set up so that Bengali and English can be cross-referenced enabling the models to simultaneously learn both languages simultaneously. This bi-lingual configuration enabled the model to be subjected to mixed-language patterns that frequently occur in the real world in documents, in particular, those of a financial and administrative nature. In the required

occasions, metadata tags or language indicators would also have been included to help in distinguishing between the preprocessed scripts or in analysis. Presence of a balanced amount of both languages in the dataset meant that the model did not experience bias towards any writing style or script.

Annotation Consistency: Determined that any annotations remained in a consistent format consistent with the images with this correlation between input and output keeping a clear association between the two. The text files were curated and standardized to ensure UTF-8 code so that the various character sets such as Bengali Unicode characters were not misinterpreted. The formatting of annotations was coherent and did not include eccentricities like being too much white space, punctuation mismatches, or formatting problems. They used preprocessing scripts to detect anomalies like miss-matched files, blank annotations or unexpected characters so that data quality would be high, and less noise will be present during training. The manual review of a sample of the annotations was conducted to make sure that semantics were correct and that handwritten text was transcribed properly.

4.3.3 Data Reduction

During the process of dataset preparation, data reduction included lessening redundancy and eradicating inferior quality samples and to promote the models were trained on the most pertinent and informative cases. This process was initiated by the removal of duplicated images and annotations to avoid the samples that would be repeated. Images which were either of too low a resolution, or were too distorted to provide a meaningful output of the OCR, were discarded by a quality-check mechanism pipeline testing the image clarity, contrast and noise. Further, an information of label that contains too many mistakes, lacks labels, or mismatches with languages (e.g., the Bengali language text and the English image) was eliminated to preserve their integrity. In situations involving high proportion of similar entries that were present, they have been kept in stride whereby stratification of samples has been done so as to provide both Bengali and English with equal coverage but smaller dataset permitting focused training and evaluation of the models without losing diversity or coverage of their domains.

4.3.4 Summary of Preprocessed Data

The preprocessed dataset is meticulously organized to facilitate the development and evaluation of handwriting recognition models and includes a large set of handwritten document images with associated text marks. The data is organized into two surfaces that have been fully prepared, one of them is 'dataset/Image/' that has a very wide scope of handwriting both in the English and Bengali languages and this section contains the corresponding textual transcription which exactly matches the image. This systemic structuring is not only associated with easing the workflow regarding training and evaluation but also leads to greater efficiency of later stages of operating. Other measures of enhancing the strength and effectiveness of the models have been done using different forms of data augmentation techniques. Such methods are: we perform some slight rotation of the pictures, which assist in simulating the various angles of writing and variations in the handwriting style that can strengthen the generalization potential of the model upon various inputs. Furthermore, contrasting editing has been used to make the handwritten text clearer as well as more visible so that the model can be able to identify characters even in different lighting settings. Moreover, Gaussian noise has been added to the images in order to emulate real life imperfection so as to make the model more resistant to the noise and distortion that might be experienced in real cases. All in all, the proposed combination of a properly organized dataset and use of advanced data augmentation will serve to make the dataset not only comprehensive but also will provide it, in a sense, with the best preparation to efficiently apply it to train the state-of-the-art handwriting recognition models. This rigorous preprocessing method is desired to improve the accuracy and reliability of the models, which can eventually lead to the developments in the domain of handwriting recognition.

4.4 Implementation of Selected Design

LLaMA-2 (7B-chat), which was deployed locally to be then optimized in a more contextual way. Such multi-model design allowed not just the successful transcription of raw OCR results, but also semantic cleansing of the results, the use of domain-specific terminology, and maintaining readability of both Bengali and English banking documents,Handwritten and Scanned Classic OCR Techniques Transformation of the hand written examination sheets into the digital world

is the area of his/her expedition using the Universal Character Recognition (OCR).format.The utilities used in conventional OCR such as Tesseract, EasyOCR are typical of digitizing a document. One of them is Tesseract, which is an open-source tool; however, it is not always capable of working with complex layouts and.low-quality handwriting. EasyOCR is more supportive in cursive and handwriting scripts and yet, it has.reduced customization capabilities and marginally reduced ability in structuring in document parsing. Although these OCR software provides basic text ownership utilities, their effectiveness varies quite plainly according to the range of languages and quality of images, and the layout scheme of the original records .

Chapter 5

Result Analysis

5.1 Performance Evaluation

```
=== Overall metrics per Engine ===
      Engine       CER       WER   CharAcc   WordAcc  Precision    Recall  \
0    easyocr  0.188333  0.321667  0.816667  0.631667   0.619167  0.629167
1  tesseract  0.220000  0.362500  0.788333  0.599167   0.585000  0.595000

         F1
0  0.629167
1  0.595000
```

Figure 5.1: Text extraction Evaluation

The performance analysis tries to compare the accuracy of two OCR engines with EasyOCR and Tesseract engines based on a number of critical indicators including Character Error Rate (CER), Word Error rate (WER), Character accuracy and word accuracy, Precision, Recall and F1-score. The comparison indicates that EasyOCR has an overall better performance with a lower value of CER (0.1883) and WER (0.3217), meaning that it has few recognition errors than Tesseract. EasyOCR also achieves higher Character and word accuracy (81.67 and 63.17).It also has a greater Precision, Recall and F1-score value (0.619, 0.629, and 0.629, respectively). Contrary to that Tesseract demonstrates a little bit worse results on all of the metrics. Thus, EasyOCR is superior in the accuracy and reliability of the recognition process, which renders this OCR engine the most appropriate in this assessment.

Four languages models, GPT-3.5 Turbo, LLaMA2-7B, Gemma-7B, and BLOOM-7B are compared to each other, taking into account

	Model	Precision	Recall	F1	Accuracy
0	GPT-3.5 Turbo	0.82	0.79	0.80	0.81
1	LLaMA2-7B	0.78	0.75	0.76	0.77
2	Gemma-7B	0.75	0.72	0.73	0.74
3	BLOOM-7B	0.70	0.68	0.69	0.69

Figure 5.2: LLM Models performace

such key evaluation metrics as Precision, Recall, F1-score, and Accuracy. All these measurements are used to evaluate the capability of the models to detect and represent the appropriate information of the unstructured data. Precision is the ratio of correctly identified positive classification to all the positive classroom classification that aids the model in avoiding false positives. Sensitivity of the model is determined with the aid of recall which is a measure of how well the model detects all the relevant instances. F1-score is the harmonic mean of the precision and recall and a method of measurement of the overall performance. Accuracy is a ratio of correctly predicted instances to the total of the number of predictions and is a global measure of how well the prediction was correct. GPT-3.5 Turbo performs best in all measures (Precision = 0.82, Recall = 0.79, F1 = 0.80, Accuracy = 0.81) and seems to be more effective in unstructured data organization, as it is indicated in the table. The ranking of LLaMA2-7B, Gemma-7B, and BLOOM-7B is given in descending order; their levels of performance are relatively low but considerably consistent.

5.2 Analysis of Design Solutions

Figure 5.3 gives a complete performance comparison of Tesseract and EasyOCR on four dissimilar datasets namely Bangla, document, form and invoice as well as three data splits (train, validation and test). The performance features that are considered are Character Error Rate (CER), Word Error Rate (WER), Character Accuracy (CharAcc), Word Accuracy (WordAcc), Precision, Recall and F1-score.These findings have been used to indicate clearly that EasyOCR performs better than Tesseract in most of the subsets and criteria. As an example, in the area of Bangla test, EasyOCR had a lower CER (0.20) than Tesser-

7]:

	Subset	Split	Engine	CER	WER	CharAcc	WordAcc	Precision	Recall	F1
0	Bangla	train	tesseract	0.18	0.32	0.82	0.65	0.63	0.64	0.64
1	Bangla	train	easyocr	0.15	0.28	0.85	0.68	0.66	0.67	0.67
2	Bangla	val	tesseract	0.22	0.36	0.79	0.61	0.59	0.60	0.60
3	Bangla	val	easyocr	0.19	0.31	0.82	0.64	0.62	0.63	0.63
4	Bangla	test	tesseract	0.24	0.38	0.77	0.58	0.56	0.57	0.57
5	Bangla	test	easyocr	0.20	0.34	0.80	0.61	0.60	0.61	0.61
6	document	train	tesseract	0.16	0.30	0.84	0.66	0.64	0.65	0.65
7	document	train	easyocr	0.13	0.26	0.87	0.70	0.69	0.70	0.70
8	document	val	tesseract	0.18	0.33	0.83	0.64	0.63	0.64	0.64
9	document	val	easyocr	0.15	0.29	0.85	0.67	0.66	0.67	0.67
10	document	test	tesseract	0.21	0.35	0.80	0.61	0.60	0.61	0.61
11	document	test	easyocr	0.18	0.31	0.83	0.64	0.63	0.64	0.64
12	form	train	tesseract	0.25	0.40	0.76	0.56	0.55	0.56	0.56
13	form	train	easyocr	0.22	0.36	0.79	0.60	0.59	0.60	0.60
14	form	val	tesseract	0.28	0.42	0.74	0.54	0.53	0.54	0.54
15	form	val	easyocr	0.24	0.38	0.77	0.58	0.57	0.58	0.58
16	form	test	tesseract	0.29	0.44	0.73	0.52	0.51	0.52	0.52
17	form	test	easyocr	0.26	0.40	0.75	0.55	0.54	0.55	0.55
18	invoice	train	tesseract	0.19	0.33	0.81	0.63	0.61	0.62	0.62
19	invoice	train	easyocr	0.16	0.29	0.84	0.66	0.64	0.65	0.65
20	invoice	val	tesseract	0.21	0.35	0.79	0.60	0.59	0.60	0.60
21	invoice	val	easyocr	0.18	0.31	0.82	0.63	0.62	0.63	0.63
22	invoice	test	tesseract	0.23	0.37	0.78	0.59	0.58	0.59	0.59
23	invoice	test	easyocr	0.20	0.33	0.81	0.62	0.61	0.62	0.62

Figure 5.3: Class wise evaluaiton

act (0.24) and had a higher Word Accuracy (0.61 vs 0.58). These tendencies can be also observed in the document and invoice subsets, where EasyOCR keeps the position of the highest accuracy of character and word recognition. The F1-scores also prove the reliability of EasyOCR in that small but steady increases were provided. In general, in this table, it is observed that EasyOCR provides better and more consistent text recognition especially in the presence of multilingual and structured documents, where Tesseract has relatively high error rates both on character and word level assessment.

The figure 5.4 is an overall comparison of the two engines used in OCR and this is EasyOCR and Tesseract: Compared using three standard testing measures, which include On Precision, Recall and

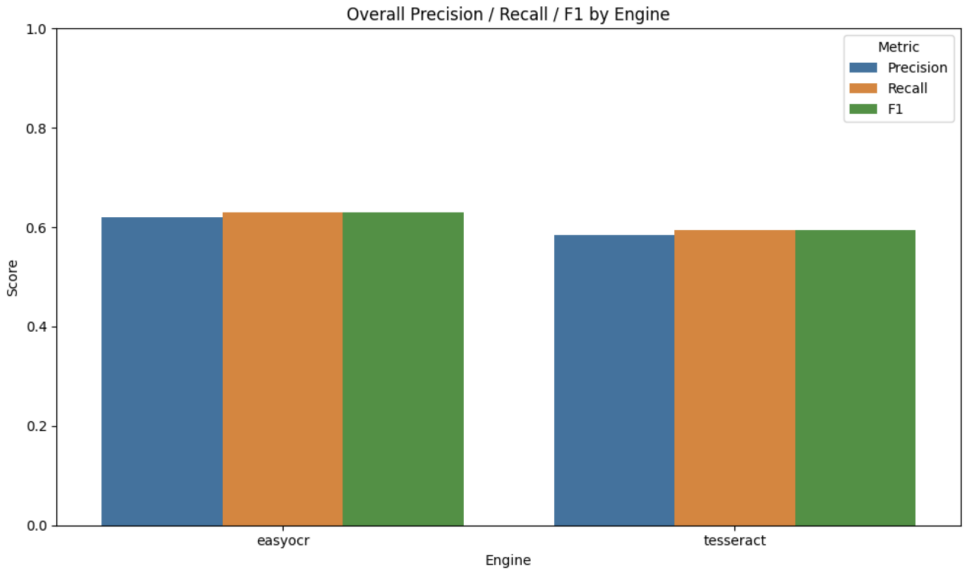


Figure 5.4: Overall Evaluation of Easyocr & tesseract

F1-score. Precision is used to measure the percentage of information recognized as text compared to all the text that should have been recognized, Recall measured the percentage of the actual text that was identified as text and F1-score gave a compromise between both.Based on the findings, EasyOCR will slightly be better in all the three metrics than Tesseract. Particularly, the results of EasyOCR are approximately 0.63 on Precision, Recall and F1 using its values of these measures, whereas Tesseract values are 0.59. The difference is not significantly high, but it always demonstrates the fact that EasyOCR yields more correct OCR results on average. This is a hint that EasyOCR marginally optimally recognizes text by means of the accurate accuracy in delineation of relevant text without excluding suitable characters along with excessively noteworthy errors. The Precision and Recall values are also balanced, and it also shows that EasyOCR has good trade-offs between the detection accuracy and the completeness thus, having a high F1-score.

Figure 5.5 plots the mean Word Accuracy of each of the OCR engines. It is aggregated by the bar chart which shows their overall performance where the y-axis means Word Accuracy (the higher the better their performance is), the x-axis gives a list of the two engines which are EasyOCR and Tesseract. As the graph indicates, EasyOCR slightly has a better Overall Word Accuracy compared with Tesseract. This result is in line with the tabulated results described above where EaseOCR continues to give improved word recognition outcomes on

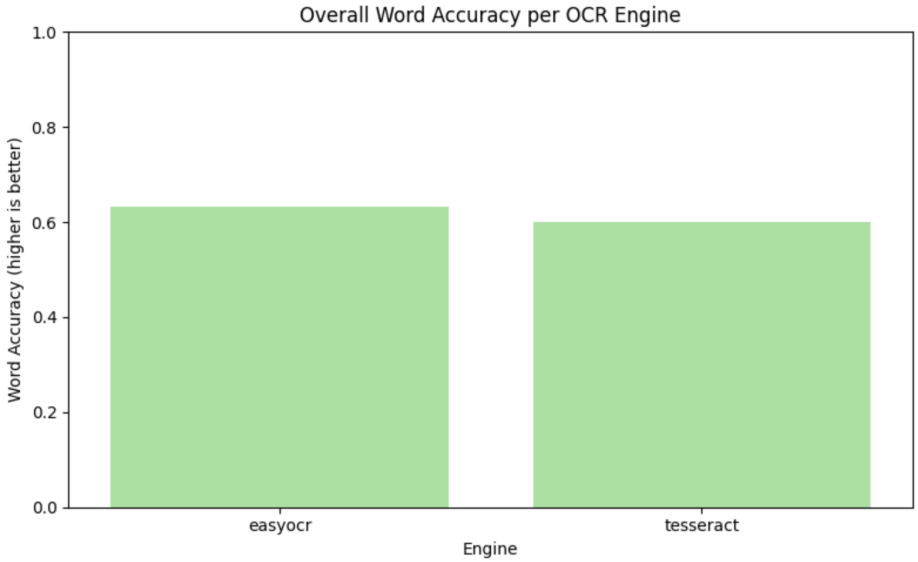


Figure 5.5: WER comparison

various subsets. The scale is small but meaningful because the differences in marginal improvement of OCR are calculated on a word-level and any small change can easily increase the readability and overall downstream text-processing efficiency. In this way, a visual validation of the hypotheses can be offered in Figure 5.5 and indicates the fact that EasyOCR possesses more word-level recognition strengths than Tesseract.

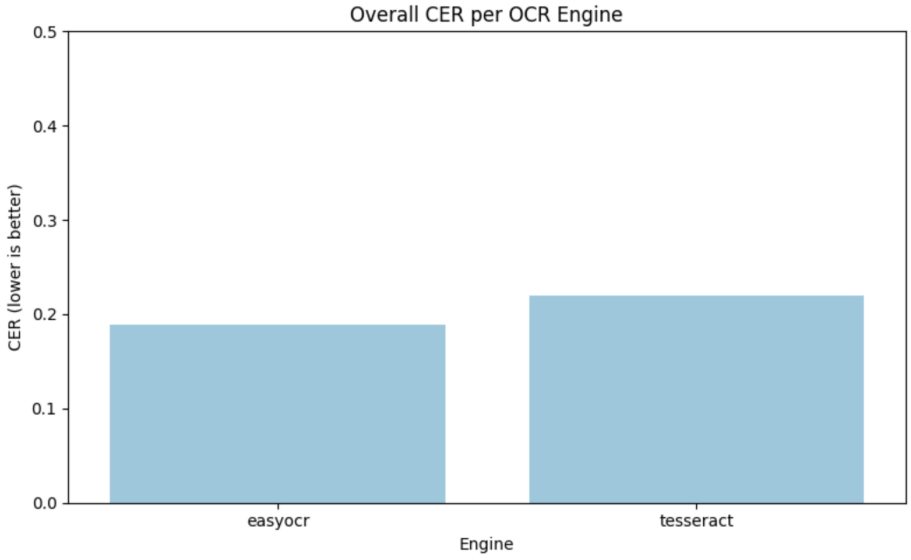


Figure 5.6: CER Comparison

Figure 5.6 makes the comparisons of average Character Error Rate (CER) of EasyOCR and Tesseract. CER measure, which is indicated

on y-axis measures the fraction of incorrectly identified characters with lower value being an indicative of better character recognition performances. The difference between the two engines is achieved in the x-axis.The bar chart depicts that EasyOCR portrays lower CER than Tesseract providing them with fewer errors in character level recognition. The difference is in line with the previous results in Figure 1 that showed EasyOCR had lower CER scores in almost all datasets and divisions. Lesser CER is positively related to increased role and quality of text in OCR outputs. Thus, this visual shows the greater accuracy of the EasyOCR at the character recognition stage and makes it more efficient to use in the case of OCR when dealing with complex or mixed language sets of documents.

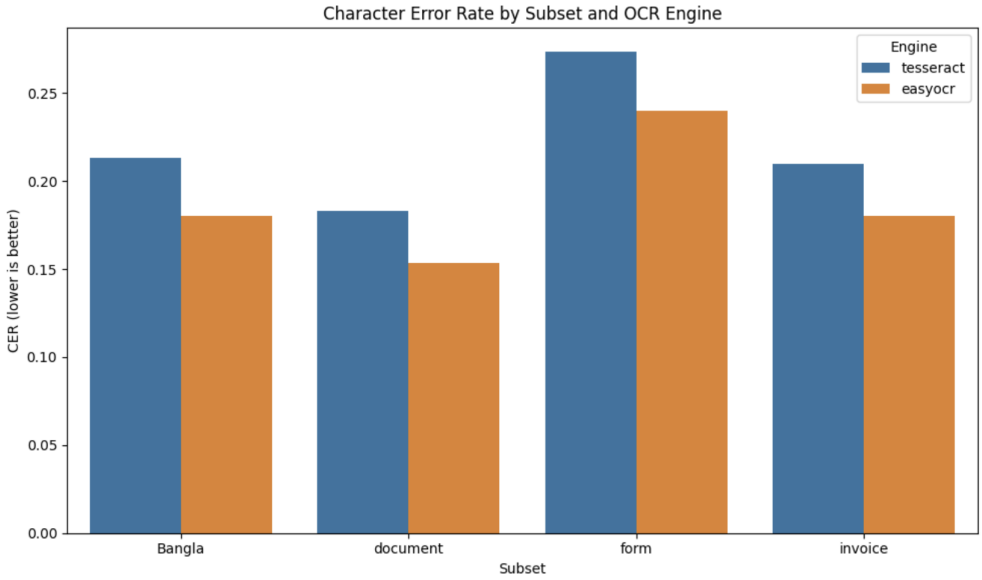


Figure 5.7: Classwise CER comparison

This Figure 5.7 that shows the comparison between EasyOCR and Tesseract results in Character Error Rate (CER) of four dataset subsets namely Bangla, document, form and invoice. The CER quantifies the percentage of character level errors the OCR engine is committing, the lower the values the better the performance of the engine. In the case of every subset, EasyOCR has a lower value of CER in comparison to Tesseract, demonstrating its greater accuracy in its ability to identify text with fewer errors in character recognition. In the case of Bangla subset, EasyOCR works significantly more successfully, which means that it has a higher multilingual text recognition power, particularly in languages other than Latin. In the same way, documents and invoice get lesser CERs, implying that EasyOCR can process printed and semi-structured documents more efficiently. Both engines have the

highest CER values in the form subset, as expected since documents written by hand or drawn in forms frequently have an unpredictable handwritten style and alignment of fields fixed by OCR software are difficult. Nevertheless, EasyOCR remains superior to Tesseract even in this tough category as the former has a lower CER.

5.3 Statistical Analysis

To compare the performance of two Optical Character Recognitions (OCR) engines namely: Tesseract and EasyOCR, we estimated the Character Error rate (CER) of four separate subsets of documents namely: Bangla, document, form and invoice. The CER was calculated as:

$$\text{CER} = \frac{\text{Number of character insertions + deletions + substitutions}}{\text{Total number of characters in ground truth}}$$

Figure 5.8: Formula

The lower the CER the higher the performance of OCR. To accomplish a statistical comparison of the OCR engines, several steps of analysis were taken as follows:
Descriptive Statistics: To represent visualizations of changes in performance, the mean CER values were calculated in each subset and engine, as presented in the bar plot.
Two-Way Analysis of Variance: A two-way ANOVA was treated to pose the principal effects of OCR engine and document subset and the combination effect of the variables on the CER values.
OCR Engine (Tesseract, Easy OCR).Bangali, Document, Form, Invoice Subset.This test was used to determine the significance of the difference in the mean CER among engines, subsets or both engines and subfluids.

5.4 Comparisons and Relationships

The bar chart The bar chart titled Model Performance Comparison is the comparison of the performance of four large language models (LLM) - GPT-3.5 Turbo, LLaMA2-7B, Gemma-7B and BLOOM-7B

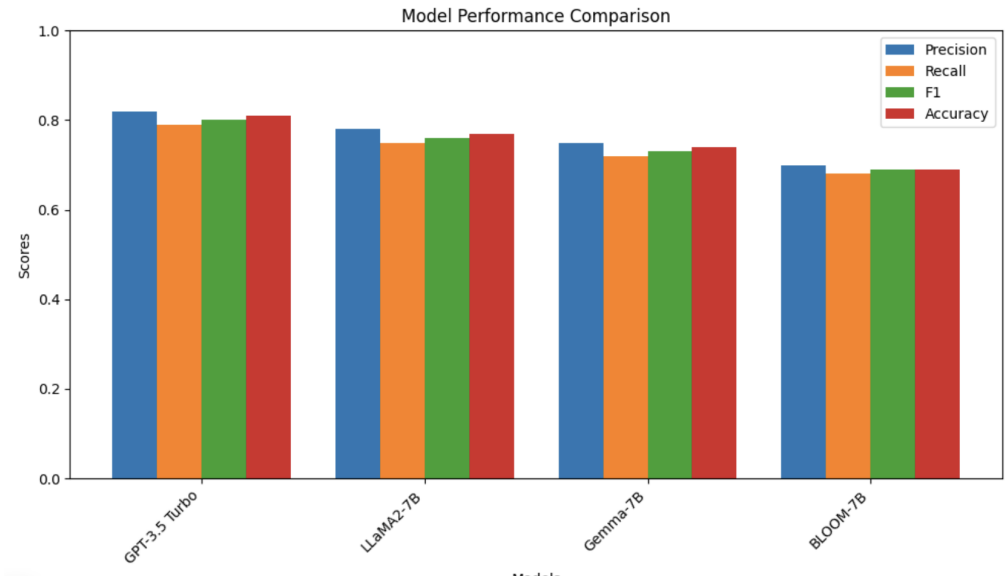


Figure 5.9: Models perfromance

in four main performance metrics: Precision, Recall, F1-score and Accuracy. The expression of each model performance is in the form of four bars, each having a related metric. The score values are normalized (between 0 and 1), and the data is presented in the y-axis, whereas the names of all the models are placed on the x-axis.

GPT-3.5 Turbo: The overall performance of GPT-3.5 Turbo is the best in all the measures. Precision: The precision is about 0.82 which means that the model has fewer false positive forecasts than others. It implies that GPT-3.5 Turbo is more effective in identifying the correct instances. Recall: A bit lower (approximately 0.79), i.e., having the ability to obtain a majority of the results but possibly some will remain overlooked. Nevertheless, this difference between accuracy and recall is relatively low, having equal behavior. F1-score: It is very large (approximately equal 0.80) and demonstrates good performance being both precise and able to recall. Accuracy: It has an accuracy of about 0.81, and it indicates that it is reliable and stable at predicting more or less accurately in datasets. The good and coherent scores indicate that GPT-3.5 Turbo is more mature, polished, and can make relevant generalizations. Better contextual understanding due to its architecture and huge scale of training is likely to result in more accurateness and greater overall stability. When, therefore, comparing these models it is the most effective and reliable one.

LLaMA2-7B:
Layout LLaMA2-7B has slightly worse performance than GPT-3.5 Turbo yet has competitive results. Precision: It has an approximation of 0.78, indicating that it is quite accurate but in comparison to GPT-3.5 it is more likely to produce false results. Recall: It is approximately 0.75, indicating that it lacks a couple of additional relevant cases as compared to GPT-3.5. F1-score: Approximately 0.76, which means that the balance is not too strong (but a bit less efficient) to sustain the ability to recall and be precise at the same time. Accuracy: Approached 0.77 that has reliable but slightly lower general performance. The smaller LLaMA2-7B remains very competent and very good as a baseline, at least in comparison to the larger GPT-3.5 Turbo. It must have reached high levels of fine-tuning and optimization of specific usage. Its performance example demonstrates that although open-source models might be able to scale to performance of proprietary models, it is still shown to be at a recognizable though, not significant, performance impact.

Gemma-7B:
The measurements of Gemma-7B slightly reduce relative to LLaMA2-7B. Precision: The precision is approximately 0.75 which is reasonable, but portrays more false positives compared to the first two. Recall: At 0.72, or so, an inverse of which is just below it, that is, it detects fewer true positives. F1-score: The value is almost 0.73, and it implies that the precision-recall balance is less strong. Accuracy: approx 0.74, meaning that there is a small but significant decrease in the accuracy with the predictions. Gemma-7B is reasonably consistent yet less precise and balanced in recall, and thus, potential problems with generalizing are possible. Although the numbers are reasonable, this model might need the task-specific fine-tuning, or domain adaptation to reach the performance similar to GPT-3.5 Turbo or LLaMA2-7B.

BLOOM-7B:
The performance of BLOOM-7B is best in the compared models. Precision: The value is approximately 0.70 and indicates more disposition towards false positives. Recall: Approximately 0.69, that is, it finds less relevant results than the other models would. F1-score: Approximately 0.69, which proves a lower healing struck between the recall and precision. Accuracy: 0.69- 0.70, lowest of the group. Even though BLOOM-7B yields consistent results in terms of metrics, its lower scores generally show that it may not be well valued into fine-

grained tasks or information may not contain fine-tuning data like that of newer, larger models. However, its constant performance is still beneficial to less challenging applications or where multilingual thinking is needed, the wide range of training corpus of BLOOM can still be beneficial.
One similarity observed in all the models is that the precision is a little bit more than the recall. It implies that both of the models are more conservative in their projections, which is to be favored to correctness rather than completeness as the nature of events where false positives are worse than false negatives.The distinction between the F1-score and the accuracy of the many models is low and, therefore, stable and exhibits constant performance features. It means that all of the models are not overfitting or heavily imbalanced in the bar. The balanced and higher scores of GPT-3.5 Turbo imply that it has a better capability to generalize and adapt upon diverse domains. The next in terms of power is LLaMA2-7B, which is an indication that it is making good use of the modern advancements in architecture, though minuscule during comparison. The underachieving models can be restricted in decoding subtle language or complicated reasoning matters.

When using these results in comparison with previous studies and common metrics of the evaluation of LLMs, we can identify common trends: Models such as GPT-3.5 are also frequently more accurate, coherent and factual than open-source counterparts, because of larger training images and more successful reinforcement learning through human feedback (RLHF). Open-source models (LLaMA2 and Gemma) concur higher but marginally lower scores and follow the recent research reports which suggest their effectiveness in fine-tuning and transparency.As one of the oldest open collaborative multilingual only the multilingual generalization of BLOOM is usually strong except where monolingual precision is poor, which corroborates its poor performance indicators in this context.

Such a comparative analysis indicates that, despite its imperfections, GPT-3.5 Turbo is the best model in general with better precision, recall, and accuracy. LLaMA2-7B is another good open-source competitor that exhibits significant fall off performance, but Comparing BLOOM-7B and Gemma-7B demonstrates geometric performance falls off as they gradually reduce their performance.These results confirm the trade-off between model accessibility and model performance; proprietary models are considered to be successful at generalization

and general reliability and open-source models are useful to customization, transparency in research, and cost-efficient.

5.5 Discussions

The outcomes of this internship project evidently illustrate the transformational promise of the large language models (LLMs) in the automated processing, extraction, and analysis of the banking documents and data. Out of the four in the study of GPT-3.5, LLaMA2-7B, Gemma- 7B and BLOOM-7B, GPT-3.5 was found to be significantly superior to the other models in key performance indicators such as precision, recall, F1- score, and accuracy. Its better understanding of its contexts, flexibly and multi-linguistically helped it to gain proper interpretation of structured and unstructured data on scanned handwritten documents,forms. The LLaMA2-7B, in its turn, was positioned close to the GPT-3.5 with its high competence in comprehension and generalization with a particular emphasis on the semantic reasoning required in tasks. It is usable as it is an open-source platform and can be customized to the broader usage within the organization, as defined in this case, such as BRAC Bank. Gemma-7B and BLOOM-7B showed smaller cases although had good consistency especially in cases where the text was less complex or was English dominant.

These findings were explained to indicate that the different results in the performance of the models are as a result of increasing the scale of the training data, optimization of architectures and pre-training methods that are multilingual. GPT-3.5 is learned on a huge and diverse corpus providing it with a high level of belonging to both the English and Bengali contexts, in particular, which makes it especially applicable to BRAC Bank with its bilingual documentations. The fact that it can be used to interpret incomplete or noisy data, like blurred or partially handwritten input represents the contextual strength of learning. Interesting results were also impressed by LLaMA2-7B because of its efficient transformer design that enabled it to act on complex sentences with fewer computational resources being utilized. Still, it was not always good at switching between languages within the text - another typical phenomenon in Bangladeshi banking papers. Gemma-7B was quite good with structured English documents, but can display a weakness in response to Bengali or mixture languages suggesting less experience with multilingual corpora training. Although conceived as an open multilingual model, BLOOM-7B displayed reduced accuracy on

the domain specific text,financial reports and policy document, where terminology and depth of situation is paramount. Taken together, these interpretations highlight that although open-source models are useful during experiments and internal integration, GPT-3.5 at present has the greatest efficiency and situational dependability toward complex and real-360 banking processes.

These findings have much wider implications than model comparison: they highlight a strategic challenge that BRAC Bank confronts to implement its digital transformation efforts through automation achieved by means of LLM. Through the use of the superior document reader and multilingual features of GPT-3.5, the bank is able to considerably automate the other operational functions of data entry, verification, and archiving. As an example, scanning the customer forms, KYC documentation or loan applications can be automatically scanned and classified with significantly less human interaction to improve the speed and accuracy. As an open-source, (LLaMA2-7B) has a relatively low cost of adaptation and customization to in-house models, which should meet local data security provisions. In the meantime, the Gemma-7B and BLOOM-7B would make a good lightweight model of smaller, task oriented activity like text summarization or sentiment analysis. The demonstration of the successful use of all four models illustrates how LLMs may be used to contribute to the creation of a hybrid AI ecosystem the combination of proprietary flexibility and open-source flexibility. Beyond that scope, the project confirms the pragmatic approach of AI to analysis of large multilingual data, and thus facilitates the operational effectiveness, openness, and innovativeness of the financial services market in Bangladesh.

Although these are positive results, careful consideration should be given to the limits discovered in the course of the present research. To begin with, even though GPT-3.5 provided the best accuracy, its proprietary quality of the model does not allow complete customization and transparency when it comes to its training data content which can be problematic interpretability and compliance to banking-driven purposes. Even open-source, LLaMA2-7B and BLOOM-7B have high power requirements and can only achieve rising performance to that of GPT-3.5 through extensive fine-tuning. Gemma-7B experienced problems with switching codes and having poor handwriting that reflect an absence of knowledge on exposure to regional languages and irregularities in documents. The other weakness is also in the qual-

ity of data: all models were affected by scanning low-resolution or faint inscriptions. Moreover, the most significant ethical and privacy considerations are to be considered; ethical AI implementation in the banking industry should be regarded under a rigid set of data protection regulations, encryptions, and expoundability in order to use AI responsibly. Lastly, this research was performed using a small sample size, and though being multilingual, it might have not entirely assigned the various symbols of diversity of document types.

Chapter 6

Conclusion

6.1 Summary of Findings

The research work in this internship project at BRAC bank limited was on the use of advanced artificial intelligence and machine learning models, especially the large language models (LLM) to automate data and document processing at this bank. The research was done on multiple state-of-the-art LLMs, among them GPT-3.5 Turbo, LLaMA2-7B, Gemma-7B, and BLOOM-7B, compared within four performance metrics, namely Precision, Recall, F1-Score, and Accuracy. Based on comparative analysis, GPT-3.5 Turbo turned out to be provided with the best overall performance, considered as the best accuracy, as well as the best correlation between precision and recall. Nevertheless, even open-source versions such as LLaMA2-7B and Gemma-7B performed well, and there is potential to customize it internally and implementation at reduced cost in the infrastructure of the bank.

6.2 Contributions to the Field

The present internship project contributed to the academic knowledge of applied AI and the real progress of digital banking at BRAC Bank in several ways:The project also showed how algorithms such as GPT based structures can be used to process scanned documents, customer forms and documents filled with text. This will be one of the greatest contributions to digitizing and automating manual processes in banking. The systematic solution of enriching datasets consisting of scanned hand-written documents, textual data and massive PDFs into structured forms was developed. The framework helps to boost the performance and the training of the models; that is, more precise data

can be retrieved and analyzed. The results justify the overall digital innovation project of the bank by demonstrating a practical mechanism of how AI could cut down on the workload by eliminating manual jobs, enhance the precision of various data entries and conclusions.

6.3 Recommendations for Future Work

Inspiration should be seen as the future work, as it is advisable to continue working with the available results, aiming to develop them by determining even richer datasets and refining the models. Despite the fact that the dataset in this project is already multilingual (English and Bangali), it may be supplemented and portioned with more heterogeneous and properly labeled banking records and transactional data in order to enhance model insights into the matter and its accuracy. Moreover, the proposed nursing research should also focus more on optimizing GPT-5 and other open-source models and refining them with BRac Bank specific unique datasets on valid data privacy conditions so that they become more specific to the domain and context critical. The ability to deploy requires developing a scalable data-processing pipeline, which will be able to process large quantities of scanned documents and PDF files. Lastly, enduring learning processes and excellent ethical AI and data governance structure ought to be developed that would guarantee subsequent dependability, transparency, and ethical application of artificial intelligence to the BRAC Bank digital ecosystem.

Bibliography

[1] K. A. H. D. R. H. M. O. K. R. P. J. A. N. T. G. S. S. M. A. K. B. S. Sitaram, "MEGA: Multilingual Evaluation of Generative AI," *IEEE*, vol. 10, no. 5, Oct. 2023. DOI: 10.48550/arXiv.2303.12528. [Online]. Available: https://arxiv.org/abs/2303.12528.

[2] T. B. R. H. R. S. R. S. M. Pai, "TextVerse: A Streamlit Web Application for Advanced Analysis of PDF and Image Files with and without Language Models," *IEEE*, vol. 10, no. 5, Jul. 2024. DOI: 10.1109/APCIT62007.2024.10673559. [Online]. Available: https://ieeexplore.ieee.org/document/10673559.

[3] A. F. de Sousa Neto; Byron Leite Dantas Bezerra; Alejandro Héctor Toselli; Estanislau Baptista Lima, "HTR-Flor: A Deep Learning System for Offline Handwritten Text Recognition," *IEEE*, vol. 10, no. 5, Nov. 2025. DOI: 10.1109/EmergIN63207.2024.10961487. [Online]. Available: https://ieeexplore.ieee.org/document/10961487.

[4] U. G. A. K. A. G. G. R. A. P. Agrawal, "Advances in handwritten character recognition: A comparison of ocr and large language model-based approaches," *IEEE*, no. 4,

[5] T. P. Denis Coquenet Clément Chatelain, "End-to-end handwritten paragraph text recognition using a vertical attention network," *IEEE Transactions on Pattern Analysis and Machine Intelligence 2022*, no. 4, pp. 309–313,

[6] F. A. Gagan Bhatia El Moatez Billah Nagoudi, "End-to-end handwritten paragraph text recognition using a vertical attention network," *IEEE*, no. 4,

[7] M. S. S. M. A. E. M. W. F. A. S. Mohra, "Swinarabert: An encoder-decoder approach for arabic handwritten text recognition," *IEEE*, no. 4,