Enhanced Hybrid Technique for Efficient Digitization of Handwritten Marksheets

by

Junaid Ahmed Sifat
23241117
Abir Chowdhury
20201112
Hasnat Md.Imtiaz
20141004
Sayma Akter Lubna
20301450

Department of Computer Science and Engineering
Brac University
February 2025

Declaration

It is hereby declared that

1. The thesis submitted is my/our own original work while completing degree at Brac University.
2. The thesis 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 thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.
4. We have acknowledged all main sources of help.

Student's Full Name & Signature:

Abir Chowdhury 20201112	Junaid Ahmed Sifat 23241117
Hasnat Md. Imtiaz 20141004	Sayma Akter Lubna 20301450

Approval

The thesis titled *"Enhanced Hybrid Technique for Efficient Digitization of Handwritten Marksheets"* submitted by

1. Junaid Ahmed Sifat (23241117)
2. Abir Chowdhury (20201112)
3. Hasnat Md.Imtiaz(20141004)
4. Sayma Akter Lubna (20301450)

Of Fall, 2024 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science on February, 2025.

Examining Committee:

Supervisor:
(Member)

Md. Imran Bin Azad
Senior Lecturer
Department of Computer Science and Engineering
Brac University

Program Coordinator:
(Member)

Dr. Md. Golam Rabiul Alam
Professor
Department
Brac University

Head of Department:
(Chair)

Dr. Sadia Hamid Kazi
Chairperson and Associate Professor
Department of Computer Science and Engineering
Brac University

Abstract

The digitization of handwritten marksheets presents huge challenges due to the different styles of handwriting and complex table structures in such documents like marksheets. This work introduces a hybrid method that integrates OpenCV for table detection and PaddleOCR for recognizing sequential handwritten text. The image processing capabilities of OpenCV efficiently detects rows and columns which enable computationally lightweight and accurate table detection. Additionally, YOLOv8 and Modified YOLOv8 are implemented for handwritten text recognition within the detected table structures alongside PaddleOCR which further enhance the system's versatility. The proposed model achieves high accuracy on our custom dataset which is designed to represent different and diverse handwriting styles and complex table layouts. Experimental results demonstrate that YOLOv8 Modified achieves an accuracy of 92.72%, outperforming PaddleOCR 91.37% and the YOLOv8 model 88.91%. This efficiency reduces the necessity for manual work which makes this a practical and fast solution for digitizing academic as well as administrative documents. This research serves the field of document automation, particularly handwritten document understanding, by providing operational and reliable methods to scale, enhance, and integrate the technologies involved.

Keywords: Handwriting Recognition, PaddleOcr, OpenCV, YoloV8, Table Detection, Document Digitization

Acknowledgement

Firstly, all praise to the Great Allah for whom our thesis have been completed without any major interruption.
Secondly, to our supervisor Md. Imran Bin Azad sir for his kind support and advice in our work.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Introduction

In an era where digital solutions are changing in every aspect of society , the digitalization of handwritten documents still remains an important challenge particularly in academics. Even with the increased use of automation, marksheets are still handled manually which makes the process time consuming, full of delays and mistakes. Consequently, it has led to a growing demand for an integrated system for scanning handwritten mark sheets which can be digitized with high accuracy. In response to this problem, this study uses computer vision approaches and deep learning algorithms to automate handwritten mark sheets to digital format. Manual marksheets are difficult to process because they use different handwriting, and the tables included are difficult to process regularly. Such inconsistencies were always an issue in traditional document digitization approaches and affected the recognition of both textual and tabular contents. Previous approaches are incapable of providing the flexibility needed to detect marksheets and other handwriting, as they can become complex to generalize as they are used across a wide range of handwriting styles, where marksheets include both alphabetical numbers and well tabulated, finely handwritten structures. Advanced deep learning techniques like the use of PaddleOCR help to overcome these challenges. PaddleOCR is designed for accurate and efficient optical character recognition (Ocr) which makes it well suited for recognizing handwritten contents on marksheets. Its ability to handle different and diverse handwriting styles and adapt to varying layouts ensures precise transcription of handwritten text. In this research, to detect rows, columns, and table structures OpenCV is utilized and it is providing a foundational framework for identifying the layout of marksheets. Once the structures are identified, the recognized structures lead PaddleOCR to perform a handwritten character recognition task inside the detected areas.By using OpenCV for table identification with PaddleOCR for text recognition, both the content of the marksheet and structural organization remain intact throughout the digitization process.The method enhances accuracy in data extraction through solutions that handle problems in inconsistent handwriting and diverse table frameworks. This proposed system adapts to various marksheet layout structures and achieves high mark transcription accuracy from handwritten sources. The findings of this research contribute in manual data entry while improving precision and uniformity as well as reducing the time and effort involved. In addition to the use of PaddleOCR, this research also explores the implementation

of YOLOv8 (You Only Look Once version 8) for text detection in marksheets.The state-of-the-art object detection model YOLOv8 excels in object recognition while maintaining fast operation speed alongside precise results. The task of character recognition remains PaddleOCR's core function while YOLOv8 specializes in detecting and localizing text regions within marksheet documents. YOLOv8 identifies textual regions through its customized training process for marksheets where precise text recognition leads to later transcription processing. The tool demonstrates broad applications through its capacity to recognize multiple document formats and it functions effectively with complex markup layout structures.YOLOv8 is employed in this study as a standalone method, independent of PaddleOCR, to evaluate its effectiveness in detecting and segmenting text regions efficiently. YOLOv8 demonstrates superior performance for text identification alongside text segmentation but its application only extends to detecting text areas.For transcription of handwritten content, PaddleOCR is used in a separate pipeline. Through individual deployment of these models users gain the freedom to select which method best suits their unique purposes. This research implements YOLOv8 for text detection along with PaddleOCR for text recognition to create a complete framework for marksheet digitization. These methods function separately to solve unique elements of the problem and maintain the system's modular design along with efficiency and adaptability to multiple requirements. The combination of these techniques enables automated academic workflow enhancement which supports overall document digitization advances.

1.2 Research Problem

With the growing demand for automated solutions in document digitization, specially in administrative sectors alongside academic sectors, the challenges of processing these handwritten documents and marksheets remain significant. Current and existing methods are somewhat limited and can fail to address both multiple table structures and different kinds of handwriting. There remains a need for a general approach to processing tables whose structure may vary while recognizing sequential handwritten characters and numerals is still a challenge to current technologies. In recent studies, researchers have proposed OpenCV-based techniques for table detection, where rows and columns are effectively identified in document images. OpenCV assists table detection because it processes data images with detailed accuracy. but these method posses some limitations. One of them is the resizing of images which can introduce distortions and loss of the imformations in the following documents. The powerful image processing tool found in opencv struggles to efficiently handle processing of an extensive number of documents like marksheets because it needs massive computing resources. Furthermore, this OpenCV based method can sometimes struggle to generalize different and diverse tablular formats when dealing with some irregularities like merged cells and nonstandard layouts. For this, Additional postprocessing ways are often required to refine features, ensuring accurate segmentation results. this makes further complexity to this system and it will demand layers which are computational and also it will processing time. Handwritten character recognition research uses deep learning methods including Convolutional Neural Networks (CNNs) and Convolutional Recurrent Neural Networks (CRNNs) which have become extensively studied across different domains. Multiple studies

show promises that deep models recognize isolated handwritten characters alongside printed characters. However, their application to structured document processing, such as marksheets, poses challenges. While effective for detecting single characters, these models often fall short in handling sequences of handwritten characters or digits that are presented in structured formats. The constraints are especially evident when processing high-resolution images or lengthy text sequences, as these models are often computationally demanding and time-consuming.Sequential recognition models, such as PaddleOCR, have been designed to address these issues. PaddleOCR provides a comprehensive solution for processing character sequences and enhances the recognition of handwritten text in a structured format. It is especially appropriate for marksheets, where tabular data and handwritten sequences coexist. Nevertheless, despite its benefits, PaddleOCR incurs significant computational expenses, rendering it less feasible for real-time applications or extensive processing jobs. This trade-off demands meticulous evaluation of its implementation, especially in situations that involve both rapidity and precision.In light of the dual challenges of identifying table structures and recognizing handwritten data in marksheets, our research suggests a hybrid and multimodal solution. Instead of depending exclusively on deep learning techniques for table detection, we employ OpenCV, which is more computationally efficient than its deep learning alternatives. OpenCV proficiently detects rows, columns, and table configurations in document pictures, establishing a basis for subsequent processing.This method reduces computational burden while preserving the necessary precision for table detection.We incorporate PaddleOCR as a sequential recognition model for the identification of handwritten characters. By integrating the capabilities of OpenCV with PaddleOCR, we establish a system proficient in processing marksheets with exceptional accuracy and efficiency. OpenCV is utilized to identify all cells within the marksheet, which are subsequently stored as separate images in a specified folder. Each cell is then analyzed with PaddleOCR to retrieve handwritten characters. The gathered data is subsequently organized and exported in Excel format, yielding a structured digital output of the marksheet.This hybrid methodology efficiently addresses the issues of table detection and handwritten character recognition. Utilizing OpenCV for structure identification and PaddleOCR for text recognition, we develop a system that optimizes computational efficiency and precision. The proposed methodology diminishes processing time and computational requirements, hence enhancing the feasibility for academic institutions to digitize marksheets on a broader scale. Simultaneously, it enhances the accuracy and consistency of data extraction, facilitating progress in document processing and the automation of academic procedures. In our endeavor to reconcile a computationally efficient table detection method with a dependable handwritten text recognition system for practical applications, it is prudent to evaluate the implications of the theoretical upper bound of computational complexity for each algorithm in a real-world context. This research examines the application of YOLOv8 (You Only Look Once version 8) for text detection in marksheets, in addition to the previously stated approaches. YOLOv8 is a cutting-edge object identification model recognized for its efficacy and accuracy in recognizing objects, including text regions. This study utilizes YOLOv8 as a method to identify text regions in marksheets, emphasizing precise localization of text areas for further processing. In this research, YOLOv8 is utilized independently of PaddleOCR, offering a modular approach for text identification. The model is optimized to identify handwritten text areas in

marksheets. YOLOv8 has superior capability in managing varied layouts, encompassing abnormalities like merged cells and inconsistent handwriting styles. The system generates a pre-processed output by identifying text sections with YOLOv8, which may further be evaluated using additional tools or algorithms. While YOLOv8 emphasizes the detection of text regions rather than direct text recognition, its incorporation into the workflow improves the overall precision and flexibility of the marksheet digitization process. Notwithstanding the progress achieved by YOLOv8 and PaddleOCR in marksheet digitization, numerous problems remain. Generalizing across different handwriting styles and sizes, especially in academic marksheets, continues to be a substantial challenge, as handwriting can significantly differ in complexity and intelligibility.Moreover, the substantial computational expenses linked to models such as YOLOv8 and PaddleOCR impose constraints, particularly for large-scale applications that utilize extensive datasets. Achieving a balance between efficiency and accuracy in resource-limited settings is essential for tackling these difficulties. Furthermore, although YOLOv8 demonstrates proficiency in text localization, the smooth integration of its outputs with subsequent processes, such as character recognition and structured table reconstruction, necessitates a resilient workflow that prevents data loss or misinterpretation. Ultimately, attaining real-time or near-real-time processing for marksheet digitization is a difficulty due to the computational intricacy of text detection and recognition tasks, which are essential for time-sensitive academic and administrative operations.

1.3 Research Objective

The main goal of this research is to establish an efficient approach for recognizing table structures and handwritten text in marksheet images. The study aims to address the dual challenges of detecting diverse table layouts and recognizing sequential handwritten characters and numerals. The research aligns its hybrid approach with OpenCV for table structure detection alongside PaddleOCR as well as YoloV8 for handwritten text recognition to minimize both overhead and achieve precise results and adaptable performance.The specific research objectives for this work:

1. Develop a method that detects and isolates individual cells in marksheet visuals so systems work effectively while accommodating different types of table arrangements.
2. Identify detected cells by assigning correct boundary boxes which preserve the document structure so tabular data can be correctly extracted.
3. Recognize each detected cell contains handwritten input which PaddleOCR and YoloV8 process separately and also maintains the high accuracy in the diverse and varying handwriting styles and formats.
4. Systematically extract data from identified cells for transfer into an Excel file and also provide a structured as well as digital representation of the marksheet.

This set of objectives emphasizes the development of a practical, efficient, and scalable solution for digitizing marksheets while overcoming challenges related to diverse table structures and handwriting variations. By achieving these goals, this research contributes to the advancement of document digitization technologies, particularly in academic and administrative contexts.

1.4 Thesis Structure

Our thesis, titled "Enhanced Hybrid Technique for Efficient Digitization of Handwritten Marksheets", is intelligently designed as follows: It starts with Chapter 1: Introduction, where it outlines the main idea of the research, clarifies the research problems and the objective of the study Chapter 2: Literature section provides an overview of text recognition, digit recognition using OCR, and table detection methodologies. Chapter 3: Dataset description where diversive 10000 primary data were used. Chapter 4: Methodology follows, describing methods for detecting cells in each row and column, applying OCR to extract text, and systematically storing the results in an Excel file. Chapter 5: Result Analysis presents a comparative analysis of our models against other approaches, evaluates performance on test cases, and examines the effectiveness of PaddleOCR in handling complex sheets compared to a web application. Finally, Chapter 6: Finally, the conclusions reveal what we found, the importance of our research, and what possible additional work is to be done.

Chapter 2

Literature Review

Shahab et al.[3] introduced an open benchmarking framework, emphasizing ground truth preparation and performance evaluation metrics for table structure recognition. Their method included manual annotation steps with heuristic row-column segmentation methods, which needed specific performance adjustments for each dataset. Consequently, this approach proved limited in dealing with irregular or handwritten tables. The evaluation approach they created gave structure to the assessment but proved inadequate in handling both dynamically changing tables and unstructured free-form handwritten inputs in practical conditions. The deep learning-based techniques applied by our research solve these restrictions in the method. We implement a mix of algorithms that utilize OpenCV for table structure detection together with PaddleOCR for text recognition and also separately use YOLOv8 and our proposed model for text recognition to abolish the need for predefined rules and manual feature engineering requirements.

Text recognition from images has been a significant research area, evolving from rule-based methods to deep learning-driven approaches. The researchers[6] have proposed a Document Image Analysis (DIA)-based method, where scanned text undergoes preprocessing, segmentation, feature extraction, and classification using Artificial Neural Networks (ANNs) for character recognition. While their approach effectively processes structured text with predefined font characteristics, it struggles with handwritten variations, skewed text, and complex tabular structures, which are common in marksheets. Their reliance on template matching and fixed feature sets limits adaptability across diverse datasets, making real-world implementation challenging. In contrast, our research eliminates manual feature engineering by leveraging OpenCV, PaddleOCR, and separately using YOLOv8, forming a hybrid deep learning-based pipeline for automated handwritten text and table detection. Unlike their segmentation-dependent method, our Modified YOLOv8 dynamically learns from diverse handwriting styles, improving detection accuracy (92.72%) and reducing false classifications. While their study contributes to foundational OCR methodologies, our research extends its practicality by enabling scalable, high-precision handwritten marksheet digitization. By overcoming challenges related to inconsistent handwriting, structural distortions, and segmentation errors, our approach provides a robust, adaptable, and automation-friendly solution, addressing the key limitations of traditional DIA-based OCR techniques.

The extraction of text from images has been a longstanding research focus, transitioning from conventional rule-based OCR methods to advanced deep learning-driven approaches, enabling greater accuracy, adaptability, and automation in text recognition. The researchers S. Umatia, A. Varma, A. Syed, K. Tiwari, and M. F. Shah, in [30] have discussed a combination of OpenCV and Tesseract OCR to extract text from invoices and scanned documents, aiming to improve recognition accuracy through preprocessing, segmentation, and text classification. Their approach involves converting color images to binary, applying edge detection, and using contour detection for text region extraction, followed by Tesseract OCR for final text recognition. While this method proves effective for structured printed text, it struggles with handwritten variations, text misalignment, and complex tabular structures, where segmentation errors and misclassification can occur. Conversely, our research emphasizes the digitization of handwritten marksheets, utilizing PaddleOCR and Modified YOLOv8 independently and separately for text recognition, thereby achieving an optimal synergy between OCR-based text retrieval and deep learning-powered localization. While both studies focus on text recognition from images, the reviewed work relies on Tesseract's rule-based OCR, whereas our approach integrates OpenCv for table detection and PaddleOCR for handwritten text detection as well as YOLOv8 for text detection within complex dataset, allowing for greater adaptability across handwriting styles and document layouts. Unlike their pipeline, which depends on predefined contours and traditional image processing techniques, our hybrid deep learning-driven method dynamically learns patterns, significantly reducing segmentation errors. While the reviewed study presents a practical OCR pipeline for printed documents, our research extends its applicability by ensuring scalability and high precision in handwritten marksheet digitization, bridging the gap between conventional OCR and deep learning-based document automation.

The authors, M. Laine and O. S. Nevalainen showed in [2] that a simple OCR system on a Nokia 6630 camera-phone can be implemented without relying on dedicated hardware or network facilities. The researchers have shown the implementation of OCR (optical character recognition) on mobile platform. This paper shows some subsystems to work with. Image Capturing, Preprocessing, Segmentation, Feature Extraction, Character Recognition are the subsystem of this work. In case of Image Capturing, test are initially done in RGB color images, and then it would be converted to grayscale. In case of Preprocessing, binarization is the only Preprocessing process used in this system. In case of Segmentation, the process was done using skew angle detection and X-Y-Tree decomposing. The skew angle detection is basically a hypothesis testing for two algorithm. The first technique uses a fast algorithm to approximate the skew angle. Although the second technique is slower, it only tests a small number of angles because the approximate angle is already known. A straightforward segmentation approach that works well for images with mostly text lines as their content is called X-Y Tree decomposition. In case of Feature Extraction, the paper suggests using Centroid-to-Boundary Distances. It accepts the centroid as an input and outputs a vector of centroid-to-boundary distances in every direction. In case of Character Recognition, lowercase letters are correctly executed. Each character is extracted into a brief rectangle with dimensions equal to its height plus its ascent. Result wise, English capital letters written in black

on a white background were recognised by the system. The study explores the advantages and disadvantages of implementing OCR on mobile platforms equipped with image-capture capabilities.Limitation wise, the system is limited to only recognize with English capital letters printed in Black against a white background. The system is for Nokia 6630 in limit. Performance and accuracy are not that much discussed here.In our Work, We have Done table extraction using OpenCv and text detection using paddleOcr and yolov8 separately in english digit where the accuracy looks decent compare to this.

The researchers R. Jana, A. R. Chowdhury, and M. Islam [5] adopt a texture and topological feature-based approach, where characters are segmented, and their geometric properties, such as corner points, convex hulls, and pixel distributions, are analyzed for recognition.The system succeeds in structured environments but struggles to process various complicated writing styles together with poorly written sections in different quality lighting. Additionally, its reliance on predefined templates for character matching restricts its adaptability to diverse datasets. In contrast, our research introduces a deep learning-driven hybrid approach that eliminates the constraints of static feature sets. By integrating OpenCV for table structure detection, PaddleOCR for robust handwritten text detection and also YOLOv8 for detection, our method adapts to diverse handwriting styles and complex tabular structures without requiring manual feature engineering. Unlike the reviewed study, which processes characters in isolation, our model leverages deep learning's ability to recognize contextual relationships between handwritten text and table structures, improving both detection accuracy and segmentation robustness. While their research contributes valuable insights into feature-based OCR, our approach enhances real-world applicability in recognizing handwritten marksheets, reducing segmentation errors, and ensuring higher adaptability across varying document formats. By addressing the limitations of character misclassification, and structural inconsistencies, our work advances the field by offering a scalable, efficient, and automated solution for digitizing handwritten marksheets.

The Author S. Malakar, S. Halder [4] presents a segmentation-based approach, emphasizing heuristic-driven table structure identification and optical character recognition for extracting handwritten content.Their approach gives structured detection methods for tables, yet it faces challenges working with complex layouts, tilted text, and different handwriting styles because of dependent predefined rules and fixed heuristics. The system faces performance restrictions that prevent it from offering suitable solutions for numerous document forms, thus reducing its productivity in practical usages. In contrast, our research moves beyond predefined feature sets by incorporating deep learning-based automation, integrating OpenCV for table structure detection, PaddleOCR for robust text detection and YOLOv8 for object detection. Unlike the reviewed study, which follows a rule-dependent segmentation process, our approach employs a custom model inspired by the YOLOv8 model that adapts to different handwriting styles and minimizes false detections. Furthermore, while their research primarily focuses on evaluating segmentation techniques, our study goes further by ensuring practical usability in large-scale handwritten marksheet digitization. By overcoming challenges related to text misclassification, handwriting inconsistencies, and complex table layouts, our approach delivers a scalable,

adaptive, and high-precision solution, effectively bridging the gap between conventional OCR techniques and modern deep learning advancements.

The researchers J. Jebadurai, I. J. Jebadurai, G. J. L. Paulraj, and S. V. Vangeepuram in [20],proposed a model for Handwritten document recognition which aims to convert handwritten information into digital text.Their model utilizes Convolutional Neural Networks (CNN) to study features, Long Short Term Memory (LSTM) to handle sequential data, and Connectionist Temporal Classification (CTC) loss to deal with text placement variations. It achieves 94% accuracy and a loss of 0.147 on training data and 85% accuracy and a loss of 1.105 on validation data after training on the IAM Handwriting Database. They have expressed the same concern for a myriad number of people who still prefer writing on paper due to the lack of digital note-taking devices or personal preference. If this written information needs to be converted into digital text, a snap of the text can be taken and given to a programmed system, which uses deep learning, specifically neural networks, to understand human handwriting and convert it into digital text. Neural networks are designed based on the functionality of neurons in the human brain and learn patterns in the training data to predict the output for new handwriting inputs. Their model is trained with sample human handwriting to understand how humans write and generate accurate digital text outputs. Convolution is a mathematical operation used in deep learning to analyze images and extract features from them. It is particularly effective in image recognition tasks.Handwritten text recognition involves identifying the features of human handwriting and using them to predict the text. Simply identifying each character is not enough, as words are a sequence of characters and each word contains similar characters in different sequences and in order to understand these sequences, a specific type of Recurrent Neural Network (RNN) called Long Short Term Memory (LSTM) is used. LSTM divides the image into time steps and attempts to understand where each character occurs and what it means. They have trained their model with grayscale images which contain only shades of gray and no color information because Handwritten text recognition requires the model to learn the features of handwriting which are independent of the color of the ink and used paper. Moreover, grayscale images work well with neural network models and hence, all the images of words from the IAM handwriting database are converted into grayscale images. The neural networks work with values between 0 and 1, and therefore, the pixel brightness values of all the images of the word are scaled down from 0-255 to 0-1. This is done by dividing each pixel brightness value by 255, which normalizes the values to a range between 0 and 1. They also suggested that CNNs work well with max-pooling layers which calculate the largest value of a feature map generated by the convolutional layer. This helps in considering only the important and prominent features while ignoring minor features leading to faster training of the model. In our work, we have done paddleocr alongside Yolov8, and also inspired by Yolov8, we have created a custom model to detect text where we get decent accuracy on it.

Text line detection in handwritten documents has been a persistent challenge in document analysis, with various methodologies attempting to enhance accuracy and robustness. Pach and P. Bilski [7] have introduced a Block-Based Hough Transform method, which segments text by detecting connected components (CCs) and

refining text line detection using an extended Hough Transform. This approach effectively addresses handwritten text alignment issues and overlapping characters, particularly in historical manuscripts. However, its dependence on histogram-based analysis and local feature extraction limits its adaptability to complex table structures and modern handwritten datasets, where variations in handwriting styles and tabular layouts pose significant challenges. In contrast, our research incorporates deep learning-driven techniques, combining OpenCV, PaddleOCR, and YOLOv8, to eliminate the constraints of traditional heuristics and manual feature engineering. While the reviewed paper improves text line detection through connected component enhancement, it does not fully address structural inconsistencies in tabular data, which is critical in marksheet digitization. Our Modified YOLOv8 model dynamically learns from diverse formats in handwritten text recognition while reducing false segmentations. Moreover, while their study focuses on historical document analysis, our work extends real-world applicability to large-scale automated marksheet processing, offering a scalable, adaptive, and high-accuracy solution that bridges the gap between classical document analysis techniques and modern deep learning-based OCR systems.

The authors[24] have presented a Convolutional Neural Network (CNN)-based approach trained on the NIST dataset, where the model extracts and learns features from handwritten characters to classify them into different categories. Their methodology includes preprocessing, segmentation, and feature extraction, followed by classification using a multi-layer CNN. The research demonstrates promising results, achieving 90.54% accuracy, yet remains limited in its ability to handle handwritten text within structured tabular formats due to segmentation errors and handwriting variability. Conversely, our research emphasizes the digitization of handwritten marksheets, utilizing PaddleOCR and Modified YOLOv8 independently for text recognition, thereby achieving an optimal synergy between OCR-based text retrieval and deep learning-powered localization. While both studies focus on deep learning-driven text recognition, the reviewed work applies CNN solely for character classification, whereas our approach ensures context-aware recognition within structured documents. Unlike their method, which processes isolated handwritten characters, our hybrid approach separates text detection from table, allowing PaddleOCR and Yolov8 and Modified YOLOv8 to specialize in handwritten text detection separately, achieving higher accuracy (92.72%) in Modified YOLOv8. While the reviewed paper successfully demonstrates CNN's capability for handwritten character classification, our research bridges the gap between isolated character recognition and real-world document digitization, ensuring scalability, adaptability, and high-precision extraction of handwritten marksheets.

According to the authors S. R. N and N. Kennedy Babu C of [26], the challenges of handwritten character and digit recognition are addressed. This is very crucial for various applications like postal mail sorting and bank check processing. In this paper, the ARAN model enhances the recognition through employing a much improved energy valley optimization algorithm, which is IEVOA for the parameter tuning and alongside it, improving accuracy and precision of recognition of this paper are also added. This ARAN model illustrates the superior performance by handling the distorted image and optimizing parameters like epoch size as well as

hidden neuron count and the steps for per epoch count. Furthermore, this paper solves significant limitations of the models, like recognizing overlapping, distorted characters as well as digits with different scales with orientations. IEVOA is used for tuning parameters which makes the model adaptable and efficient. Afterwards, this paper's experimental results show that ARAN model performes better than LSTM, RNN and ResNet in terms of their F1-score and accuracy. This happens when datasets like TamilNet, Devanagari Handwriiten character dataset are used. The improvement of accuracy after using this dataset goes up by 7.48%, and the F1-score improved 76.41% for the dataset 2.In our work, we have detected tables using OpenCV and detected text by using paddleocr, Yolov8, and modified Yolov8, where we get decent accuracy for each model.

The author M. Salaheldin Kasem, A. Abdallah, A. Berendeyev[28] depicted the findings in the field of table detection and structure recognition using deep learning techniques. Table detection and recognition is central to information retrieval and document understanding.The paper focuses on how deep learning solutions have replaced heuristic and machine learning algorithms. HDR research has incorporated and used Machine Learning (ML) models including MLP, SVM and CNN and lastly CNN appears to work very well. The feature extraction and classification are fused in the architecture and are more effective compared to the traditional methods. Recent insights include adding CNNs, Region based CNNs (R-CNN), and transformer based models to add even better performance.The paper points to remaining challenges of table detection on a variety of layouts, document formats and poor quality scans. While accurately identifying table boundaries, differentiating tables from other graphical elements, and resolving multi-page tables are still difficult. However, deep learning has enhanced the detection, but accurate localization of table structures is still required in all formats.Recent approaches by object detection and semantic segmentation are used to recognize table structures. The detection of the complex structures can be done more robustly using graph based networks, hierarchical models and encoder-decoder architectures. Transformer based models, which have seen great success in natural language processing, are generalized to tables and their spatial relationships.The paper reviews datasets like ICDAR, Marmot, PubTabNet, and TableBank, which benchmark table detection models.These datasets consist of scanned documents, digital PDFs and synthetic data, for full evaluation. The approaches work as evaluated by metrics like precision, recall, Intersection over Union (IoU), and mean Average Precision (mAP). Trends like multimodal learning, consisting of textual and visual information, are covered in the survey. The application of pre-trained language models and synthetic data presents encouraging avenues. The fact that there are still issues with complicated table layouts, generalizing across document types, and dataset limitations shows how deep learning is still developing in this area.In our work, we have detected tables using OpenCV and detected text by using paddleocr, Yolov8, and modified Yolov8, where we get decent accuracy for each model. We have worked on our primary dataset where modified Yolov8 get the best accuracy comapring others.

The authors in the paper[21] explore a CNN-based HCR approach, leveraging layered feature extraction and classification to recognize handwritten characters from the NIST dataset. Their methodology involves preprocessing, segmentation, and

training a CNN model to differentiate between character shapes, achieving a 92.91% accuracy with 1000 training images. This approach effectively enhances recognition in structured datasets but remains limited in handling complex tabular documents and diverse handwriting styles, where segmentation errors and character overlap pose challenges. Our research emphasizes the digitization of handwritten marksheets, utilizing PaddleOCR and Modified YOLOv8 independently for text recognition, consequently achieving an optimal synergy between OCR-based text retrieval and deep learning-powered localization. While both studies focus on deep learning-driven recognition, the reviewed paper applies CNNs for isolated character classification, whereas our method ensures context-aware recognition within structured documents. Unlike their approach, which relies solely on CNN-based classification, our hybrid model integrates opencv for table detection and PaddleOCR, Yolov8, modified Yolov8 separately for precise text detection, achieving higher accuracy (92.72%) in modified Yolov8. While their research effectively demonstrates CNN's strength in character recognition, our study bridges the gap between isolated character classification and real-world document processing, ensuring scalability, adaptability, and high-precision extraction of handwritten marksheets.

In the paper [27], author M. Raheem and N. Abubacker have explored a comparative study of CNN, MLP, and SVM models for handwritten digit recognition, aiming to determine the most efficient approach. Their methodology follows data preprocessing, feature extraction, and classification, where CNN outperforms the other models, achieving 99.31% accuracy due to its superior ability to learn spatial features. While their research effectively demonstrates CNN's capability in digit classification, it is primarily designed for isolated digit recognition, making it less adaptable to handwritten text within structured tabular formats, where segmentation errors and character misclassification are common. Conversely, our research emphasizes the digitization of handwritten marksheets, utilizing PaddleOCR and Modified YOLOv8 independently for text recognition, thereby achieving an optimal synergy between OCR-based text retrieval and deep learning-powered localization. While both studies focus on deep learning-driven recognition, the reviewed work applies CNN for isolated digit classification, whereas our method ensures context-aware recognition within structured documents. Unlike their approach, which relies on a single CNN model for classification, our hybrid pipeline separately leverages PaddleOCR and Modified YOLOv8 for precise text detection. While the reviewed paper highlights CNN's effectiveness in digit recognition, our research bridges the gap between isolated classification and real-world document automation, ensuring scalability, adaptability, and high-precision extraction of handwritten marksheets with 92.72% accuracy.

The paper author Paliwal and other researchers[17] propose TableNet, an end-to-end deep learning model that simultaneously performs table detection and structure recognition using a shared VGG-19 encoder with two separate decoders for segmenting tables and columns. This approach improves table boundary localization and column-wise segmentation while leveraging semantic rule-based row extraction for refined table structure identification. However, despite its efficiency in structured table extraction, TableNet is primarily designed for printed tables and struggles with recognizing handwritten text, limiting its adaptability to marksheets and handwrit-

ten documents. Our research builds upon the foundation of deep learning-driven text recognition but diverges in methodology by offering both PaddleOCR and modified YOLOv8 separately for text recognition. Initially, PaddleOCR is used to extract handwritten text, leveraging its robust OCR capabilities. We further integrate Modified YOLOv8 to enhance performance, ensuring better accuracy in complex handwriting scenarios. Unlike TableNet, which processes tables holistically, our approach separates text recognition from table segmentation and detects our tables rows and columns using OpenCV, allowing each model to specialize in its respective task. The result is a flexible, high-precision system that dynamically selects the best recognition model based on performance. While TableNet efficiently extracts tabular data from structured printed documents, our research extends beyond this by ensuring scalability for handwritten marksheets, combining OCR and deep learning-based detection to provide the best possible recognition accuracy.

Xie,Mao,Nan and other co authors [8] introduced an image processing-based approach that focuses on preprocessing, binarization, and edge detection to extract and recover tables from camera-captured images with complex backgrounds. By utilizing image rectification and mathematical morphology, their method successfully addresses issues like distorted tables and uneven lighting, making it particularly useful for scanned documents. However, the reliance on handcrafted rules and morphology-based segmentation poses a significant limitation—while it works well for printed tables, it struggles with handwritten variations and intricate table structures where text alignment is inconsistent. This is where our research takes a different path. Instead of depending on static feature extraction, we employ a deep learning-driven hybrid approach using OpenCV, PaddleOCR, and YOLOv8, allowing our model to learn table structures dynamically rather than relying on predefined rules. Unlike their method, which requires extensive image preprocessing to correct distortions, our model can adapt to different handwriting styles and table formats with minimal manual intervention and reducing segmentation errors. While their research offers valuable insights into reconstructing tables from challenging backgrounds, our approach goes beyond structured extraction, making marksheet digitization scalable, adaptive, and far more efficient for real-world applications. By integrating deep learning, we have transformed the process from rule-dependent to learning-driven, ensuring a robust, high-precision solution that overcomes the constraints of traditional OCR techniques.

Handwriting recognition has long been a field of interest, with earlier research, such as Beigi's "An Overview of Handwriting Recognition," [1] laying foundational methods for Optical Character Recognition (OCR) and signature verification. This paper primarily explored statistical models, neural networks alongside hidden Markov models (HMMs) for feature extraction and classification, emphasizing preprocessing techniques like slant correction, segmentation strategies, and language models to enhance recognition accuracy. The limitations of this method were due to its dependence on heuristic segmentation and old computational power limitations and also because deep learning transformations have advanced this field since its inception. The handwritten marksheet digitization system we developed uses a contemporary combination of OpenCV, PaddleOCR and YOLOv8 deep learning frameworks which addresses the constraints from previous methods. The approach improves detection

precision by enabling our models to accurately detecting text, thus addressing challenges faced by Beigi's method. Our method ensures high-speed marksheet processing along with improved accuracy of 92.72% while remaining adaptable thus making it suitable for large-scale digitization tasks. Our research solves the issues that traditional OCR could not handle particularly with unique writing styles combined with complex tables while also handling character detection errors to create an advanced solution with high scalability.

The author Honnegowda and others of this paper [15] discusses CNN, RNN, and a custom model called Deep-Writer to classify and recognize handwritten characters across multiple languages. Their methodology involves preprocessing, segmentation, feature extraction, and classification, where CNNs are primarily used for character recognition, and RNNs enhance sequence learning for structured handwriting patterns. While their approach efficiently recognizes individual characters, it faces limitations in handling text within complex tabular structures and struggles with segmentation errors in overlapping handwritten content. In contrast, our research extends beyond character recognition by addressing both text recognition and table structure extraction separately, utilizing PaddleOCR,YOLOv8,Modified YOLOv8 for handwriting text detection within marksheets. While both studies emphasize deep learning-driven recognition, the reviewed paper focuses on sequential text processing, whereas our method ensures context-aware recognition within structured documents. Unlike their approach, which relies on a combined CNN-RNN framework for learning character sequences, our work leverages specialized models—PaddleOCR for precise text extraction and YOLOv8 for accurate tabular segmentation resulting in improved recognition accuracy (92.72%). Although their research highlights an effective deep learning pipeline for handwriting classification, our study bridges the gap between isolated character recognition and real-world document automation, ensuring scalable, high-precision extraction of handwritten marksheets while reducing segmentation errors.

The Author Mupparaju and other researchers [25] provides a depiction of the development and features of YOLOv8. YOLOv8, the latest version in the YOLO family, represents a significant advancement in object detection which offers enhanced features such as greater accuracy, compact size, and exceptional speed, making it suitable for diverse hardware and cloud-based deployment, extending across various domains, including image classification and instance segmentation.The architecture of YOLOv8 consists of 2 components: head and backbone, both built on convolutional neural networks. The backbone utilizes an adapted CSPDarknet53, which improves information transfer via cross-stage partial linkages. The newly added C2f module enhances the system's capability to capture contextual and high-level characteristics. SPPF and convolution layers enhance feature processing across various scales. The head takes the output from the backbone to provide final predictions, including bounding boxes and classifications of objects. It utilizes a modular framework that enhances efficiency and precision by separately handling tasks such as classification and regression. YOLOv8 introduces several noteworthy architectural innovations Anchor-Free Detection, Modified Convolution Layers (C2f), Improved Backbone with CSPDarknet53 Adjustments, Detachable Head DesignDetachable Head Design, Optimized Feature Fusion. YOLOv8 introduces significant advance-

ments with its easy installation through a downloadable Python package via pip which simplifies integrating the framework into various environments. Also YOLOv8 CLI allows for fine-tuning, distributed training, and model evaluation. The Python Software Development Kit (SDK) simplifies integration with custom scripts, allowing developers to execute complex tasks like detection, segmentation, and classification with minimal effort. The tasks provided by YOLOv8 extend from object detection to segmentation along with classification and pose estimation and each operation has available pre-trained models. The precision capabilities of YOLOv8 stand out through meaningful results achieved in COCO and RF100 standard testing platforms. The COCO val2017 dataset shows high mAP scores because it evaluates detection accuracy across objects with diverse sizes and elaborate compositions. Its assessment on RF100, shows reliable accuracy across semantically distinct images, like satellite and microscopic data. The findings demonstrates YOLOv8's versatile performance in multiple applications while showing consistently strong adaptability according to varied settings. The speed performance of YOLOv8 exceeds its previous iterations under CPU and GPU execution conditions. Employing TensorRT on NVIDIA A100 GPUs, the model achieves minimal latency and significant throughput. Using FP16 data points enables additional acceleration of processing speed. The performance evaluation of YOLOv8 reveals superior ability across essential metrics which establishes it as the forefront object detection framework in the field.

Advancements in object detection have led to increasingly sophisticated deep learning models, each aiming to balance detection accuracy and computational efficiency. The author Talib and Co [29] have demonstrated YOLOv8-CAB, an improved version of YOLOv8 that integrates Context Attention Blocks (CAB) and modified Coarse-to-Fine (C2F) modules to enhance the detection of small objects. Using multi-scale feature extraction and spatial attention mechanisms, YOLOv8-CAB optimizes object detection, achieving a 97% mean average precision (mAP) on the COCO data set. However, despite its impressive improvements in object detection, this model is primarily optimized for real-time detection of small objects and does not explicitly address the challenges of handwritten text recognition, particularly within structured tabular formats. In contrast, our research emphasizes the digitization of handwritten marksheets, utilizing PaddleOCR and Modified YOLOv8 independently for text recognition, thereby achieving optimal synergy between OCR-based text retrieval and deep learning-powered localization. Although both studies focus on deep learning-driven feature extraction, the reviewed work applies YOLOv8-CAB for enhanced object detection, while our approach ensures context-aware recognition within structured documents. Unlike YOLOv8-CAB, which enhances small object detection using CAB and spatial attention modifications, our hybrid pipeline uses openCV for table detection and PaddleOCR,yolov8 and modified yolov8 separately for precise handwritten text detection, which leading to higher accuracy (92.72%). Although YOLOv8-CAB presents a breakthrough in object detection, our study extends its application to real-world document automation, bridging the gap between structured OCR-based recognition and deep learning-powered text extraction for handwritten marksheet digitization.

Gohen, Afshar[9] introduced EMNIST, an extension of the widely used MNIST dataset, which incorporates handwritten letters alongside digits to create a more

challenging classification task. By applying image preprocessing, binarization, and Gaussian filtering, the dataset converts characters from the NIST Special Database 19 into a format compatible with MNIST-based models, enabling direct integration with existing neural networks. This structured approach effectively standardizes handwritten character recognition; however, it is primarily designed for isolated character classification rather than contextual handwritten text within complex document structures, limiting its usability for real-world document digitization tasks. In contrast, our research addresses not only character recognition but also table structure detection and segmentation, integrating OpenCV for table detection, PaddleOCR, and Yolov8 modified yolov8 for text detection. While EMNIST provides a controlled dataset for training deep learning models, our work focuses on real-world handwritten marksheets, where text must be accurately extracted from varied table formats and complex handwriting styles. Unlike the reviewed study, which relies on fixed-sized, centered character images, our approach dynamically learns spatial dependencies between table cells and text in practical applications. By extending beyond character-level recognition, our research presents a scalable, adaptable solution that bridges the gap between traditional OCR datasets and real-world document automation, ensuring improved accuracy and contextual understanding in handwritten marksheet digitization.

Dufourq and Bassett [10] introduced EDEN in their paper which is a neuro-evolutionary approach that automates neural network architecture search using genetic algorithms (GA). By leveraging mutation and selection mechanisms, EDEN dynamically evolves convolutional, max pooling, and fully connected layers alongside hyperparameters like learning rate and embedding dimensions. This approach effectively reduces the reliance on manual tuning, making deep learning more accessible and computationally efficient. However, its dependency on evolutionary algorithms introduces a limitation—longer training times and computational overhead, as each generation requires extensive evaluation before converging to an optimal architecture. In contrast, our research adopts a deep learning-based hybrid method combining OpenCV, PaddleOCR, and YOLOv8, where model optimization is driven by structured learning rather than evolutionary randomness. While both studies prioritize automation and accuracy, EDEN focuses on model architecture discovery, whereas our research tackles handwritten text and table structure recognition in real-world marksheet digitization. Unlike EDEN, which evolves networks for general classification tasks, our approach is tailored for structured data extraction, allowing Modified YOLOv8 to adapt dynamically to complex table layouts and diverse handwriting styles, achieving 92.72% accuracy. While EDEN demonstrates an innovative step toward self-optimizing deep networks, our work presents a more task-specific, scalable solution that directly addresses the challenges of handwritten marksheet recognition, bridging the gap between OCR and intelligent table detection.

According to Peng and Yin [11], a hybrid method for effective image classification that combines Convolutional Neural Networks (CNNs) and Markov Random Fields (MRFs). To improve feature extraction, CNNs incorporate MRFs. MRFs are well-known for their capacity to model contextual constraints and spatial dependencies using local conditional probabilities. Using MRFs, the suggested architecture,

known as MRF-CNN, creates filters based on global image representations and texture patterns. These filters are then employed as prefixed filter banks in CNNs.This method effectively combines the statistical modeling strengths of MRFs with the hierarchical feature learning of CNNs. The MRF-CNN framework was evaluated on the MNIST, EMNIST, and CIFAR-10 datasets, showing significant performance improvements compared to state-of-the-art methods. For MNIST, MRF-CNN achieved an error rate of 0.38%, outperforming traditional CNNs and advanced pooling methods. Likewise, on EMNIST subsets, the model also had improved performance, obtaining an error rate of 8.75 percent on CIFAR-10, demonstrating its effectiveness with grayscale and color image datasets. By making use of prefixed MRF based filters with the CNNs, this study illustrates concepts that would lead to lower parameter complexity while not losing out on classification accuracy. This research hopes to further expand on these ideas by gathering more extensive datasets and adopting unsupervised learning with set filter banks.

The authors, Singh and Paul [12] have experimented a Deep Convolutional Neural Network (DCNN) implemented on CUDA, leveraging parallel computing with GPUs to accelerate digit recognition on the MNIST and EMNIST datasets. By utilizing multiple convolutional layers, max-pooling, and fully connected networks (FCNs), their approach significantly enhances accuracy (99.49% for MNIST, 99.62% for EMNIST) while reducing execution time through GPU-based parallelization. However, despite its efficiency, the model is primarily designed for isolated digit classification, limiting its applicability to real-world handwritten document recognition, where characters exist in complex tabular structures. In contrast, our research extends beyond character classification by integrating OpenCV for structural analysis, PaddleOCR, Yolov8, and modified YOLOv8 for text recognition separately, enabling the digitization of entire handwritten marksheets rather than just individual numbers. While both studies emphasize accuracy and computational efficiency, the reviewed work prioritizes parallel processing for speed, whereas our approach ensures context-aware recognition, handling variations in handwriting styles and table layouts. Unlike their static dataset-trained model, our Modified YOLOv8 dynamically learns spatial relationships within tables, achieving 92.72% accuracy in real-world handwritten text extraction. By addressing the limitations of standalone digit classification, our research introduces a scalable and application-driven approach that bridges deep learning advancements with practical handwritten document automation.

The authors Baldominos and Co[13] have used a hybrid neuroevolutionary approach, where genetic algorithms (GA) optimize the topology and hyperparameters of Convolutional Neural Networks (CNNs), while committees of multiple evolved models enhance recognition accuracy. By leveraging transfer learning, their method reduces the computational cost of neuroevolution by reusing optimized architectures across different datasets. This strategy effectively balances accuracy and efficiency but struggles with high computational overhead during the evolution process and lacks adaptability for handwritten text within structured tabular formats. In contrast, our research follows a task-specific deep learning-driven approach, integrating OpenCV for table structure identification, PaddleOCR, YOLOv8, and modified YOLOv8 for text detection separately, ensuring high accuracy for digitizing hand-

written marksheets. While both studies prioritize automation and accuracy, the reviewed work focuses on model evolution and ensemble learning, whereas our research emphasizes context-aware recognition within structured documents. Unlike their approach, which optimizes networks for character recognition, our Modified YOLOv8 model dynamically learns spatial dependencies within tables, significantly improving recognition accuracy (92.72%) and reducing segmentation errors. By addressing the limitations of standalone character recognition and high computational costs in neuroevolution, our study presents a scalable, real-world solution that bridges the gap between deep learning advancements and practical handwritten document digitization.

The author Cavalin and Oliveira [14] have experimented a confusion matrix-based hierarchical classification system in their paper which refines predictions by identifying frequently misclassified classes through transformed confusion matrices and Pearson correlation metrics. By constructing a hierarchical structure, their method improves character differentiation and recognition accuracy, particularly in datasets where certain classes are prone to confusion. However, this approach is limited by its dependency on predefined class relationships, making it difficult to generalize to unstructured handwritten text and complex table layouts. In contrast, our research tackles handwritten marksheet digitization through a deep learning-driven hybrid approach, where PaddleOCR and Modified YOLOv8 are used separately for text recognition. While both studies focus on minimizing classification errors, our methodology does not rely on fixed hierarchical refinement but instead ensures accurate text extraction at the detection stage itself. PaddleOCR, YoloV8, Modified Yolov8 specializes in recognizing handwritten text separately. Unlike their method, which refines misclassified cases post-prediction, our approach prevents misclassification before it occurs across diverse handwriting styles. While the reviewed study contributes a valuable hierarchical classification refinement strategy, our research extends its applicability by offering a scalable, real-world solution that effectively handles handwritten text within structured tabular documents, bridging the gap between traditional OCR refinement and modern document automation.

Handwritten character recognition remains a challenging task, particularly for languages with limited labeled datasets. The author Jayasundara and Co [16]presents TextCaps, a Capsule Network-based approach designed to train deep learning models with very small datasets by generating synthetic training samples. Through controlled perturbations in instantiation parameters, their approach introduces realistic handwriting variations, reducing the dependence on large datasets while achieving state-of-the-art accuracy on EMNIST and MNIST datasets. However, despite its innovative data augmentation strategy, TextCaps primarily focuses on isolated character classification, making it less suited for real-world handwritten document digitization, where text exists within structured formats like tables. In contrast, our research expands beyond character-level recognition by integrating PaddleOCR and Modified YOLOv8 separately for text recognition, ensuring both handwritten text extraction and accurate tabular segmentation in marksheets. While both studies emphasize accuracy and dataset efficiency, TextCaps aims to compensate for data scarcity using Capsule Networks, whereas our research leverages deep learning-based hybrid models to improve recognition within structured documents. Unlike

TextCaps, which excels in synthetic character augmentation, our approach processes complex document layouts dynamically, enabling context-aware extraction of handwritten text from structured marksheets with higher accuracy (92.72%). While the reviewed study presents a compelling solution for character recognition in data-limited settings, our work offers a more scalable and adaptable solution for practical handwritten document automation, bridging the gap between dataset-efficient learning and real-world text recognition challenges.

Pad, Narduzz and Co, [19] showcases a new strategy to improve the performance of convolutional neural networks (CNNs) by introducing a hybrid optical-digital implementation of Convolutional Neural Networks (CNNs), leveraging optical convolution layers to reduce computational overhead by encoding convolutional kernels directly onto an imaging aperture. By performing the first convolutional layer in the optical domain, their system significantly lowers energy consumption and processing latency, making it particularly effective for low-power real-time applications. However, despite its efficiency in reducing computational costs, this approach is limited in adaptability for complex handwritten text recognition, as it primarily focuses on image acquisition and feature extraction rather than dynamic learning from diverse handwriting styles. Conversely, our research emphasizes the digitization of handwritten marksheets, utilizing PaddleOCR and Modified YOLOv8 independently for text recognition, thereby achieving an optimal synergy between OCR-based text retrieval and deep learning-powered localization. While both studies emphasize accuracy and computational efficiency, the reviewed paper achieves this by offloading convolutional processing to an optical system, whereas our approach enhances recognition performance through an adaptive hybrid model, dynamically selecting between PaddleOCR as well as Modified Yolov8 for fine-grained character recognition. Unlike their system, which is designed for structured optical convolution, our method is more versatile for real-world applications in handwritten text recognition across varied document layouts. While their work presents a compelling approach to energy-efficient deep learning, our research extends its applicability by ensuring scalability and precision in complex document processing, bridging the gap between structured OCR and deep learning-based document automation.

Jeevan[22] has introduced WaveMix, a novel neural architecture that leverages a multi-scale 2D discrete wavelet transform (DWT) for spatial token mixing in vision tasks. Unlike Vision Transformers (ViTs) that require quadratic self-attention or CNNs that depend on deep networks for spatial feature extraction, WaveMix reduces computational overhead while achieving comparable performance. By using predefined multi-level DWT layers, the model enhances long-range spatial dependencies without self-attention, making it an efficient alternative for vision tasks. However, while WaveMix achieves state-of-the-art generalization in image classification, it primarily focuses on structured image processing and does not explicitly tackle handwritten text recognition within tabular documents. In contrast, our research emphasizes handwritten marksheet digitization, employing PaddleOCR and Modified YOLOv8 independently for text recognition to optimize performance. Both studies prioritize efficiency and accuracy, but while the reviewed work focuses on spatial token mixing for image classification, our method ensures robust text recog-

nition and structured table extraction. Unlike WaveMix, which relies on wavelet transforms for feature extraction, our dual-model approach integrates PaddleOCR for detailed character recognition and Modified YOLOv8 for text localization within complex table structures. While WaveMix presents a compelling resource-efficient alternative to ViTs and CNNs, our study bridges the gap between OCR-based handwritten text recognition and deep learning-driven document automation, ensuring scalability and precision in marksheet digitization.

The Author Yuning Du and Co [18] presents an OCR framework that delivers remarkable efficiency while maintaining a small model size—3.5M for Chinese characters and 2.8M for alphanumeric symbols. By focusing on computational efficiency and flexibility, PP-OCR is designed for resource-limited settings such as embedded devices. The architecture of the system combines a text detection model that employs Differentiable Binarization (DB) and a lightweight backbone, a classifier for determining text direction, and a text recognizer utilizing CRNN which extracts features and models sequences. This was influenced by the utilization of advanced methods such as MobileNetV3 backbones, data augmentation, cosine learning rate decay, and quantization techniques to reduce resource requirements while preserving accuracy.The practical usefulness of PP-OCR is clear from extensive experimental tests on multilingual datasets, showcasing its flexibility for various languages and uses, such as real-time scene text recognition. Their work demonstrates how they achieve minimal model size without compromising their identification and recognition capabilities. The system could gain additional advantages from improvements aimed at tackling issues in managing high-density or low-resolution text situations. In general, PP-OCR marks a considerable advancement in lightweight OCR technology, providing a scalable answer for situations requiring high precision and efficiency.

The Author Jaya Krishna Manipatruni and Co [23] gives detailed insights of leveraging the PaddleOCR framework to automate text extraction, especially in the cases of invoice formats. Through the combination of deep learning models and strong preprocessing methods, the study establishes an effective pipeline for extracting data, on the basis of 'structured' and 'unstructured'. Metrics like Character Error Rate (CER) and Word Error Rate (WER) have been utilized to confirm the approach which was vital in enhancing precision and operational effectiveness. The approach includes preprocessing steps, such as normalization and noise elimination. The practical element of the implementation is further improved by its connection with extensive database systems, guaranteeing dependable management and accessibility of financial data. The research also connects manual data management with automation. This presents considerable possibilities for scalable uses in various sectors. Although the study is well-conducted, insights into the difficulties encountered during implementation or a wider comparative analysis with leading OCR frameworks would enhance its practical significance. This paper demonstrates the transformative capabilities of PaddleOCR in automating invoice processing and also sets the stage for future advancements in document digitization, including the use of augmented reality for restoring degraded text. It is an admirable enhancement to AI-powered efficiency in managing financial data.

Chapter 3

Dataset Description

3.1 Description of the data

To develop a robust and efficient model capable of recognizing diverse handwriting styles, we created a custom dataset of handwritten digits. The dataset was meticulously assembled from various sources and methods to ensure it encompassed a wide range of variations. Initially, we collected images provided by our instructors as part of their materials.

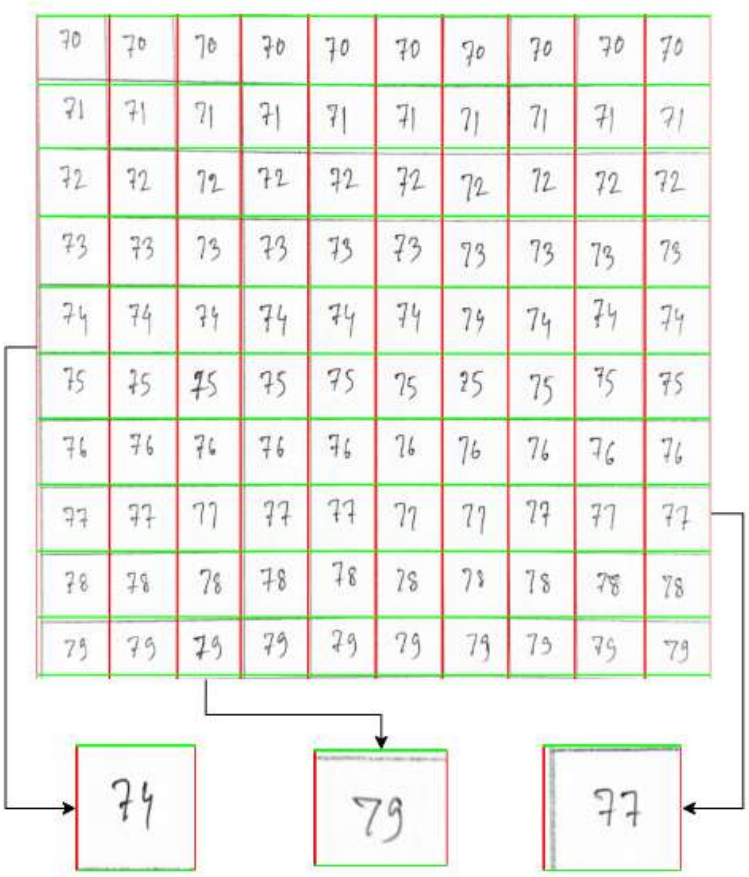


Figure 3.1: Handwritten Dataset And Grid Cells.

In addition, we printed lines both horizontal and vertical to form grids with rows and columns, and within these grids, handwritten entries were made. The dataset

incorporates various styles and tools, such as different colors and pencil inputs, to enhance diversity and representation. The dataset utilized for this work comprises a total of 10,000 images, carefully curated to ensure a diverse and representative sample. The figure 3.2 begins by converting the handwritten dataset into a grayscale image to simplify processing, followed by applying a blurring filter to reduce noise and enhance clarity. In figure 3.3, threshold was applied and then converted the image into a binary format, isolating text and grid lines. In figure 3.4, Vertical and horizontal lines are detected separately to identify the table structure, and these are combined to reconstruct the complete grid. The robustness of the dataset and also the generalisability of the dataset were improved through the use of data augmentation techniques.

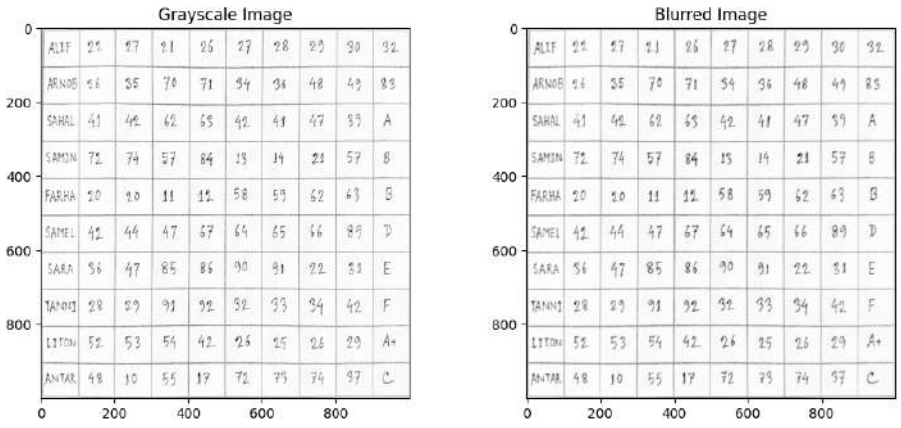


Figure 3.2: Visualization of Handwritten Dataset: Grayscale and Blurred Variants

Slight rotations in figure 3.6 of the images within a range of -15° to +15° of these techniques also simulate real world variations that occur in handwritten text. In addition to rotation, the dataset augmentation process includes flipping and cropping techniques to enhance model robustness and generalization. This step was important to ensure that the model could adapt to subtle distortions and inconsistencies present in real-world handwriting.

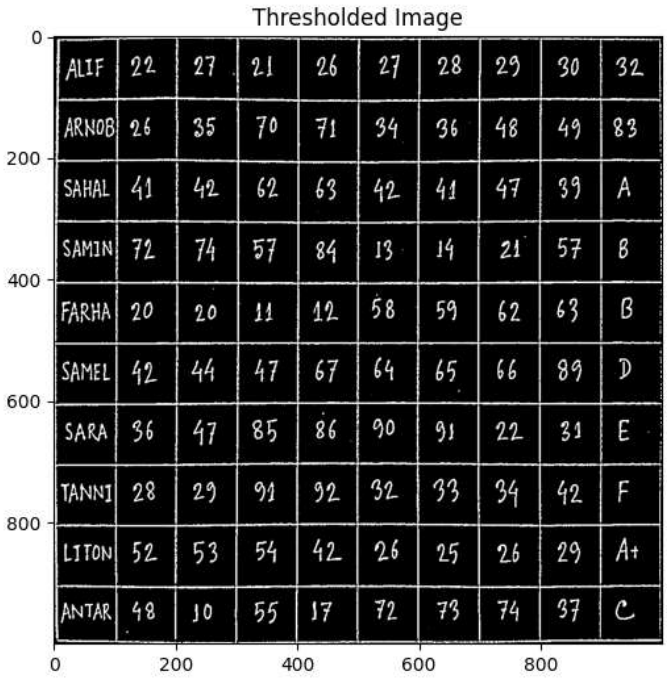


Figure 3.3: Visualization of Handwritten Dataset :Threshold Image

The dataset was then systematically divided into three subsets: 70% of the images were allocated for training, 20% for testing, and 10% for validation. The structured data split enabled us to successfully develop a comprehensive model training process alongside promising data for unbiased assessment and model refinement after which we incorporated an image table detection process.

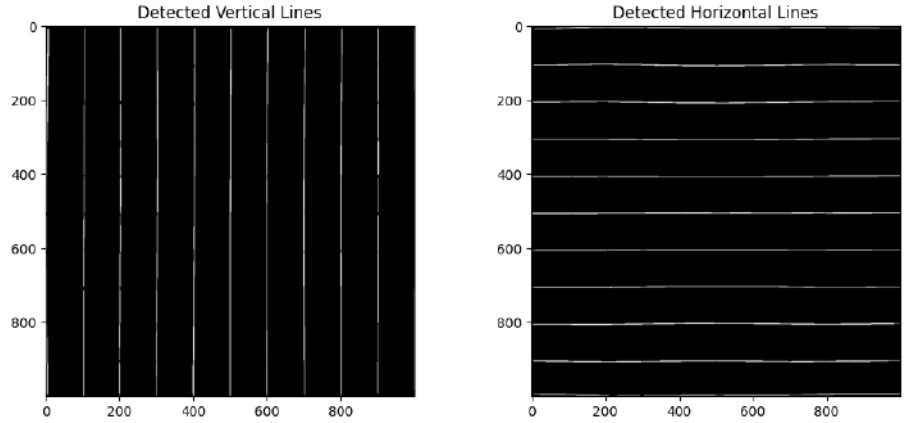


Figure 3.4: Visualization of Handwritten Dataset: Detected Lines

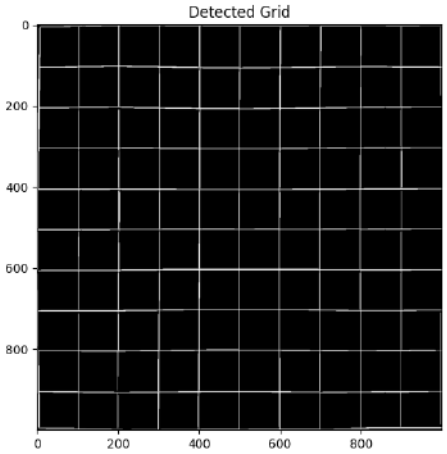


Figure 3.5: Visualization of Handwritten Dataset: Detected Grid

Figure 3.6: Visualization of Handwritten Dataset: Extracted Cell

We achieved better data point extraction by separating tables through individual cell divisions. The step facilitated efficient labeling and secured a more precise dataset preparation. The method we implemented produced a dataset fully prepared to train a model that demonstrates proficiency with different handwriting variations as well as writing instruments and document structures to enable successful real world implementations.

Chapter 4

Methodology

4.1 Working Process

A systematic process for extracting text and identifying data from tabular images begins by standardizing image format and noise reduction before detailed text recognition and data extraction operations. Verification of image compatibility takes place after image loading. When an incorrect file path or unsupported format leads to image loading failures the system displays an error message before termination. A standard image dimension of 1000×1000 pixels serves to maintain consistent visualization across all resolutions after image resizing begins. The standardization process reduces variations in grid dimensions that result from images of different dimensions. This step minimizes variability in grid dimensions caused by differing image sizes. After that the resized image is converted to grayscale and reducing the computational complexity by eliminating the need to process three color channels. To further enhance the image, a Gaussian blur with a kernel size of 5×5 is applied, effectively reducing high-frequency noise while preserving the structural features of the grid. To distinguish grid lines and table structures from the background, adaptive thresholding is applied to the blurred grayscale image. The block size is set to 15, which determines the size of the neighborhood used for threshold calculation, and a constant value of 8 is subtracted from the mean to fine-tune the threshold. This results in a binary image where grid lines are highlighted as white regions, and the background appears as black. Morphological operations are used to isolate vertical and horizontal grid lines separately. For vertical lines, a rectangular kernel of size 1×30 is employed, while for horizontal lines, a kernel of size 30×1 is used. Morphological opening is performed to remove noise and isolate elongated structures in the respective orientations, ensuring the integrity of grid lines. The vertical and horizontal line images are combined to create a unified grid structure using addition. This combined grid highlights the intersections of vertical and horizontal lines, marking potential cell boundaries. Contours are detected on the separated vertical and horizontal line images to determine column and row positions, respectively. For vertical lines, contours are identified to extract the x-coordinates of bounding rectangles, representing potential column boundaries. Similarly, the y-coordinates of bounding rectangles derived from contours on the horizontal line image represent row positions. To eliminate duplicate or closely spaced lines caused by imperfections in the grid, rows and columns are grouped based on proximity. A custom grouping logic considers a line unique if its position differs from the previous line

by more than 30 pixels. This ensures only distinct rows and columns are retained. The detected and grouped rows and columns are visualized on the original image for validation. Vertical lines marking columns are drawn in red, while horizontal lines representing rows are overlaid in green. This visual feedback confirms alignment with the grid structure in the input image. Individual cells are extracted and saved as separate images. The boundaries of each cell are defined using the coordinates of adjacent rows and columns. The current position selects the left corner coordinates, then the following position selects the bottom right corner coordinates. Each and every cropped region cell is saved with the filenames indexed by its row and column positions. Then each crop selection becomes an independent image file for next-step analysis. The extracted cell images are analyzed using a sophisticated optical character recognition (OCR) system, which deciphers the text content within each cell. The recognized text, along with its confidence scores, exists in a structured table configuration that maintains the original document structure. In text recognition, PaddleOcr operates as a crucial role. A preprocessing functionality applies binarization along with edge detection to improve text readability by enhancing visible text regions. A detection model like PPOCRDet identifies text areas within cell regions and a recognition model like PPOCRRec executes a text interpretation within defined regions. Text orientation accuracy is enhanced through angle classification when it is enabled to automatically correct detected text skew.

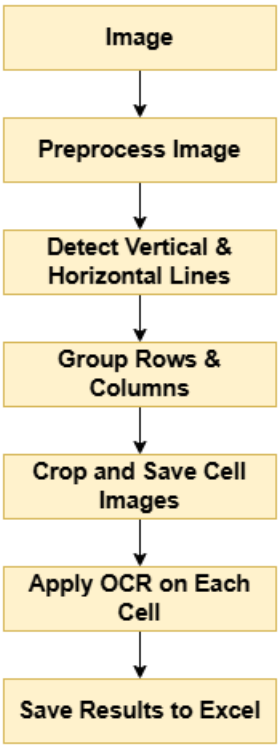


Figure 4.1: Entire work process

Finally, the extracted textual data is saved into an Excel spreadsheet.The workflow maintains accurate reconstruction of tabular structures through its capability to match recognized text with proper cell positions. By combining image preprocessing with structural analysis and OCR technology this comprehensive methodology generates precise structured digital data from tabular images suitable for analysis

or reporting purposes.

4.2 Description of Model

4.2.1 PaddleOcr

Optical Character Recognition (OCR) technology is a crucial tool for transforming textual data from images and scanned documents into machine-readable formats.OCR has replaced the hassle of manual transcription of the data by automating the text extraction process.As a result, it has drastically improved efficiency in various industries such as education,healthcare and finance. Various types of OCR frameworks have been developed over the years to enhance the accuracy and speed of text recognition in complex scenarios involving handwritten or distorted text. PaddleOCR is significantly the most promising OCR framework in this domain. It is a deep learning framework which is designed to precisely recognize both printed and handwritten text. It has opened the doors for various applications such as document digitization, automated form processing and specialized tasks such as recognizing handwritten marksheets which is the primary focus of our research.PaddleOCR's modular architecture divides the process into several distinct stages like text detection,text recognition and post processing. The architecture of PaddleOCR is designed with a hybrid structure which contains blackboard and pipe-and-filter design patterns.This unique structure guarantees modularity, scalability and adaptability and also enhances the reusability and maintainability of its components.The figure below illustrates how PaddleOCR leverages this hybrid architectural style by effectively processing input images through sequential stages in the pipe-and-filter model by transforming raw visual data into structured textual representations. PaddleOCR employs a modular pipeline for recognizing textual data in images where is focuses on sequentially transforming raw inputs into structured textual outputs. The architecture combines blackboard design patterns with pipe-and-filter patterns for modular operation and efficient scaling because of its hybrid approach. The text recognition process in PaddleOCR is structured into separate stages that work well together to achieve high accuracy handling different application scenarios including document digitization and form processing along with handwritten marksheet recognition. High-quality outcomes result from the three crucial stages of the pipeline which include text detection followed by direction classification and completing with text recognition.The system starts by detecting areas with text content through text detection in order to identify regions of interest (ROIs) in the input images. Identifying text areas is vital in the detection process when working with complex or irregular document layouts. PaddleOCR selects Differentiable Binarization (DB) to function as its core detection algorithm. The neural network engine of DB goes through input images to create probabilistic maps that indicate possible text areas. The text location area visibility is enhanced through thresholding processes that convert probability maps into specific binary versions indicating text positions. The text detection procedure of the DB algorithm starts by compressing detected areas into their minimal volume during training followed by boundary reconstruction at inference time. The exact positioning of elements becomes achievable through this approach which also minimizes location mistakes from overlapping or unclear areas. The detected text areas get enclosed within rectangular boxes to maintain

the original document structure.Once the text regions are detected, the next stage is direction classification, which ensures that the detected text is correctly oriented for further processing. Misaligned or rotated text blocks are common in scanned documents and photographs. PaddleOCR addresses this challenge by leveraging a Convolutional Neural Network (CNN) trained to categorize the orientation of text regions. The CNN evaluates the extracted features and determines whether the text needs to be rotated or transformed. If necessary, suitable adjustments are applied to align all text horizontally, facilitating seamless progression to the recognition stage. This ensures that even complex layouts or skewed inputs are accurately processed. After text regions get identified the system proceeds to direction classification to properly reorient detected text before processing. PaddleOCR resolves orientation detection problems through its Convolutional Neural Network (CNN) that identifies text region positioning. The CNN examines the obtained features through analysis to determine if the text needs rotation or transformation. The recognition process requires all text to be aligned horizontally which can be achieved by suitable adjustments to optimize the reading flow. The system maintains precise processing of complex layouts together with skewed inputs while doing so.

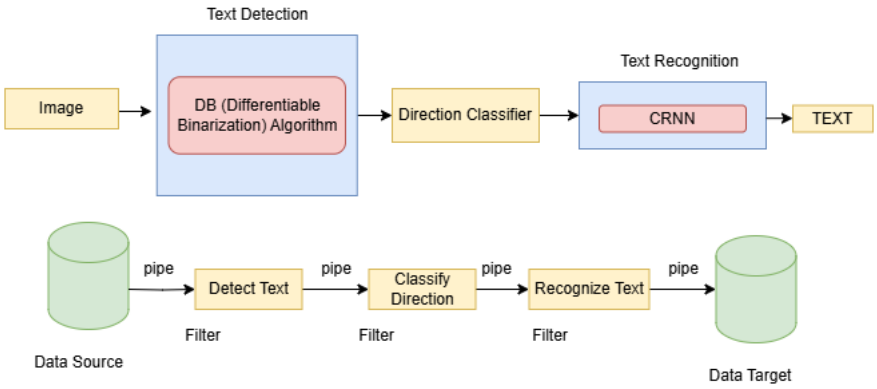


Figure 4.2: PaddleOCR Architecture

The design of PaddleOCR enables this system to work as a flexible tool which suits both standardized document types including invoices and marksheets alongside intricate documents like receipts and handwritten forms. The pipeline uses designed stages which refine outputs while maintaining results accuracy and precision. The hybrid architecture lets the system combine additional components as well as algorithms to achieve speed and accuracy balance according to particular requirements. PaddleOCR provides users with versatility that enables successful handling of various text recognition use cases.

4.2.2 YOLOV8

You Only Look Once or YOLO is a very popular family of object detection models, and the latest of them is the 8th version called YOLOv8. Although it has been used in Australian Micro Electric Vehicle mimics, Intelligent Health, Urbvan, and others; autonomous driving, medical imaging, security and surveillance, YOLOv8 is now being used for OCR by us and particularly for handwriting. This mainly includes to able to identify and accurate characters or specific words in a picture in order to

get textual data in images. Handwritten OCR is inherently difficult because many ways to write text, letter alignment is not always perfect and text scanned or photographed is often distorted in some way. Unlike many traditional OCR systems, YOLOv8 is capable of surmounting such challenges thanks to its structures, such as the absence of anchor boxes. This design allows the model to identify low contrast small, cursive, irregular scripts within handwriting, which itself might be misaligned or contain uneven letter spacing. As one of the YOLO modes, YOLOv8 also boast of a highly efficient training pipeline for transfer learning. This also makes the model to be trained using a number of sets such as marksheets or forms to solve problem that involves different handwriting traditions and styles. For multiple and diverse objects, YOLOv8 holds a high success rate under various conditions since it was fine-tuned in the detection of a specific dataset. Furthermore, real-time data processing means that the time taken to identify an abnormality is highly desirable for high-throughput applications. By identifying these particular benefits, YOLOv8 is set to revolutionize OCR, especially for areas such as form or document scanning, digitization of handwriting, and identification of marks sheets. Its speed, precision, and adaptability make it an indispensable component of the new text recognition systems that significantly change the methods for detecting and digitizing handwritten text.The YOLOv8 architecture comprises three core components which are the Backbone, the Neck, and the Head. These elements contribute to perform successful and precise object detection which makes YOLOv8 suitable to detect the handwritten text in emerging datasets.CSP Darknet53 is the baseline of YOLOv8 with convolutional layers and Cross-Stage Partial (CSP) connections. These connections enhance the reusability of features and instead of evaluating each image pixel by pixel, which is computationally intensive, the architecture can process large data sets. The use of the C2f module in YOLOv8 comes with a new improvement in feature fusion because it incorporates both contextual and semantic data. Also, the Spatial Pyramid Pooling Faster (SPPF) feature extraction method is performed to obtain these features at more than one scale, hence making it easier to detect handwritten characters of different sizes and from different sources such as faxes. These innovations make the Backbone ideal for extracting the basic structure from patterned images of handwritten text.The Neck in YOLOv8 connects the Backbone and the Head to enhance and fuse multi-scale features. With Path Aggregation Network (PANet), it effectively combines low-level and high-level features for the detection of objects of disparate sizes. This was particularly important to aid in the recognition of handwritten text characters where size and orientation differ greatly. The C2f which often forms a part of the Neck strengthens feature integration from multiple scales additional to what has been expounded above.

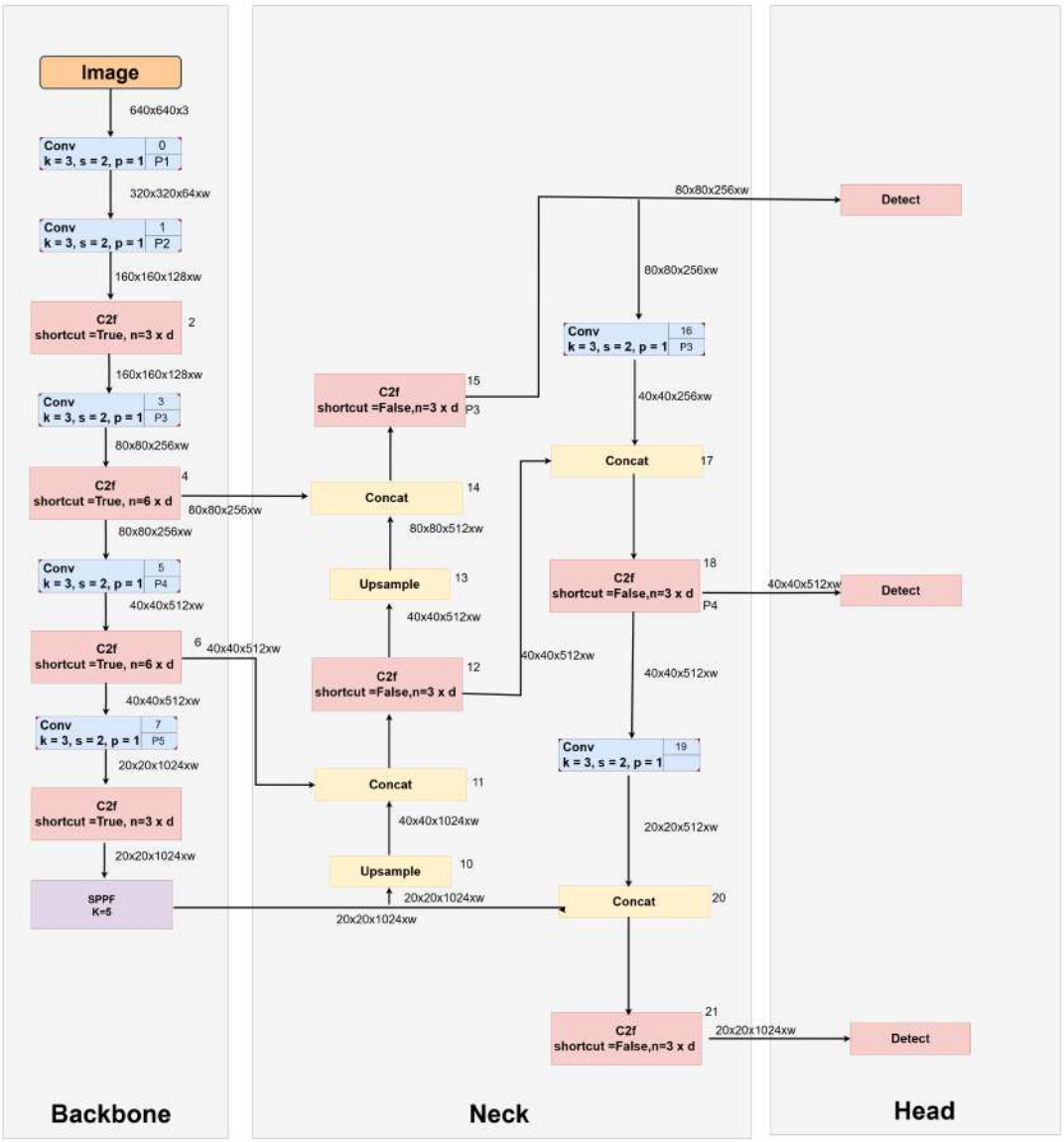


Figure 4.3: YoloV8 Architecture

This multi-scale fusion is able to provide the contextual features which are important in complicated handwriting samples, thereby guaranteeing a good text detection on special datasets.The Head generates the output of the detection results of the format of the rectangular coordinate value, the predicted object category and corresponding detection probability. The Head of YOLOv8 is another that lacks the concept of anchors, which are three predefined ground-truth boxes at different scales for training. Rather than doing this, it directly predicts the center, width and height of the bounding boxes, which makes the model simpler and somewhat de-tuned. This design is particularly useful in handwriting OCR scenarios as it means the shapes and layout of the text can widely differ. The U layers (Upsample layers) expand fields of view sizes to contain more detail features for better localizing the small handwritten characters. Due to the possibility of modular design, The Head is responsible for such specific operations such as object classification and localization, which is important for attaining high accuracy in text recognition. Since we train YOLOv8 with a custom handwritten dataset in various styles, its Head performs well in different layouts for the OCR of handwritten texts.YOLOv8 is inherently

designed for real-time object detection tasks, utilizing a deep and complex architecture to process diverse, high-resolution images and detect multiple objects with high precision. However, In case of handwriting recognition which actually focuses on classifying smaller, simpler, and more structured patterns where most of this complexity becomes redundant. Layers such as the detection layers which actually are responsible for generating bounding boxes and confidence scores as well as the neck layers like PANet which used for aggregating multiscale features, and several deeper backbone layers (designed for high-level feature extraction in complex scenes) are unnecessary for this task.

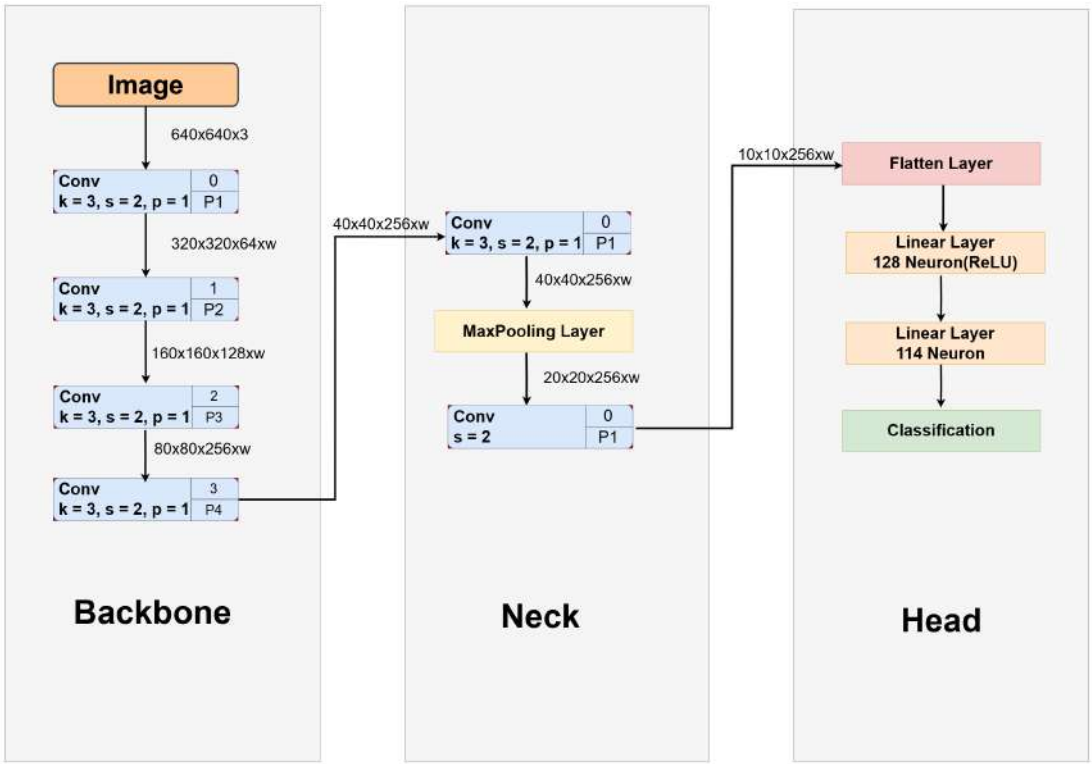


Figure 4.4: YOLOV8 Modified Architecture

Handwriting recognition typically requires only shallow and midlevel feature extraction for classification, making these components superfluous. Simplifying the architecture alongside layer removal produces substantial speed improvements while lowering resource needs because it adapts perfectly to structured handwriting patterns. The designed architecture accommodates desired performance goals while eliminating unnecessary excess components.The YOLOv8 architecture features edits for handwritten digit recognition which eliminate redundant layers while keeping core capabilities for feature extraction and classification. The backbone layers operate as convolutional networks which start by finding edge details before advancing to the detection of shapes and digit patterns. The neck portion consists of maxpooling and an extra convolutional layer with a stride value of 2 to perform efficient downsampling and retain important digit characteristics. The extracted features go through the head which flattens them into a vector before the fully connected layers activate with ReLU to improve learning potential. The model generates classification logits which correspond to digit categories in its last output layer. The modified

approach removes three complex layers C2f (bottleneck with shortcuts), SPPF (Spatial Pyramid Pooling Fast), and several upsample-concat operations from the neck section because they were designed for multi-scale object detection but not single-digit classification. Cutting these layers results in decreased computational difficulty and accelerated inference while making the model suitable for constrained resource settings. The redesigned YOLOv8 operates differently from its general object detection counterpart since the modifications prioritize essential feature extraction across multiple levels without adding excess depth. The model provides superior performance than YOLOv8 because it optimizes digit recognition speed and precision without running unnecessary procedures.

Chapter 5

Result Analysis

Our workflow is segmented into multiple stages, beginning with the detection of rows and columns, followed by the identification of individual cells. After that each and every cell is extracted as an image containing handwritten data for further processing. The complex multi-step process requires preliminary image recognition tests using different models against publicly accessible datasets before ultimately utilizing the final training and testing on our dataset. Such evaluation helps organizations find their most appropriate OCR model that matches their operational needs while delivering maximum accuracy combined with efficiency. The comparative performance of different OCR methods is summarized in Table 5.1.

Method	**Accuracy (%)**
OPIUM [9]	95.90
EDEN [10]	99.30
OptConv+Log+Perc [19]	99.43
CNN [14]	99.46
Parallelized CNN [12]	99.62
Neuro Evolved CNN [13]	99.73
Markov Random Field CNN [11]	99.75
WaveMixLite-112/16 [22]	99.77
TextCaps [16]	99.79
PaddleOCR	99.80
YOLOv8	99.40
Modified YOLOv8	99.68

Table 5.1: Comparison of Optical Character Recognition (OCR) Accuracy Across Methods on EMNIST-digit dataset

In the table of 5.1, we compared our approach with existing models to highlight its strengths in real-world scenarios. The OPIUM model [9] and CUDA-accelerated DCNN [12] excel in general handwritten character recognition on datasets like MNIST and EMNIST but fail to handle complex, structured layouts and real-world OCR challenges. The EDEN model [10] implements evolutionary algorithms for CNN topology optimization yet its high computational requirements combined with lack of domain-specific features inhibit practical OCR usage. WaveMix [22] introduces wavelet-based token mixing, achieving computational efficiency in general

image classification, but it lacks the adaptability required for tasks like handwritten marksheet recognition. The TextCaps [16] system demonstrates high performance on small datasets when using Capsule Networks yet runs into scalability problems because dynamic routing leads to increased computational requirements. Our approach for PaddleOCR is specifically designed to address challenges such different and diverse handwriting styles as well as low contrast text and overlapping regions. PaddleOCR's modularity ensures high accuracy for structured data and for the Modified YOLOv8 offers lightweight precision suitable for large-scale, real-time tasks. The individual models present a unified solution which demonstrates better performance than other models while being adaptable and efficient for practical OCR applications. Unlike other models such as TextCaps or neuroevolutionary ensembles [13], which prioritize theoretical performance, our approach bridges the gap by offering a scalable and resource efficient solution for practical implementation.

This comprehensive analysis evaluates the effectiveness of an Optical Character Recognition (OCR) system that combines OpenCV for cell detection and PaddleOCR for the recognition of handwritten text. This analysis used four images (testimg1 to testimg4) with handwritten information to determine system accuracy while detecting recurring mistakes and identifying optimization opportunities. By systematically analyzing correct detections, incorrect classifications, and missing data points, this study provides a thorough investigation of the system's strengths and limitations. The analysis examines error patterns alongside performance metrics to understand the system challenges regarding different handwriting types within various clarity ranges. The findings aim to guide future improvements in the OCR system which will enhance its robustness and reliability in real world applications.

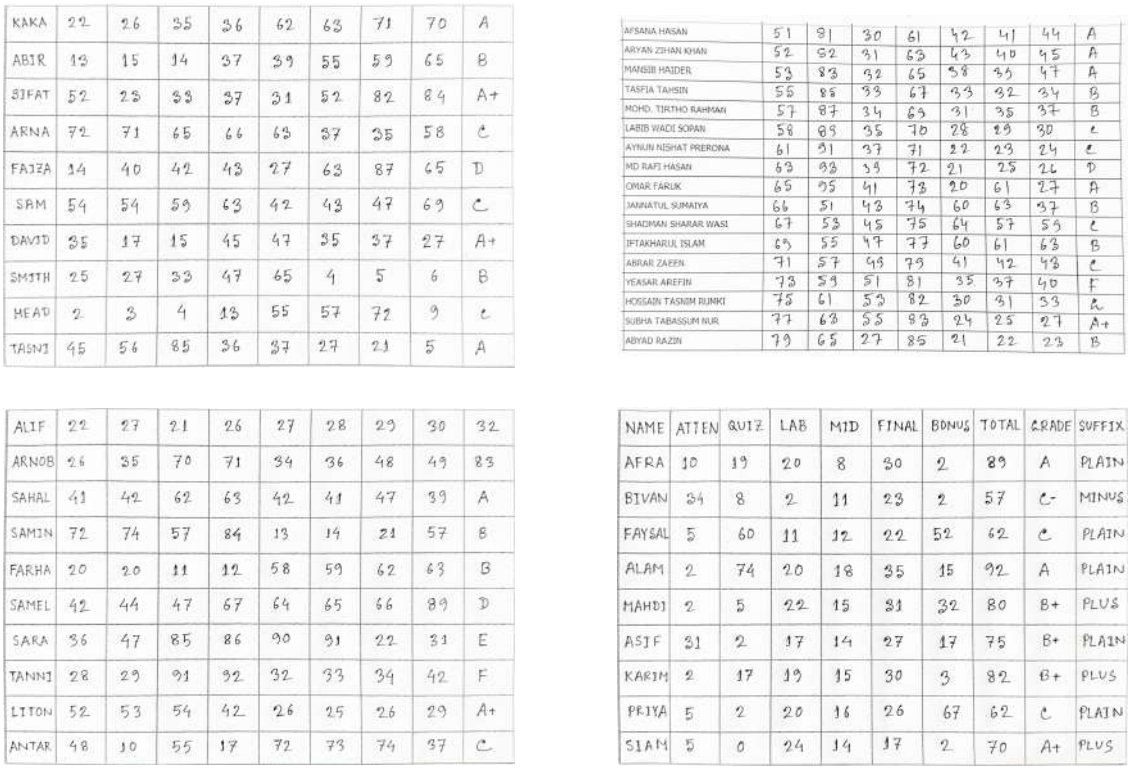


Figure 5.1: Handwritten test images for OCR analysis.

We have selected four representative test cases from our test dataset which comprises 20% of the total dataset to analyze the performance of OCR systems using PaddleOCR,YOLOv8 and modified YOLOv8. The results, along with detailed performance metrics and observations, are presented below, highlighting the systems' strengths, limitations, and overall effectiveness.

Model	**Metric**	**Test1**	**Test2**	**Test3**	**Test4**	**Average**
PaddleOCR	Total Correct Cells	89	123	98	91	100.25
	Total Incorrect Cells	3	17	1	4	6.25
	Total Missing Cells	8	13	1	5	6.75
	Precision (%)	96.74	87.86	98.99	95.79	94.35
	Recall (%)	91.75	90.44	98.99	94.79	93.99
	F1 Score (%)	94.18	89.13	98.99	95.29	94.90
	Accuracy (%)	91.75	88.49	97.99	93.81	93.01
YOLOv8	Total Correct Cells	90	120	95	88	98.25
	Total Incorrect Cells	4	15	3	6	7.00
	Total Missed Cells	6	18	2	6	8.00
	Precision (%)	95.74	88.89	94.96	91.67	92.32
	Recall (%)	93.75	90.91	95.00	90.72	92.10
	F1 Score (%)	94.74	89.89	94.98	91.19	92.70
	Accuracy (%)	93.75	89.85	94.98	90.57	92.29
Modified YOLOv8	Total Correct Cells	91	122	97	90	100.00
	Total Incorrect Cells	3	14	2	6	6.25
	Total Missed Cells	6	17	1	4	7.00
	Precision (%)	96.80	89.70	96.06	92.59	93.79
	Recall (%)	94.84	91.76	94.85	92.31	93.44
	F1 Score (%)	95.81	90.72	95.45	92.45	93.61
	Accuracy (%)	94.84	90.97	95.45	92.31	93.39

Table 5.2: Performance Metrics Across Test Images for PaddleOCR, YOLOv8, and Modified YOLOv8

PaddleOCR demonstrated high accuracy in detecting numeric and simple alphanumeric text, especially when characters were clear and well-separated. Examples include "KAKA" and "Abir," where test images like testimg1 and testimg3 had clearly defined cell boundaries. On the other hand, YOLOv8 excelled in identifying and localizing text regions through its bounding boxes, effectively handling handwriting styles and sizes in the same test images. Both systems, however, faced challenges in certain scenarios. PaddleOCR struggled with the misclassification of visually similar characters, such as "B" being mistaken for "8" or "51" as "5," particularly in testimg2. These errors were often caused by faint strokes or incomplete handwriting. In contrast, YOLOv8 encountered issues with bounding box accuracy, such as missing or incorrectly sized boxes for faint text, as observed in testimg4. These inaccuracies affected overall text detection. Another limitation of PaddleOCR was its inability to detect faint handwriting or edge-case cells, such as missing "C" in testimg1 and "44" in testimg2. These examples highlight the need for better preprocessing and boundary detection. YOLOv8 revealed vulnerabilities by showing clear detection limitations of low-contrast text areas and its dependence on additional processing steps for bounding box refinement. While PaddleOCR showed high precision in character recognition, its performance was highly dependent on clear cell boundaries. Text region localization proved robust with YOLOv8 while

the system needed adjustments to handle faint and low-contrast text instances. Table 5.2 represents a comprehensive analysis of the performance for the models of PaddleOCR, YOLOv8, and Modified YOLOv8 across four test datasets. Modified YOLOv8 outperformed the other models, achieving the highest average for total correct detections (100.00) for the respective cells and F1 Score (93.61%) which indicate its robustness in balancing precision and recall. With 93.79% precision and 93.44% recall the method established the best performance level for reducing false-positive errors without compromising detection accuracy.. Although PaddleOCR recorded the highest recall (93.99%), its precision (94.35%) and F1 Score (94.90%) were slightly lower than those of Modified YOLOv8. YOLOv8 performed well but was outpaced by its modified version, particularly in reducing incorrect cell detections and improving accuracy. Overall, Modified YOLOv8 demonstrated the most balanced and reliable performance, making it the most effective model for applications requiring high accuracy and minimal errors. The proposed model achieves high accuracy on our custom dataset, which has been meticulously designed to encompass diverse handwriting styles, varying table layouts, and challenging cases such as overlapping text and irregular cell spacing. This dataset serves as a reliable benchmark for evaluating the performance of OCR and text recognition systems in practical scenarios. Experimental results demonstrate that YOLOv8Modified achieves an accuracy of 92.72%, outperforming PaddleOCR (91.37%) and the baseline YOLOv8 model (88.91%). To provide a comprehensive analysis of the system's performance, we evaluated key metrics, including Precision, Recall, and F1-Score, for each model. As shown in Table 5.3, YOLOv8Modified consistently outperforms the other models across all metrics, achieving the highest Precision of 94.25% and an F1-Score of 93.52%, while maintaining a Recall of 92.80%. PaddleOCR follows with a balanced performance, achieving an F1-Score of 92.32% and a Recall of 91.50%, while the baseline YOLOv8 model demonstrates reasonable results with an F1-Score of 89.94% and a Recall of 89.10%. These findings highlight the robustness and efficiency of the YOLOv8Modified model, particularly in handling challenging cases present in the dataset.The results underscore the practicality of the proposed hybrid approach for real-world applications, as the combination of OpenCV, PaddleOCR, and YOLOv8-based models significantly reduces the need for manual intervention while delivering reliable and scalable solutions for digitizing academic and administrative documents.

Model	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-Score (%)**
PaddleOCR	91.37	93.15	91.50	92.32
YOLOv8	88.91	90.80	89.10	89.94
Modified YOLOv8	92.72	94.25	92.80	93.52

Table 5.3: Performance Metrics of PaddleOCR, YOLOv8 and Modified YOLOv8

The table 5.4 compares PaddleOCR, YOLOv8, and Modified YOLOv8 based on WER (Word Error Rate) and CER (Character Error Rate). PaddleOCR demonstrates good performance with an average CER of 8.54% and WER of 10.22%, excelling in character recognition particularly in Test3. YOLOv8 shows slightly higher errors, with an average CER of 11.00% and WER of 12.12% which indicate the room for improvement in both character and word level recognition. Modified YOLOv8 outperforms both of them and achieves the lowest average CER (7.75%)

Model	Metric	Test1	Test2	Test3	Test4	Average
PaddleOCR	CER (%)	10.00	14.74	2.00	7.40	8.54
	WER (%)	10.50	19.39	2.00	9.00	10.22
YOLOv8	CER (%)	12.00	16.50	5.00	10.50	11.00
	WER (%)	12.50	18.50	6.00	11.50	12.12
Modified YOLOv8	CER (%)	9.00	12.50	2.50	7.00	7.75
	WER (%)	9.50	17.00	2.20	8.50	9.30

Table 5.4: Performance Metrics WER and CER for PaddleOCR, YOLOv8, and Modified YOLOv8

and WER (9.30%), highlighting the effectiveness of its enhancements. This model provides consistent and accurate results. Overall, Modified YOLOv8 is the most reliable for precise OCR tasks, while PaddleOCR remains competitive for character recognition.

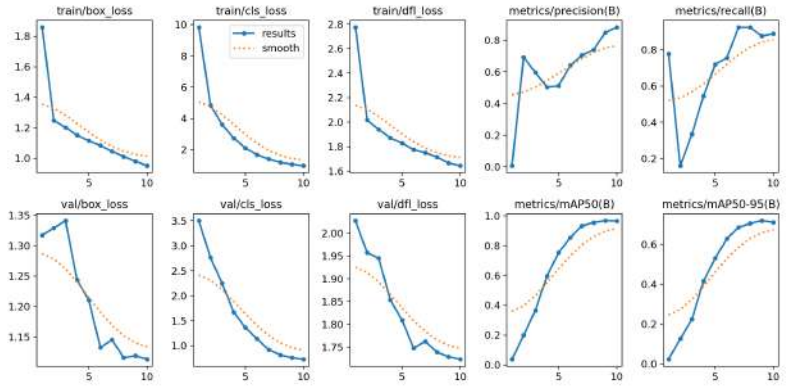


Figure 5.2: Paddleocr Training and Validation Metrics

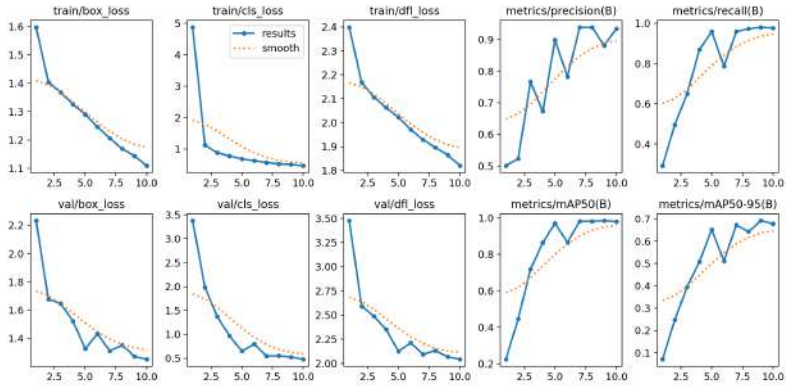


Figure 5.3: YOLOv8 Model Training and Validation Metrics

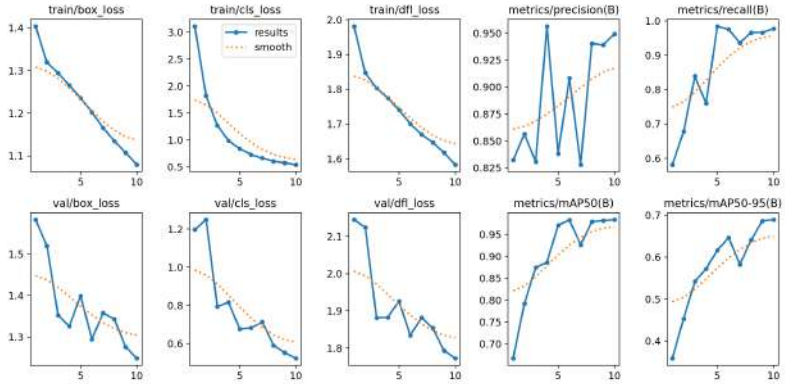


Figure 5.4: YOLOv8 Custom Model Training and Validation Metrics

The PaddleOCR confusion matrix 5.5 reveals high prediction accuracy because most predicted results gather along the main diagonal. The model displays efficient differentiation between character classes because off-diagonal values remain scarce and minimal. The system successfully detects texts with high precision when used across different text input conditions. The YOLOv8 model 5.6 exhibits excellent performance shown by

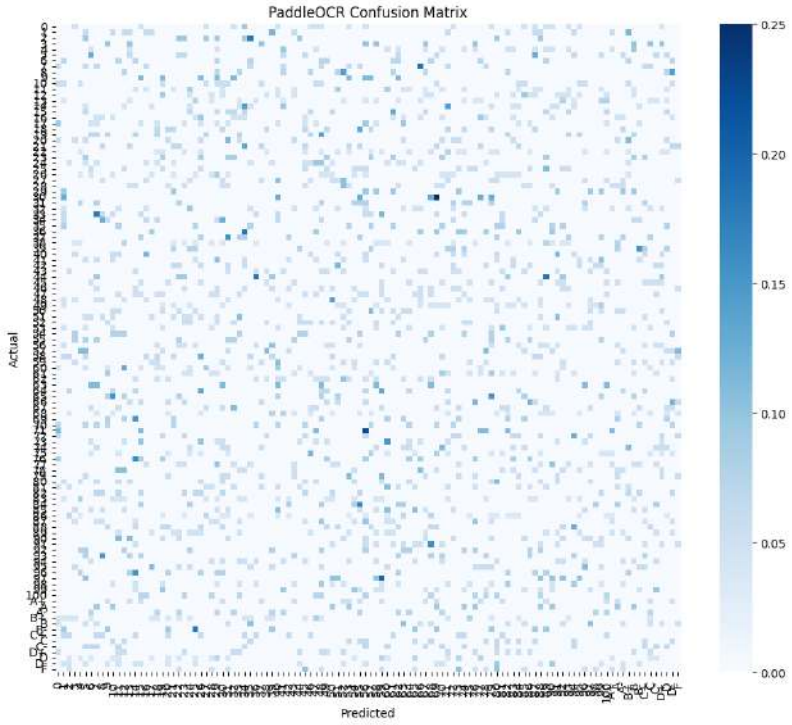


Figure 5.5: Confusion Matrix of PaddleOcr

its density of accurate predictions which appears along the diagonal axis. Few off-diagonal elements demonstrate that the model performs minimal workplace classification errors and indicates its ability to detect objects with accuracy. YOLOv8 demonstrates excellent effectiveness when used for various detection applications.

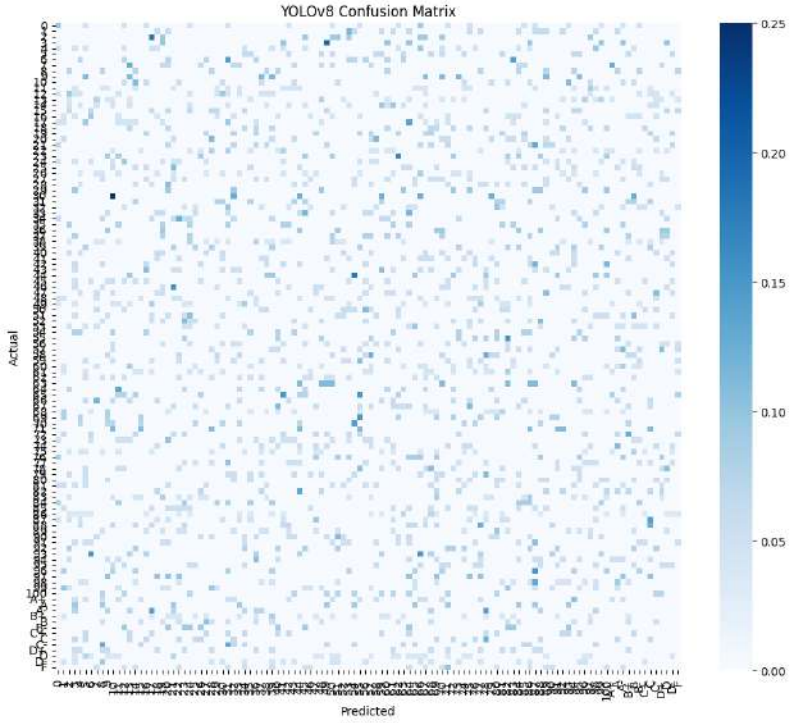


Figure 5.6: Confusion Matrix of Yolov8

A noticeable enhancement appears in the modified YOLOv8 confusion matrix 5.7 because the diagonal area displays an even stronger concentration which symbolizes improved prediction accuracy. The modifications successfully reduced off-diagonal elements in the confusion matrix because they minimized model misclassifications which indicates better detection accuracy and system reliability.

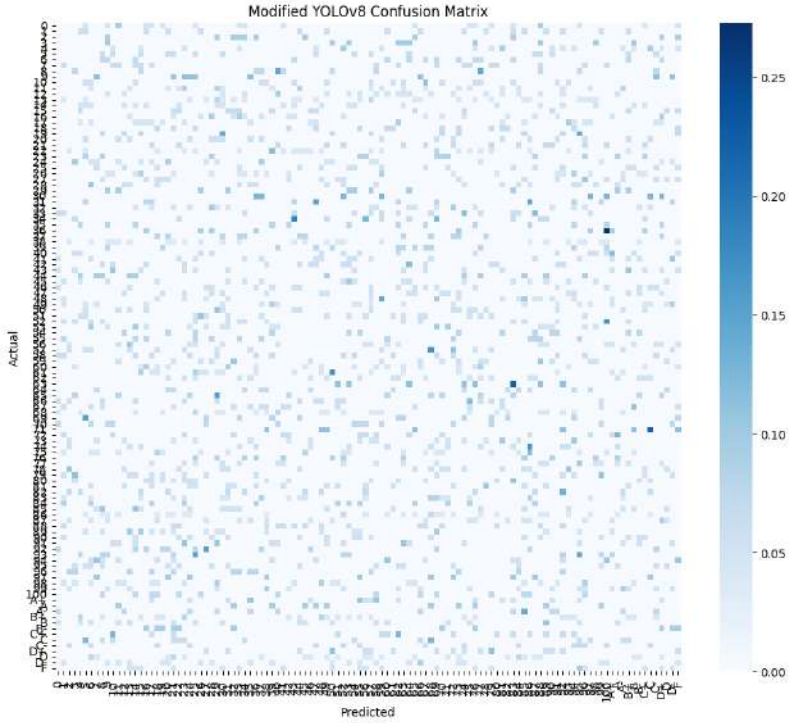
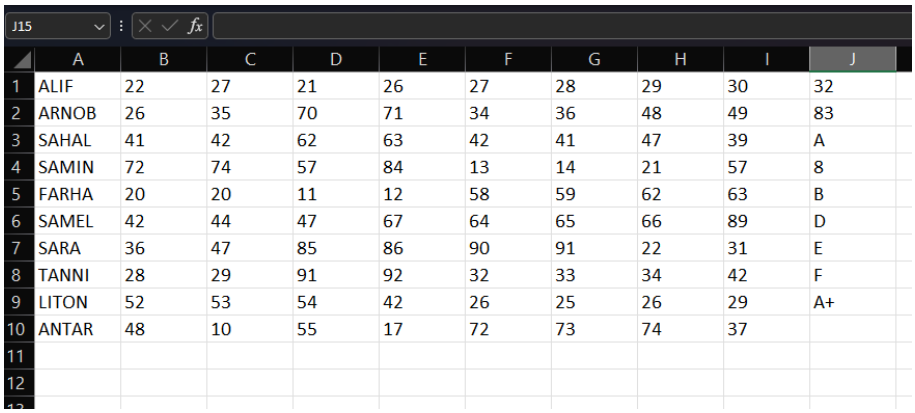
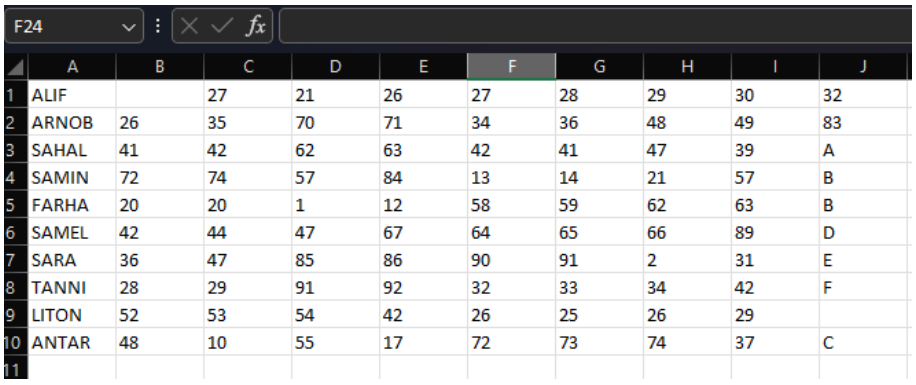


Figure 5.7: Confusion Matrix of Modified Yolov8



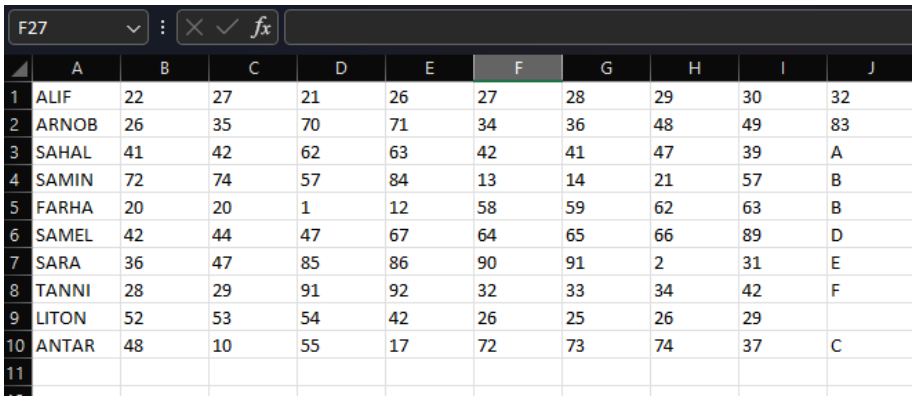
	A	B	C	D	E	F	G	H	I	J
1	ALIF	22	27	21	26	27	28	29	30	32
2	ARNOB	26	35	70	71	34	36	48	49	83
3	SAHAL	41	42	62	63	42	41	47	39	A
4	SAMIN	72	74	57	84	13	14	21	57	8
5	FARHA	20	20	11	12	58	59	62	63	B
6	SAMEL	42	44	47	67	64	65	66	89	D
7	SARA	36	47	85	86	90	91	22	31	E
8	TANNI	28	29	91	92	32	33	34	42	F
9	LITON	52	53	54	42	26	25	26	29	A+
10	ANTAR	48	10	55	17	72	73	74	37	
11										
12										

(a) PaddleOCR Results



	A	B	C	D	E	F	G	H	I	J
1	ALIF		27	21	26	27	28	29	30	32
2	ARNOB	26	35	70	71	34	36	48	49	83
3	SAHAL	41	42	62	63	42	41	47	39	A
4	SAMIN	72	74	57	84	13	14	21	57	B
5	FARHA	20	20	1	12	58	59	62	63	B
6	SAMEL	42	44	47	67	64	65	66	89	D
7	SARA	36	47	85	86	90	91	2	31	E
8	TANNI	28	29	91	92	32	33	34	42	F
9	LITON	52	53	54	42	26	25	26	29	
10	ANTAR	48	10	55	17	72	73	74	37	C
11										

(b) YOLOv8 Results



	A	B	C	D	E	F	G	H	I	J
1	ALIF	22	27	21	26	27	28	29	30	32
2	ARNOB	26	35	70	71	34	36	48	49	83
3	SAHAL	41	42	62	63	42	41	47	39	A
4	SAMIN	72	74	57	84	13	14	21	57	B
5	FARHA	20	20	1	12	58	59	62	63	B
6	SAMEL	42	44	47	67	64	65	66	89	D
7	SARA	36	47	85	86	90	91	2	31	E
8	TANNI	28	29	91	92	32	33	34	42	F
9	LITON	52	53	54	42	26	25	26	29	
10	ANTAR	48	10	55	17	72	73	74	37	C
11										

(c) Modified YOLOv8 Results

Figure 5.8: Comparison of Excel Outputs: (a) PaddleOCR, (b) YOLOv8, and (c) Modified YOLOv8 Results.

In this table 5.5, We compare the performance of PaddleOCR with an open-source OCR tool against Nanonets which is a web based OCR solution. The main focus of this is to recognition numerical values within a complex table structure containing diverse handwritten and printed digits.Unlike standard OCR tasks, this dataset presents several challenges that affect text recognition accuracy. The structure of the text and digits is poorly formatted which is leading to misalignment in the extracted text. The presence of black cells, dots, hyphens, and dashes introduces noise, which further disrupts the OCR system's ability to correctly detect and classify text.

Additionally, random dashes or dots appear after numerical values, adding to the complexity. These irregularities make it difficult for both OCR models to segment the content accurately and map the extracted values to the correct table cells.The first test demonstrated PaddleOCR achieved superior accuracy results compared to Nanonets because of its strength in processing structured data without substantial distortion. Additional noise elements in Test Case 2 caused PaddleOCR to perform poorly indicating that it has limited capabilities in dealing with extreme table structural distortions. The Nanonets system showed reduced accuracy during Test Case 1 but effectively handled Test Case 2 indicating its model shows stronger tolerance for unstandardized data formats.

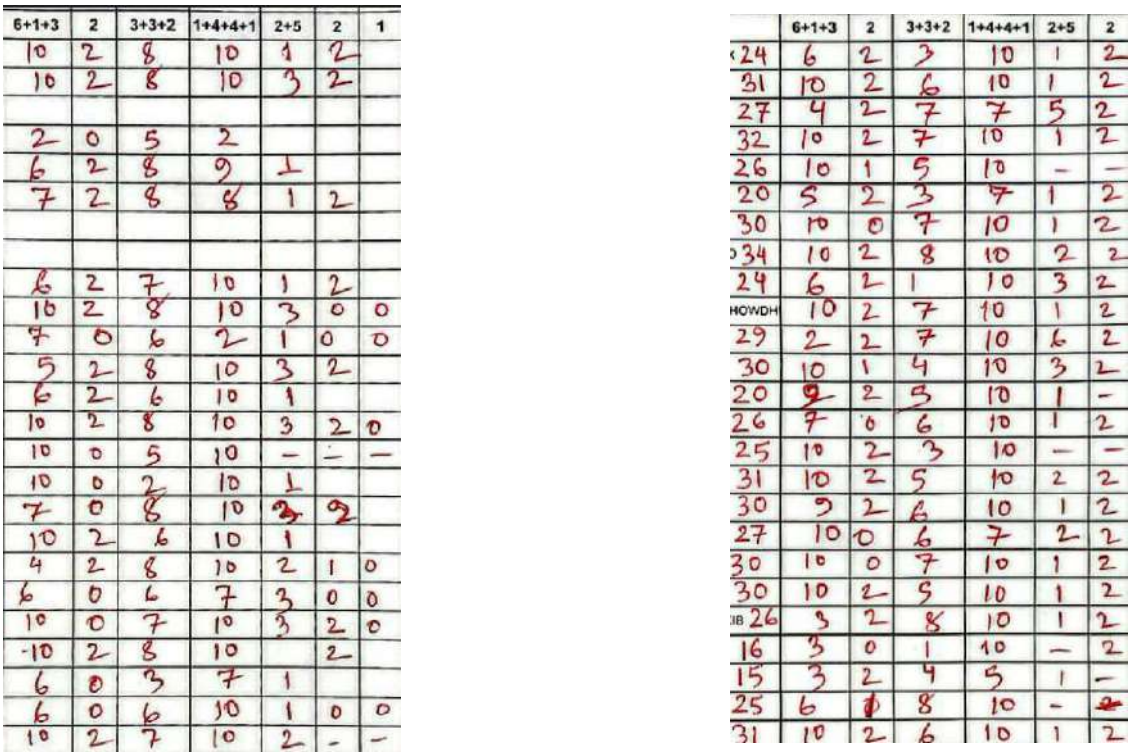


Figure 5.9: Handwritten test images for Complex OCR analysis.

The results indicate that PaddleOCR achieved a higher accuracy in Test Case 1 which is 43.21% but struggled significantly in Test Case 2 13.59%, largely due to the complexity and inconsistencies in the table structure. Nanonets, on the other hand, get lower accuracy in Test Case 1 9.04% but improved performance in Test Case 2 20.43%, likely due to its ability to generalize better across diverse structures. As for the Precision, Recall and F1-score test 1 get respectively 69.31%,53.44%,60.34% and for the Nanonets it finds 20.83%,50% and 29.41%. For the Test Case 2, Paddleocr's Precision, Recall and F1-score test 1 get respectively 21.99%,31.96%,26.05% and for the Nanonets it finds 47.50%,61.96% and 53.77%. The experimental results demonstrate that both OCR models must be optimized to properly process tabular data which contains both structured content and inconsistencies. The higher precision of PaddleOCR in this structured data indicates that it is more reliable when the format is relatively intact and on the other hand, Nanonets improvement in distorted cases suggests that it may perform better in dynamic and less structured environments. The future work would be like exploring these models to improve

Model	Metric	Test Case 1	Test Case 2
PaddleOCR	Total Correct Cells	70	31
	Total Incorrect Cells	28	109
	Total Missing Cells	61	66
	Total Extra Cells	3	22
	Total Empty Cells	46	6
	Average Character Error Rate (CER)	37.25%	85.59%
	Average Word Error Rate (WER)	39.43%	77.32%
	Accuracy	43.21%	13.59%
	Precision	69.31%	21.99%
	Recall	53.44%	31.96%
	F1 Score	60.34%	26.05%
Nanonets	Total Correct Cells	15	57
	Total Incorrect Cells	136	115
	Total Missing Cells	15	35
	Total Extra Cells	57	63
	Total Empty Cells	67	9
	Average Character Error Rate (CER)	0.44	0.7192
	Average Word Error Rate (WER)	0.91	0.7343
	Accuracy	9.04%	20.43%
	Precision	20.83%	47.50%
	Recall	50.00%	61.96%
	F1 Score	29.41%	53.77%

Table 5.5: OCR Model Performance for PaddleOCR and Nanonets

their robustness by fine tuning. It can then overcome in real world datasets where formatting inconsistencies are prevalent.

Chapter 6

Conclusion

This study successfully combines OpenCV for table detection and PaddleOCR for sequential handwritten text recognition, offering a balanced and efficient approach to marksheet digitization. OpenCV efficiently detects table structures without the computational demands typically associated with deep learning models, while PaddleOCR ensures robust recognition of handwritten sequences, effectively capturing diverse handwriting styles and document layouts. The independent use of YOLOv8 text and digit detection provided accurate region detection for text while processing digit data to enhance PaddleOCR functionality. YOLOv8 enables automatic data input while also providing an expandable flexible solution which addresses both educational administration requirements. A modified version of YOLOv8 received implementation to minimize model training and inference durations while requiring reduced computer resources and adjusting operations for handwriting data format. The upcoming work will concentrate on increasing system performance and enhancing user interface options. The system will include an application development phase to create an interface that enables document digitization for mobile and desktop users.Expanding the training dataset to include more diverse handwriting styles will improve the system's detection capabilities, while real-time processing and cloud deployment will enable efficient handling of large scale digitization tasks. Advanced preprocessing techniques will be explored to address challenges such as faint handwriting and low contrast text. Optimized algorithms for distinguishing visually similar characters, along with fine-tuned model parameters, will improve robustness across diverse scenarios. Efforts will also be directed toward modifying PaddleOCR and YOLOv8 to enhance speed and accuracy, particularly in bounding box generation and low-contrast text detection. By addressing these areas, the system aims to evolve into a comprehensive, user-friendly, and robust solution for automating handwritten document digitization. This work lays the foundation for iterative improvements in OCR systems, contributing to higher accuracy and adaptability across varied use cases.

Bibliography

[1] H. Beigi, "An overview of handwriting recognition," Feb. 1997.

[2] M. Laine and O. S. Nevalainen, "A standalone ocr system for mobile cameraphones," in *2006 IEEE 17th International Symposium on Personal, Indoor and Mobile Radio Communications*, IEEE, 2006, pp. 1–5.

[3] A. Shahab, F. Shafait, T. Kieninger, and A. Dengel, "An open approach towards the benchmarking of table structure recognition systems," Jun. 2010, pp. 113–120. DOI: 10.1145/1815330.1815345.

[4] S. Malakar, S. Halder, R. Sarkar, N. Das, S. Basu, and M. Nasipuri, "Text line extraction from handwritten document pages using spiral run length smearing algorithm," in *2012 International Conference on Communications, Devices and Intelligent Systems (CODIS)*, 2012, pp. 616–619. DOI: 10.1109/CODIS.2012.6422278.

[5] R. Jana, A. R. Chowdhury, and M. Islam, "Optical character recognition from text image," *International Journal of Computer Applications Technology and Research*, vol. 3, no. 4, pp. 240–244, 2014.

[6] P. M. Manwatkar and S. H. Yadav, "Text recognition from images," in *2015 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS)*, 2015, pp. 1–6. DOI: 10.1109/ICIIECS.2015.7193210.

[7] J. L. Pach and P. Bilski, "A robust text line detection in complex handwritten documents," in *2015 IEEE 8th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS)*, vol. 1, 2015, pp. 271–275. DOI: 10.1109/IDAACS.2015.7340742.

[8] G. Xie, Y. Mao, Z. Nan, S. Wang, Y. Li, and W. Wang, "Table recognition and recovery in the camera-captured image with complex background," Jan. 2015. DOI: 10.2991/asei-15.2015.27.

[9] G. Cohen, S. Afshar, J. Tapson, and A. van Schaik, *Emnist: An extension of mnist to handwritten letters*, 2017. DOI: 10.48550/ARXIV.1702.05373. [Online]. Available: https://arxiv.org/abs/1702.05373.

[10] E. Dufourq and B. A. Bassett, "Eden: Evolutionary deep networks for efficient machine learning," in *2017 Pattern Recognition Association of South Africa and Robotics and Mechatronics (PRASA-RobMech)*, 2017, pp. 110–115. DOI: 10.1109/RoboMech.2017.8261132.

[11] Y. Peng and H. Yin, "Markov random field based convolutional neural networks for image classification," in *Intelligent Data Engineering and Automated Learning–IDEAL 2017: 18th International Conference, Guilin, China, October 30–November 1, 2017, Proceedings 18*, Springer, 2017, pp. 387–396.

[12] S. Singh, A. Paul, and M. Arun, "Parallelization of digit recognition system using deep convolutional neural network on cuda," in *2017 Third International Conference on Sensing, Signal Processing and Security (ICSSS)*, 2017, pp. 379–383. DOI: 10.1109/SSPS.2017.8071623.

[13] A. Baldominos, Y. Saez, and P. Isasi, "Hybridizing evolutionary computation and deep neural networks: An approach to handwriting recognition using committees and transfer learning," *Complexity*, vol. 2019, no. 1, M. Scarpiniti, Ed., Jan. 2019, ISSN: 1099-0526. DOI: 10.1155/2019/2952304. [Online]. Available: http://dx.doi.org/10.1155/2019/2952304.

[14] P. Cavalin and L. Oliveira, "Confusion matrix-based building of hierarchical classification," in *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*. Springer International Publishing, 2019, pp. 271–278, ISBN: 9783030134693. DOI: 10.1007/978-3-030-13469-3_32. [Online]. Available: http://dx.doi.org/10.1007/978-3-030-13469-3_32.

[15] P. Honnegowda, N. M, S. D, S. M, and R. P, "Handwriting text recognition using neural networks," *International Journal of Innovative Technology and Exploring Engineering*, vol. 2, pp. 4088–4092, Dec. 2019. DOI: 10.35940/ijitee.B7705.129219.

[16] V. Jayasundara, S. Jayasekara, H. Jayasekara, J. Rajasegaran, S. Seneviratne, and R. Rodrigo, "Textcaps: Handwritten character recognition with very small datasets," in *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2019, pp. 254–262. DOI: 10.1109/WACV.2019.00033.

[17] S. Paliwal, V. D, R. Rahul, M. Sharma, and L. Vig, "Tablenet: Deep learning model for end-to-end table detection and tabular data extraction from scanned document images," Sep. 2019. DOI: 10.1109/ICDAR.2019.00029.

[18] Y. Du, C. Li, R. Guo, *et al.*, "Pp-ocr: A practical ultra lightweight ocr system," *arXiv preprint arXiv:2009.09941*, 2020. [Online]. Available: https://github.com/PaddlePaddle/PaddleOCR.

[19] P. Pad, S. Narduzzi, C. Kündig, E. Türetken, S. A. Bigdeli, and L. A. Dunbar, "Efficient neural vision systems based on convolutional image acquisition," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 12 282–12 291. DOI: 10.1109/CVPR42600.2020.01230.

[20] J. Jebadurai, I. J. Jebadurai, G. J. L. Paulraj, and S. V. Vangeepuram, "Handwritten text recognition and conversion using convolutional neural network (cnn) based deep learning model," in *2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA)*, 2021, pp. 1037–1042. DOI: 10.1109/ICIRCA51532.2021.9544513.

[21] I. Khandokar, M. M. Hasan, F. Ernawan, M. Islam, and M. N. Kabir, "Handwritten character recognition using convolutional neural network," *Journal of Physics: Conference Series*, vol. 1918, p. 042 152, Jun. 2021. DOI: 10.1088/1742-6596/1918/4/042152.

[22] P. Jeevan and A. Sethi, "Wavemix: Resource-efficient token mixing for images," *arXiv preprint arXiv:2203.03689*, 2022.

[23] J. K. Manipatruni, R. G. Sree, R. Padakanti, S. Naroju, and B. K. Depuru, "Leveraging artificial intelligence for simplified invoice automation: Paddle ocr-based text extraction from invoices," *International Journal of Innovative Science and Research Technology*, vol. 8, no. 9, pp. 1735–1740, 2023, ISSN: 2456-2165. [Online]. Available: https://www.ijisrt.com/.

[24] A. Mishra, A. S. Ram, and K. C, *Handwritten text recognition using convolutional neural network*, 2023. arXiv: 2307.05396 [cs.CV].

[25] S. Mupparaju, S. Thotakura, and C. V. RamiReddy, "A review on yolov8 and its advancements," in *Springer*, Jan. 2024. DOI: 10.1007/978-981-99-7962-2_39.

[26] S. R. N and N. Kennedy Babu C, "Adaptive residual attention network for handwritten character and digits recognition with improved energy valley optimizer algorithm," in *2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS)*, 2024, pp. 1–6. DOI: 10.1109/ADICS58448.2024.10533644.

[27] M. Raheem and N. Abubacker, "Handwritten digit recognition using machine learning," *Journal of Theoretical and Applied Information Technology*, vol. 102, pp. 2172–2180, Mar. 2024.

[28] M. Salaheldin Kasem, A. Abdallah, A. Berendeyev, *et al.*, "Deep learning for table detection and structure recognition: A survey," *ACM Comput. Surv.*, vol. 56, no. 12, Oct. 2024, ISSN: 0360-0300. DOI: 10.1145/3657281. [Online]. Available: https://doi.org/10.1145/3657281.

[29] M. Talib, A. H. Y. Al-Noori, and J. Suad, "YOLOv8-CAB: Improved YOLOv8 for Real-time Object Detection," *Karbala International Journal of Modern Science*, vol. 10, pp. 56–68, Jan. 2024. DOI: 10.33640/2405-609X.3339.

[30] S. Umatia, A. Varma, A. Syed, K. Tiwari, and M. F. Shah, "Text recognition from images,"