CONVERSION OF MARKS INTO DIGITAL FORMAT USING CNNS AND ANALYSIS OF MARKS

A MAJOR PROJECT

REVIEW-1 REPORT

SUBMITTED BY,

BATCH – 2A

B. Avinash **(20331A0522)**

CH Sampath **(20331A0536)**

Ch. Ravi teja **(20331A0530)**

E. Mani Bhaskar **(20331A0553)**

K. Sravani **(21335A0504)**

Project Supervisor
Mr. M. VAMSI KRISHNA
Assistant Professor

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
MVGR COLLEGE OF ENGINEERING (AUTONOMOUS)
VIZIANAGARAM-535005, AP (INDIA)
(Accredited by NBA, NAAC, and Permanently Affiliated to Jawaharlal Nehru Technological University Kakinada)

TABLE OF CONTENTS

1. ABSTRACT

We propose a specialized software solution to enhance educational institutions' performance analysis by digitizing and analysing student marks obtained from scanned mark sheets.

This system leverages Python's CV2 module for mark segmentation and utilizes customizable Convolutional Neural Networks (CNNs) for accurate digit recognition, addressing issues like image inconsistencies and unclear borders.

The stored digital data allows for historical analysis and actionable insights into student performance, aligned with Bloom's Taxonomy.

Notably, the software seamlessly integrates with existing processes without disruption.

The prototype demonstrates the software's generic applicability, with planned iterative upgrades to maximize issue resolution for this specific design, accommodating diverse institutional needs.

2. INTRODUCTION

This project aims to address the challenges associated with manual data entry and subjective analysis by introducing a robust solution for the recognition of marks and their seamless conversion into a digital format using CNNs. By harnessing the power of image processing and deep learning, this system seeks to automate the extraction of student marks from scanned mark sheets, thereby streamlining administrative processes and minimizing errors arising from human intervention.

Through the integration of the Python CV2 module and customizable CNNs, the software will segment and recognize individual marks accurately, even when faced with challenges such as varied mark sheet designs, inconsistent image sizes, and unclear boundaries. This innovation promises to not only improve the efficiency of data management but also enable educational institutions to gain valuable insights into student performance trends and areas of improvement.

3. PROJECT GOALS

This project aims to address the challenges associated with manual data entry and subjective analysis by introducing a robust solution for the recognition of marks and their seamless conversion into a digital format using CNNs. By harnessing the power of image processing and deep learning, this system seeks to automate the extraction of student marks from scanned mark sheets, thereby streamlining administrative processes and minimizing errors arising from human intervention.

Through the integration of the Python CV2 module and customizable CNNs, the software will segment and recognize individual marks accurately, even when faced with challenges such as varied mark sheet designs, inconsistent image sizes, and unclear boundaries. This innovation promises to not only improve the efficiency of data management but also enable educational institutions to gain valuable insights into student performance trends and areas of improvement.

1. Segmentation and Localization: Implement computer vision techniques using tools like the CV2 module to segment individual marks, ensuring precise localization for subsequent recognition.
2. UI-Design and Implementation : Create a user-friendly interface that allows institutions to give the input of semester papers according to their bundle numbers and barcodes.
3. Accurate Mark Recognition: Develop a convolutional neural network (CNN) model tailored for digit recognition, capable of accurately identifying marks on scanned images of varying quality and design formats.
4. Analysing : Implement data analysis techniques to provide actionable insights into student performance, utilizing established educational

frameworks in accordance to Bloom's Taxonomy to categorize performance levels.

5. Data Integrity and Security: Design a secure and efficient database system to store digitized marks while maintaining data integrity, ensuring that student information is stored and managed responsibly.

4. REQUIREMENTS GATHERING

During this project we have managed to gather there are 3 assistant examination controllers out of which the 3rd examination controller is a primary stakeholder of our project as he is a admin of evaluation of marks therefore he is reponsible for converting to digital marks

Identifying stakeholders:-

- ACE3 is responsible for evaluation of marks for the whole college and providing analysis
- NBAC / HOD'S is responsible for improving student standrads with the analysis provided by ACE3

5. REQUIREMENTS ANALYSIS

5.1 DFD Level-0:-

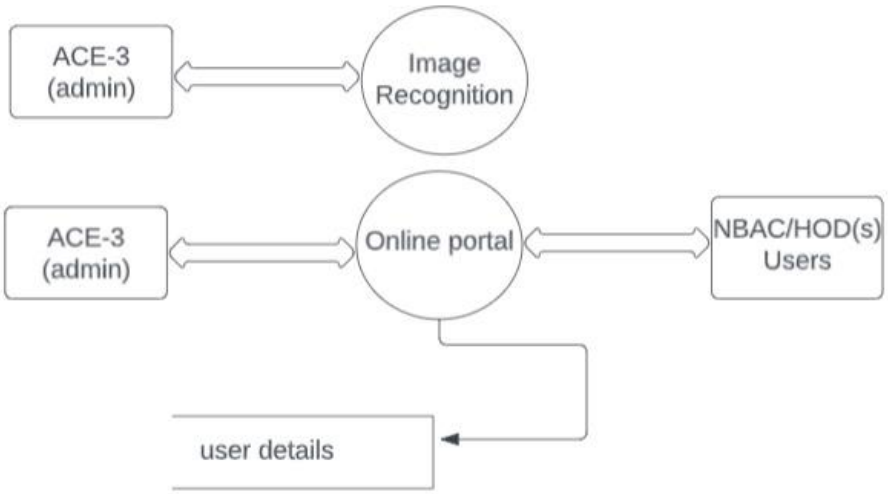


Figure 1-DFD level 0

5.2 DFD Level-1:-

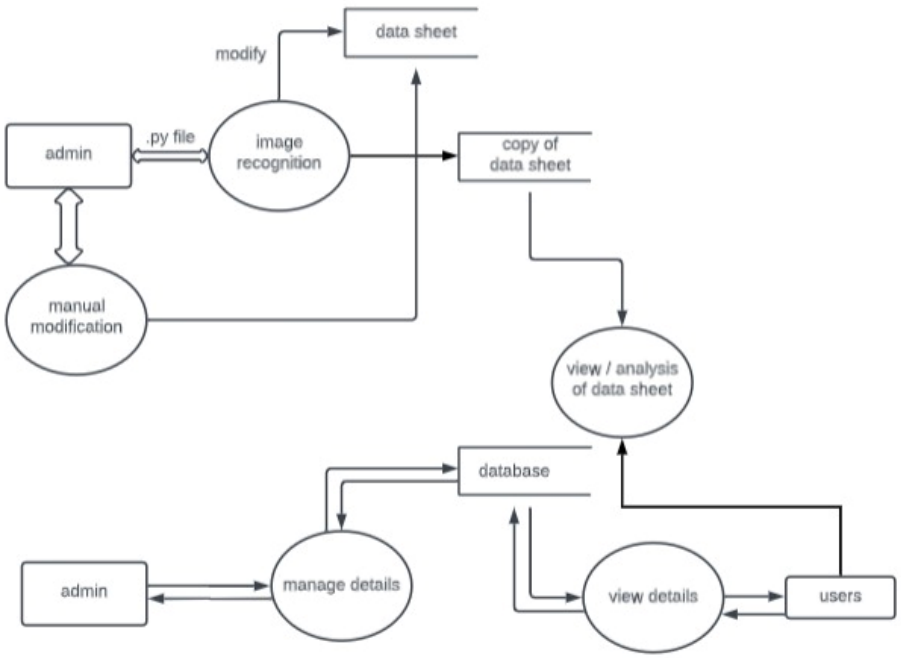


Figure 2-DFD level 1

5.3 DFD Level-2

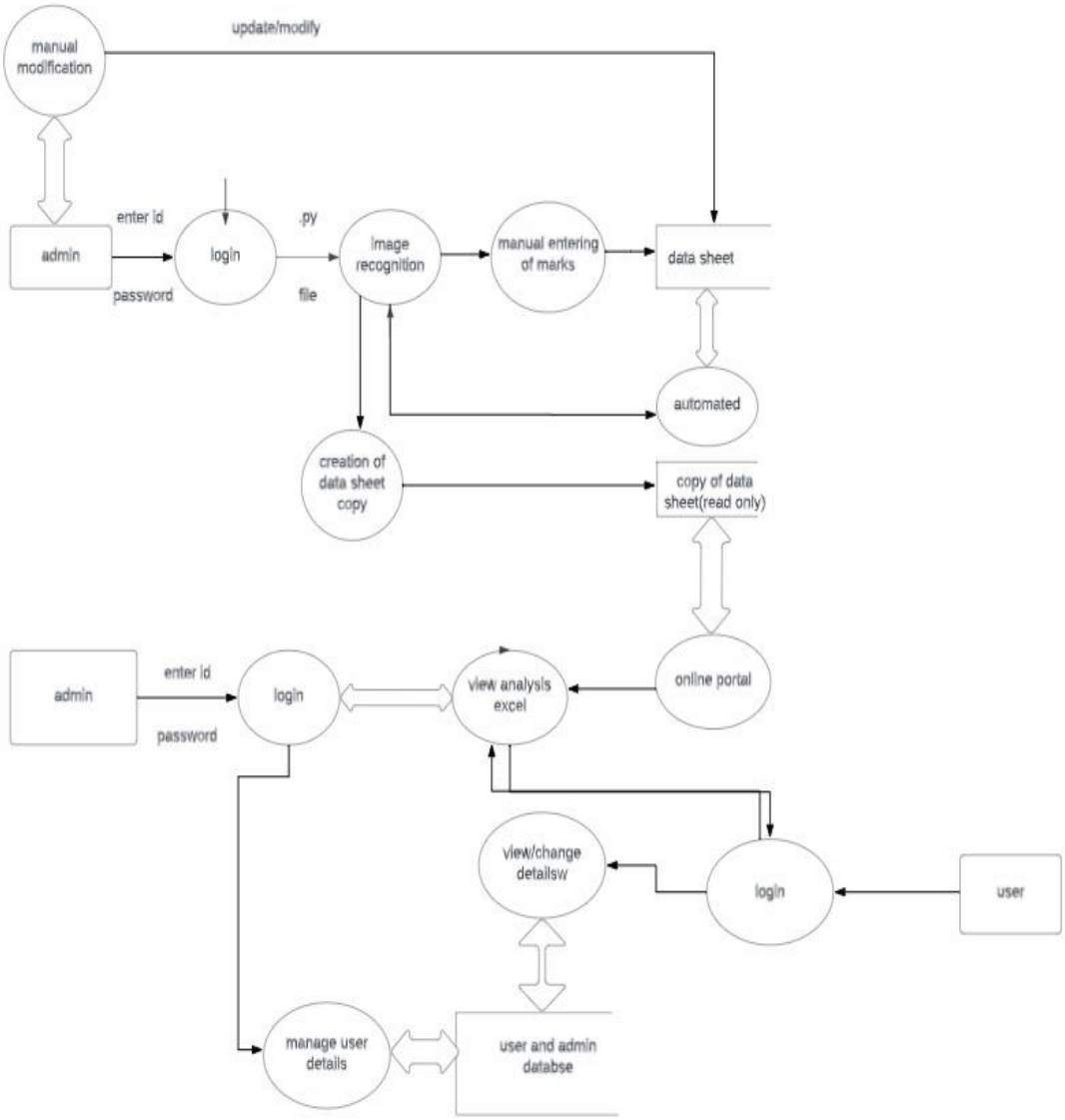


Figure 3-DFD level 2

5.4 Use case Diagrams:-

5.4.1 Use case diagram – of Users

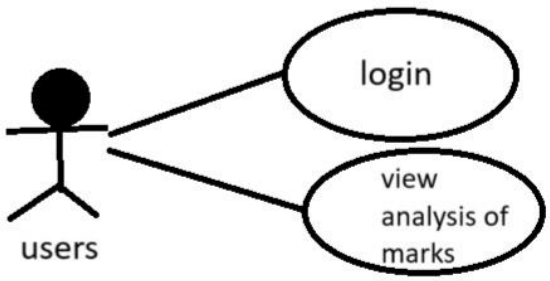


Figure 4-Use case Diagram (users)

5.4.2 Use case diagram – of Admin

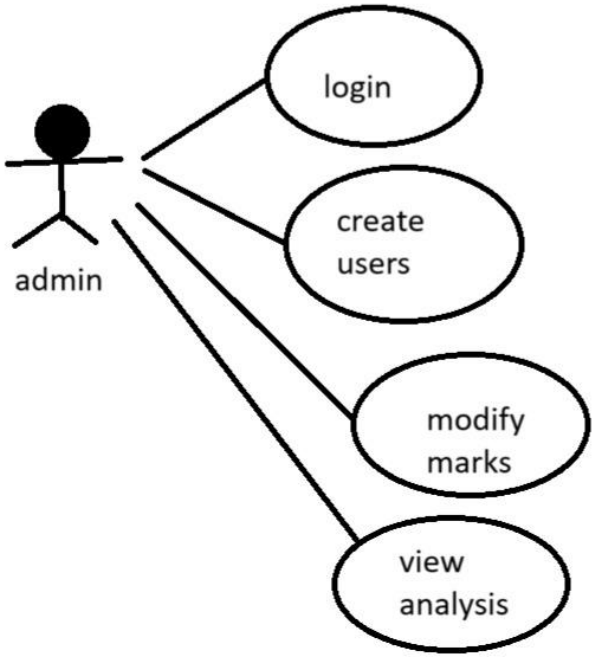


Figure 5-Use case Diagram (Admin)

6. KEY FINDINGS

The project focused on the recognition of physical marks and their subsequent conversion into a digital format. Through rigorous investigation and implementation, several key findings emerged that underscore the significance of this endeavor.

Firstly, the process of recognizing marks, whether they are handwritten or printed, presents challenges related to variations in writing styles, ink shades, and paper quality. Implementing advanced image processing and machine learning techniques was crucial to accurately identifying and extracting these marks.

Secondly, the choice of technology played a pivotal role in the success of the project. Optical Character Recognition (OCR) systems demonstrated their prowess in converting scanned or photographed marks into editable digital content. The integration of Natural Language Processing (NLP) algorithms further enhanced the system's ability to interpret and comprehend the context of these marks.

Furthermore, ensuring the preservation of data integrity during the conversion process emerged as a key concern. Developing algorithms to detect smudges, creases, or incomplete marks helped maintain the accuracy of the digital output, contributing to reliable and meaningful data representation.

The project's success heavily relied on a robust training dataset. A diverse collection of marks was compiled, encompassing various writing styles, languages, and forms of marks. This dataset was essential for training machine learning models to recognize and categorize marks effectively, transcending the boundaries of different mark types and formats.

7. TARGET COMMUNITY/BENEFICIARIES

The project's primary objective is the efficient recognition and seamless conversion of physical marks, such as handwritten annotations or symbols, into a digital format. This innovative initiative aims to significantly enhance data accessibility, organization, and utilization. The beneficiaries of this endeavour encompass a wide range of communities and individuals, including students, researchers, professionals, and institutions. By catering to the diverse needs of these target groups, the project directly addresses the challenges posed by manual data entry and retrieval processes. Through the transformation of marks into digital data, students can effortlessly digitize their notes, researchers can convert field observations with ease, and professionals can streamline data integration into their workflow. Moreover, educational institutions can maintain comprehensive digital records of assessments, facilitating efficient analysis and reporting. This project bridges the gap between conventional and digital modes of information, offering unparalleled convenience and utility. In a society increasingly reliant on digital solutions, this initiative holds the potential to revolutionize the way we interact with and manage information, empowering individuals across various domains to harness the benefits of advanced data recognition and digital transformation.

8. PROBLEM IDENTIFICATION AND DEFINITION

The project aims to address the challenge of efficiently recognizing and converting physical marks into a digital format. This problem is rooted in the need to streamline and modernize data collection processes that currently involve manual entry or interpretation of marks, which can be time-consuming, error-prone, and resource-intensive.

The core issue revolves around the identification and extraction of marks from various sources, such as paper documents or images. These marks could encompass a wide range of forms, including checkboxes, tick marks, handwritten annotations, and symbols. The lack of an automated system results in bottlenecks in data processing workflows, hindering the timely availability of accurate information for analysis and decision-making.

To tackle this, the project aims to develop a robust recognition system that utilizes advanced image processing and machine learning techniques. This system should be capable of accurately identifying different types of marks and converting them into a standardized digital format. The project's success hinges on the ability to handle diverse mark styles, sizes, and orientations, ensuring high precision and recall rates in mark recognition.

Furthermore, the project seeks to address potential challenges such as variability in mark quality, skewed or distorted input images, and integration into existing data management systems. It's essential to strike a balance between accuracy and efficiency, enabling the recognition system to swiftly process large volumes of marks without compromising on data integrity.

The envisioned solution could have far-reaching implications across various sectors, including education, healthcare, and business, where the conversion of marked information into digital form is integral to operations. By automating this conversion process, the project aims to significantly reduce manual effort,

mitigate errors, and enhance the overall efficiency of data collection and analysis.

In conclusion, the project's primary objective is to develop an advanced recognition system capable of accurately identifying and converting marks from physical documents into a standardized digital format. By addressing the challenges of mark variability, image distortion, and integration, the project aspires to revolutionize data processing workflows and contribute to more streamlined and effective information management across industries.

9. PROJECT APPROACH

The project "**Recognition of Marks and Conversion into Digital Format**" aims to streamline the process of converting physical marks or scores into a digital format for easier storage, analysis, and accessibility. The approach for this project involves several key steps.

1. **Data Collection**: Gather physical mark sheets or documents containing scores. This could include exam papers, answer sheets, or any documents with marked scores that need to be digitized.
2. **Image Scanning**: Use high-resolution scanners or smartphone apps to capture clear and accurate images of the physical documents. Proper lighting and alignment are essential to ensure quality scans.
3. **Image Preprocessing**: Apply image processing techniques to enhance the quality of scanned images. This may involve noise reduction, contrast adjustment, and cropping to isolate the relevant score regions.
4. **Digit/Text Extraction**: Utilize Optical Character Recognition (OCR) technology to extract the textual information from the scanned images. OCR software will convert the handwritten or printed marks into machine-readable text.
5. **Data Verification**: Implement algorithms or manual checks to validate the accuracy of extracted scores. This could involve comparing the extracted data with the original images to ensure minimal errors.
6. **Data Storage**: Create a structured digital database to store the extracted marks along with relevant metadata such as student IDs, exam dates, and subject details. This could be a spreadsheet, database, or any suitable storage system.
7. **Data Conversion**: Convert the extracted data into a standardized digital format, such as CSV or JSON, for easy sharing and compatibility with other software.

8. **Data Analysis**: Develop tools to perform various analyses on the digitized marks, such as calculating averages, generating graphs, and identifying trends. This can provide valuable insights for educators and administrators.
9. **User Interface**: Design a user-friendly interface that allows users to upload scanned documents, review extracted data, and perform basic interactions like editing erroneous entries.
10. **Error Handling**: Implement robust error handling mechanisms to address cases where the OCR might misinterpret handwriting, leading to incorrect data extraction.
11. **Security Measures**: Ensure the security of the digitized data by implementing encryption, user authentication, and authorization protocols to prevent unauthorized access.
12. **Integration Options**: Provide options to integrate the digitized data with existing educational management systems or software used by institutions.
13. **Feedback Loop**: Incorporate user feedback to continuously improve the accuracy of OCR and the overall system performance.
14. **Testing and Validation**: Rigorously test the system with a diverse set of mark sheets to validate its accuracy, efficiency, and reliability.
15. **Scalability Consideration**: Design the system to handle a large volume of mark sheets efficiently, considering scalability as the user base grows.
16. **Documentation and Training**: Prepare comprehensive documentation and user guides to assist administrators, teachers, and other stakeholders in effectively using the system.

10.TIMELINE AND MILESTONES OF PROJECT

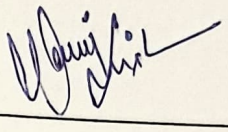
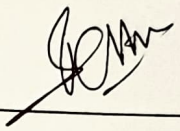
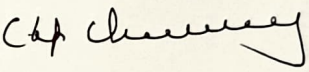
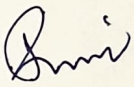
PHASES	PROCESS	DURATION
Phase – 1	Project initiation and planning	1 month
Phase - 2	System design and development	2 months
Phase - 3	Code development and testing	3 - 4 months
Phase - 4	Full – scale development and evaluation	2 months
Phase - 5	Finalization and Documentation	1 month

11. CONCLUSION

The utilization of Convolutional Neural Networks (CNNs) for recognizing and converting marks into a digital format stands as a transformative advancement in educational data management. This innovative approach addresses the challenges posed by traditional paper-based mark sheets, offering a streamlined and efficient solution for educational institutions. By harnessing the power of CNNs, the accuracy and precision of mark recognition have been significantly improved, mitigating errors stemming from image inconsistencies and handwriting variations. This achievement not only accelerates the digitization process but also establishes a foundation for data-driven insights that can drive meaningful improvements in educational practices.

12. REFERENCES

1. Advancements in Image Classification using Convolutional Neural Network https://sci-hub.se/10.1109/ICRCICN.2018.8718718
2. A Method for Improving CNN-Based Image Recognition Using DCGAN https://cdn.techscience.cn/files/cmc/2018/v57n1/20181019035712_68114.pdf
3. Recognition of Handwritten Digits https://neuralnetworksanddeeplearning.com

Role	Faculty Name	Designation	Signature
Supervisor	Mr. M. Vamsi Krishna	Assistant Professor	
Project Coordinator	Mr. P. Rama Santosh Naidu	Assistant Professor	
PRC Faculty-1	Dr. C. K. Chakravarthy	Professor	
PRC Faculty-2	Mrs. M. B. Raani	Assistant Professor	

Head of the Department
Computer Science and Engineering
MVGR College of Engineering(A)
Vizianagaram