

Marksheet Digitization using OCR and Report Generation for QAA

Team members

Anish Timsina (THA077BEI007)

Bishal Giri (THA077BEI014)

Sugam Khatiwada (THA077BEI046)

Supervised By:

Associate Professor Er. Shanta Maharjan

Department of Electronics and Computer Engineering
Institute of Engineering, Thapathali Campus

March, 2025

Presentation Outline

- Motivation
- Objectives
- Scope of Project
- Project Limitation
- Methodology
- Result and Analysis
- Future Enhancements
- Applications
- Conclusion
- Project Timeline
- Project Budget
- References

Motivation

- Paper-based and labor-intensive storage of documents like marksheets.
- Difficulty for students as well as administrators to access the required academic records.

Objectives

- To digitize hardcopy of marksheet using OCR and provide secure access through user-friendly web platform
- To develop a progress report as per the needs of QAA

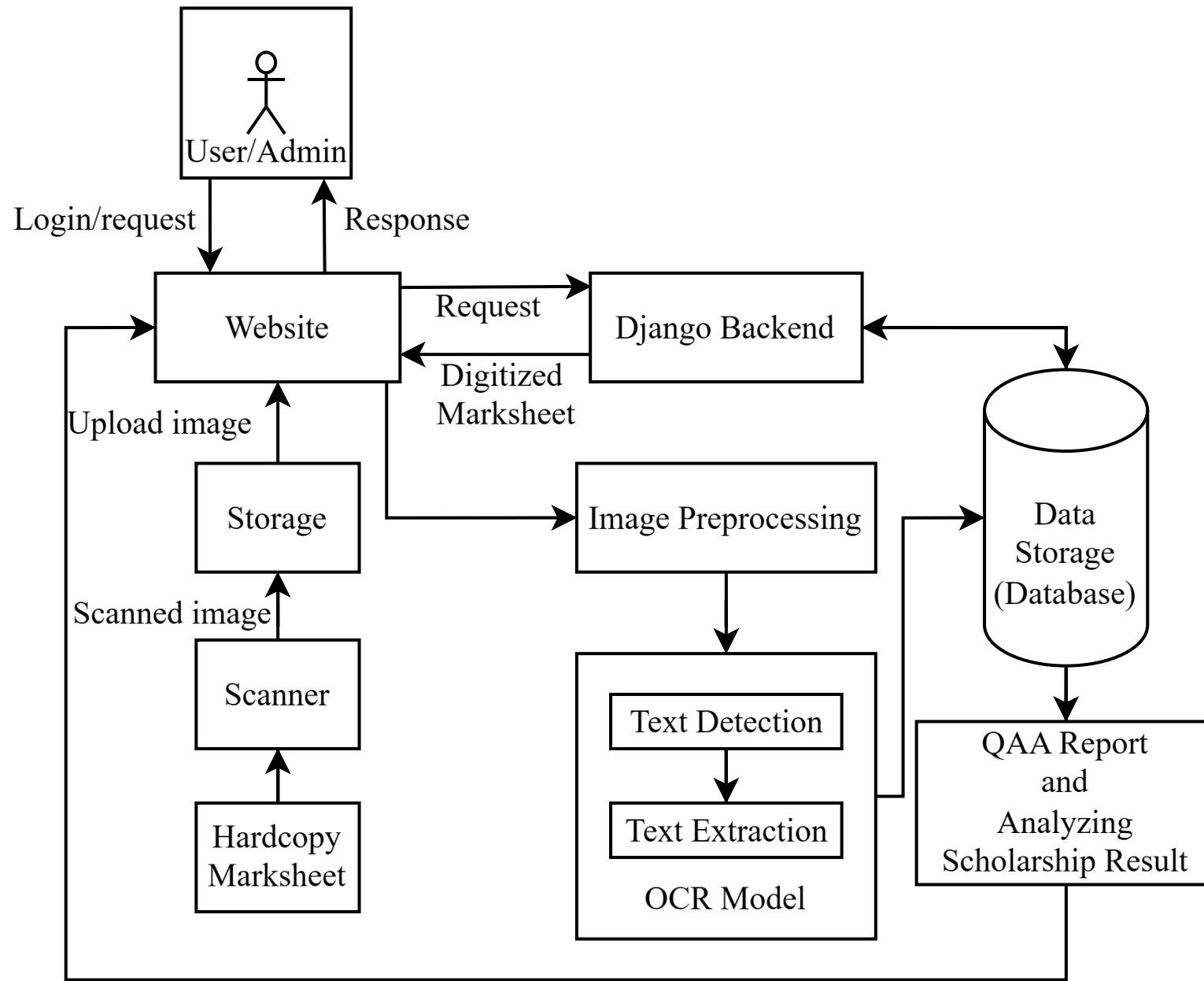
Scope

- High-resolution scanning and for accurate digitization.
- Secure database for storing and organizing digitized marksheets
- User-friendly web interface for accessing digitized marksheets.

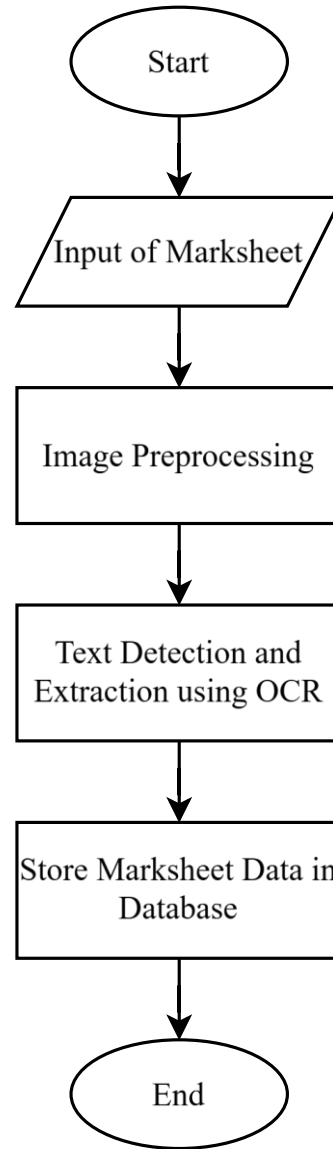
Limitation

- Designed for IOE marksheets, limiting adaptability to other institutions' formats.
- Faces challenges extracting text from noisy images with stamps or logos.
- Limited versatility and accuracy with diverse marksheet styles and formats.

Methodology - [1] (System Block Diagram)



Methodology - [2] (Flowchart of System)



Methodology - [2]

(Block Description)

- **Image Processing and Upload:**

Scanned marksheets are uploaded via a user-friendly web interface and temporarily stored.

- **Preprocessing with OpenCV:**

Images undergo preprocessing (noise reduction, contrast enhancement, skew correction) using OpenCV.

- **Text Extraction with OCR:**

Preprocessed images are converted into machine-readable text with Tesseract OCR and Paddle-OCR

- **Data Management with SQLite:**

Extracted text is stored in SQLite database, enabling access to academic records and query execution.

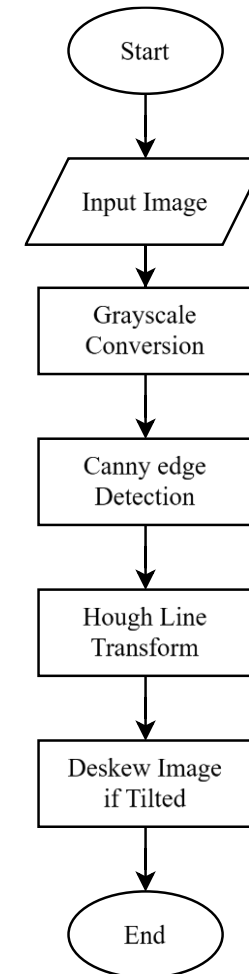
- **Reporting and Analytics:**

The system generates QAA reports and analyzes scholarship results, utilizing stored data for insightful analytics.

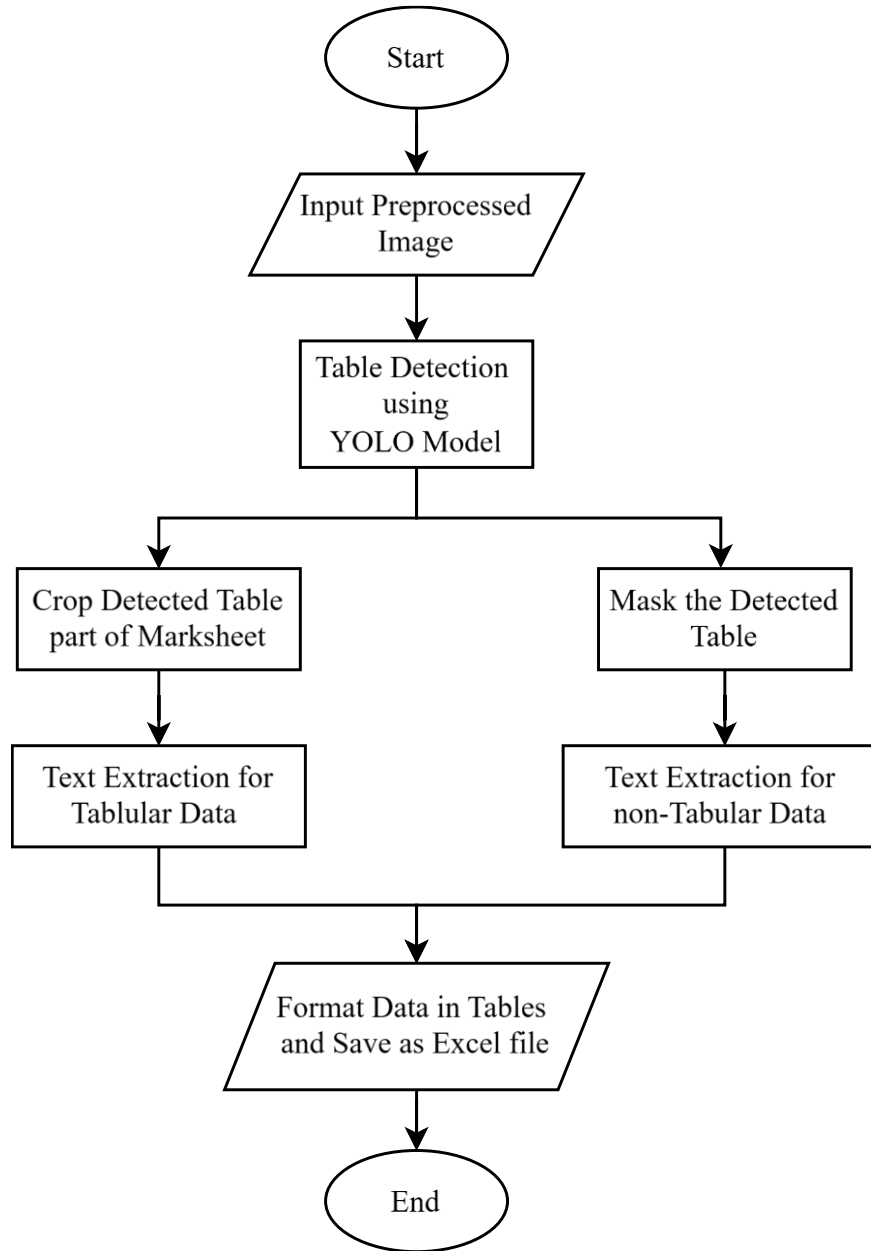
Methodology - [3]

(Image Preprocessing)

- For Image preprocessing, OpenCV is used.
- Image preprocessing includes:
 - Grayscale Conversion
 - Canny Edge Detection
 - Hough Line Transform
 - Rotate Image If Tilted



Methodology – [4] (Flowchart of Text Extraction)



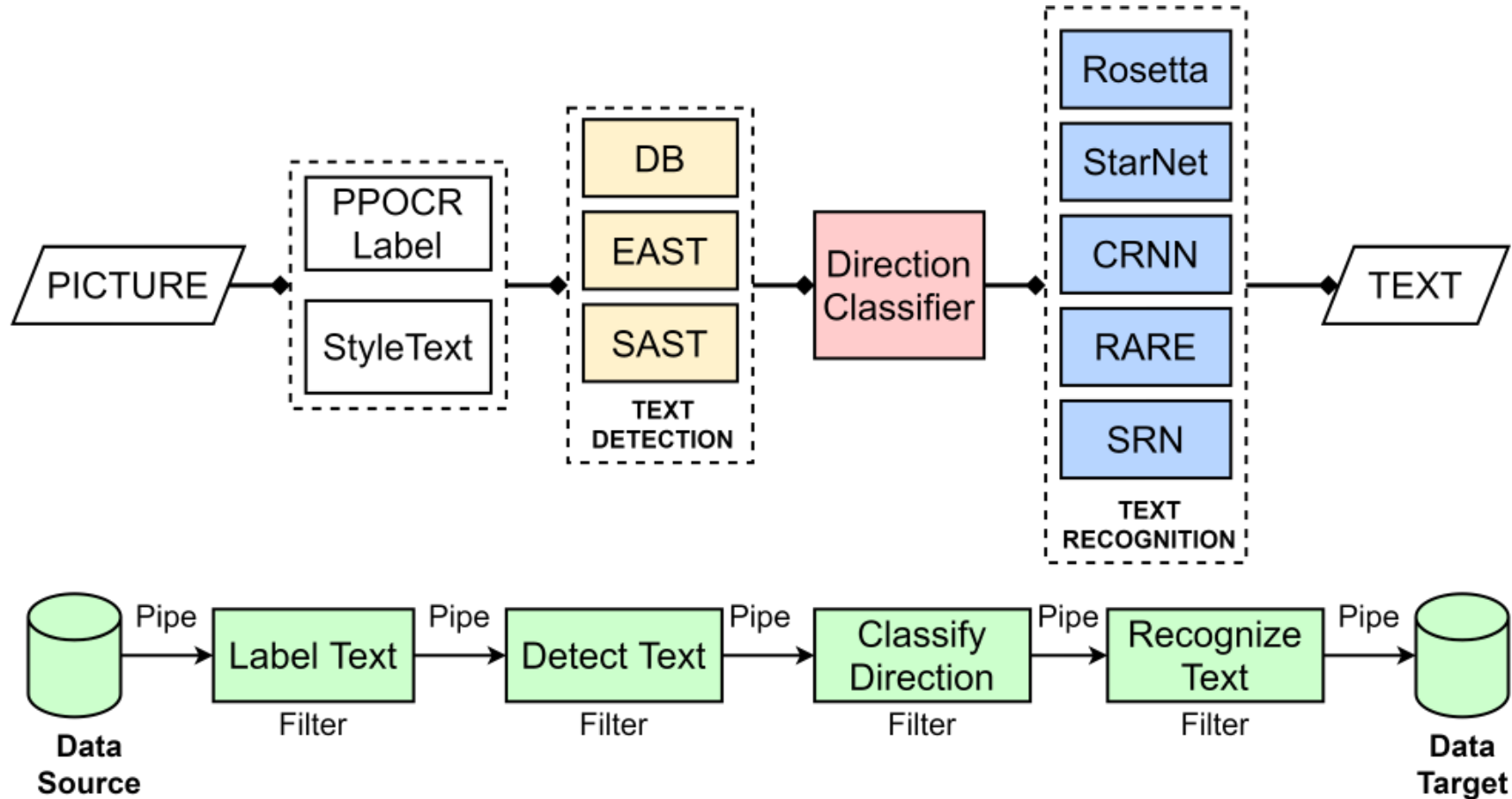
Methodology - [5]

(Table Detection and Extraction)

- Table detection is done using YOLOv8 model.
- It is pretrained on dataset that includes table structures.
- The model gives bounding boxes around the table to detect table region which is then separated as cropped image.

Methodology – [6]

(Paddle-OCR Working Procedure)



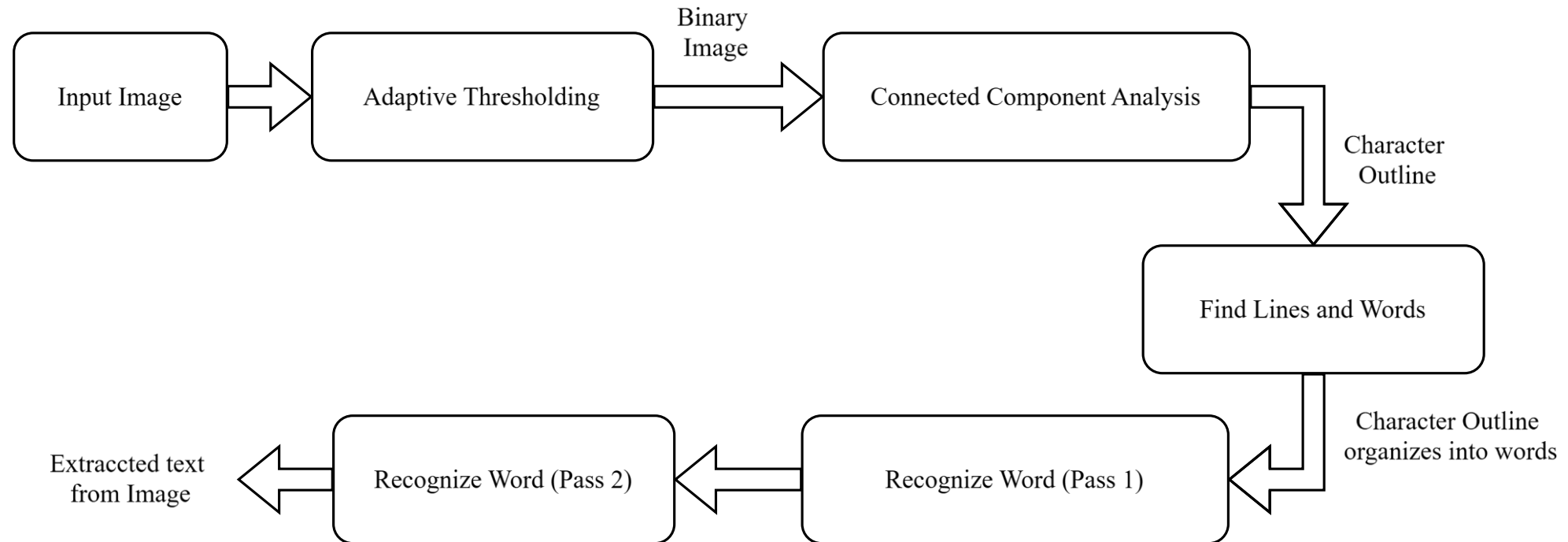
Methodology – [6]

(Paddle-OCR Working Procedure)

- **Input Data:** Images containing text.
- **Label Text:** Identify and mark text regions.
- **Text Detection:** Use models like DB, EAST, SAST to detect text.
- **Classify Direction:** Classify the orientation of detected text using a direction classifier.
- **Text Recognition:** Recognize text with models such as Rosetta, StarNet, CRNN, RARE, SRN.
- **Output Data:** Store the recognized text in a database.

Methodology – [7]

(Tesseract-OCR Working Procedure)

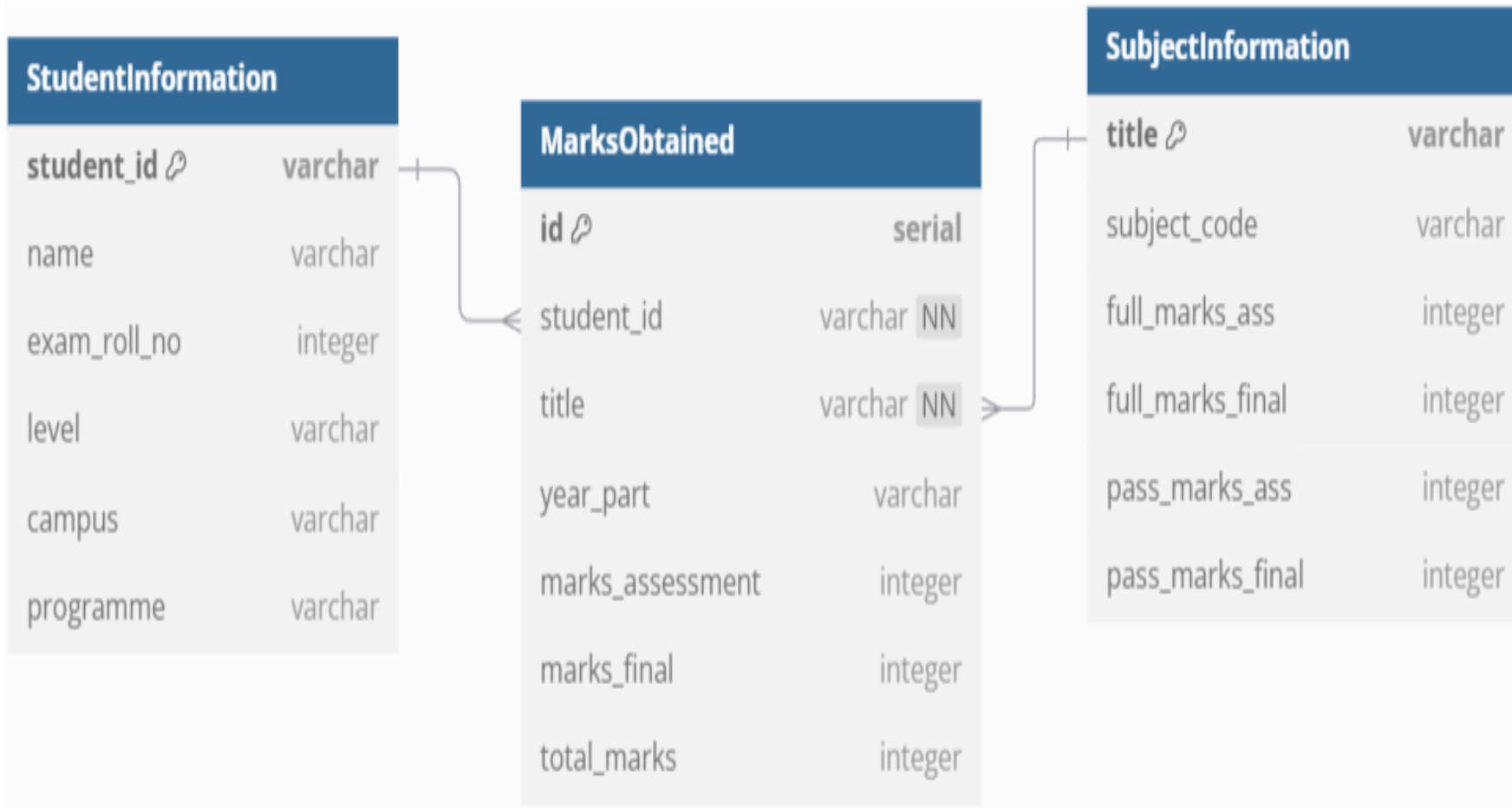


Methodology – [7]

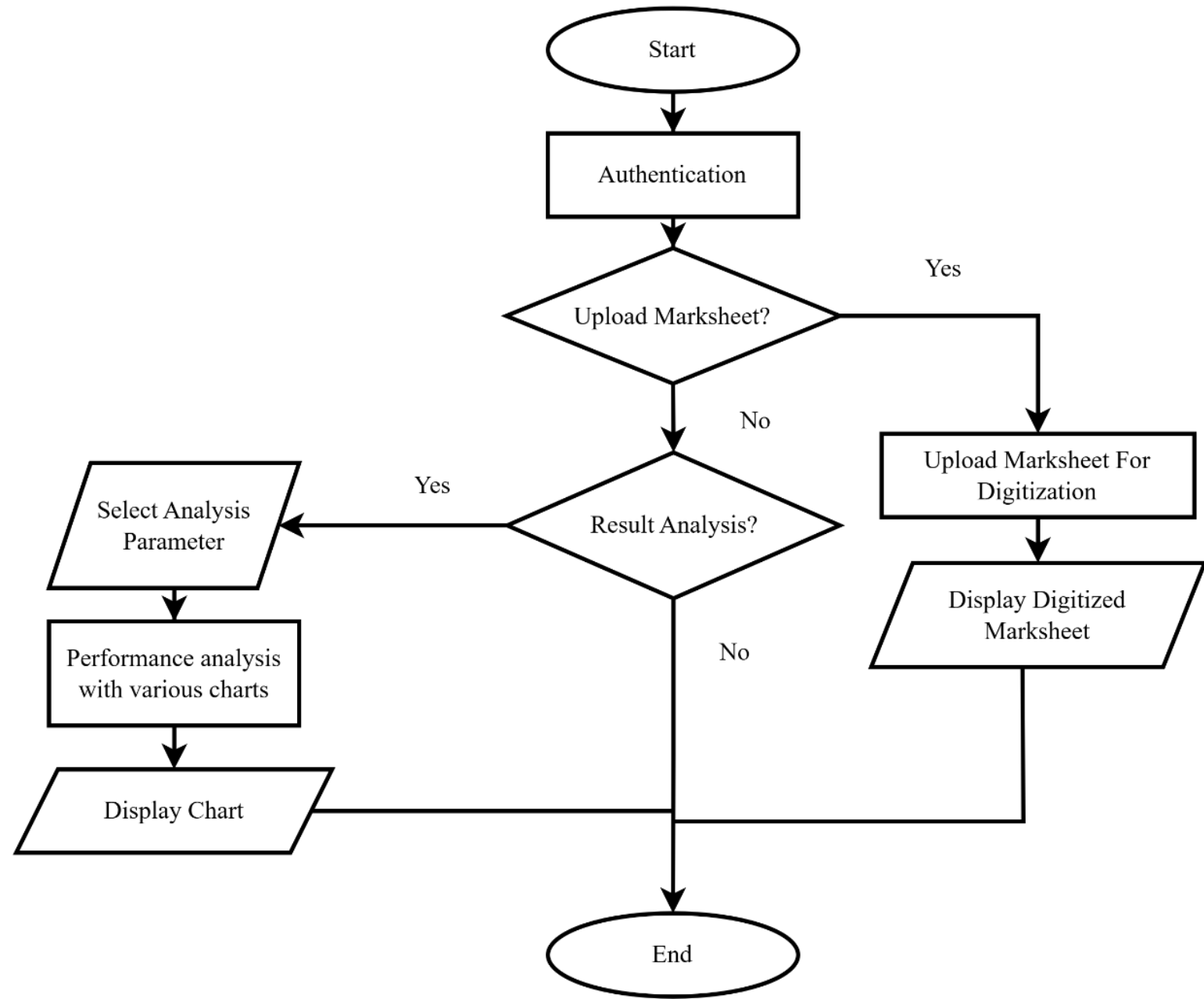
(Tesseract-OCR Working Procedure)

- **Input Image:** Start with the image containing text.
- **Adaptive Thresholding:** Convert the image to a binary format (black and white).
- **Connected Component Analysis:** Analyze the binary image to find connected components, which correspond to character outlines.
- **Find Lines and Words:** Organize the character outlines into lines and words.
- **Recognize Word (Pass 1):** Perform an initial recognition of words.
- **Recognize Word (Pass 2):** Refine the recognition of words to improve accuracy.
- **Extracted Text from Image:** Output the final recognized text from the image.

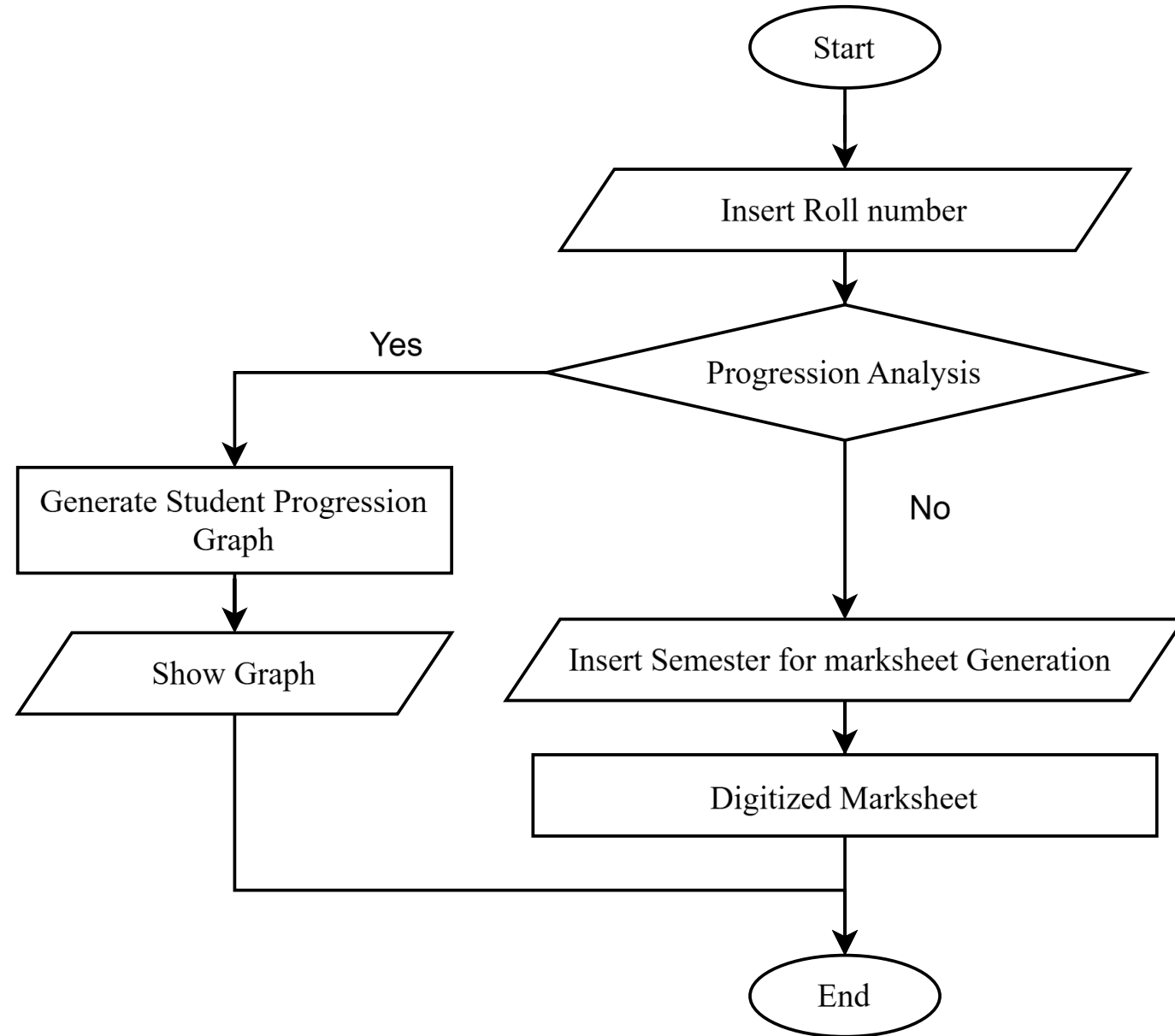
Methodology - [8] (Database Design – ER Diagram)



Methodology - [9] (User Interface- Admin Dashboard)



Methodology - [9] (User Interface- Student Dashboard)



Methodology - [10]

(User Interface-Performance Analysis)

- Admins can access a dedicated analysis page to evaluate student performance.
- Analysis includes class ranks, subject-wise and semester-wise pass ratios, overall pass/fail ratios, and more.
- Also allows comparisons of department-wise results and other performance metrics.

Methodology - [11]

(Requirements)

- **Hardware, Software, Libraries and Frameworks Requirements:**

- Scanner
- Google Collaboratory
- Visual Studio Code
- Keras
- Scikit-Learn
- NumPy
- Pandas
- Matplotlib
- OpenCV
- Pytesseract
- Django
- HTML/CSS
- SQLite
- Django ORM
- PyTorch
- PP Structure

Result and Analysis-[1] (Table Detection And Extraction)



Tribhuvan University
Institute of Engineering
Examination Control Division
Chakapat, Lalitpur
Back-paper Examination 2080 Ashwin
STATEMENT OF MARKS

Name:- Suman, B. B. Shrestha
Level :- Bachelor's in Engineering
Campus:- Thapathali Campus
Year/Part- III/II

Exam Roll No:- 72254
CRN:- 2075/BCT/037
T.U. Regd. No:- 3.2.26.404.2018
Programme:- Computer Engineering

Subjects		Full Marks		Pass Marks		Marks Obtained		Total	Remarks
		Asst.	Final	Asst.	Final	Asst.	Final		
CE655	Engineering Economics	20	80	8	32	12	A	—	
CT651	Object Oriented Analysis & Design	20	80	8	32	15	A	—	
CT652	Database Management System	20	80	8	32	14	A	—	
CT653	Artificial Intelligence	20	80	8	32	17	A	—	
CT655	Embedded System	20	80	8	32	16	A	—	
CT656	Operating System	20	80	8	32	16	A	—	

Marks Enter By:-	Grand Total
Verified By:- <u>S</u>	Result Absent
Date:- 11 JAN 2024	<u>[Signature]</u>
* - Fail A - Absent	Asst. Dean

Result and Analysis-[2]

(Text Detection And Extraction)

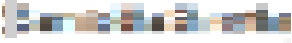
- Extracted table text Formatted into Excel file using PP Structure

	A	B	C	D	E	F	G	H	I	J
1	Subjects		Full Marks		Pass Marks		Marks Obtained			
2	Code	Title	Asst.	Final	Asst.	Final	Asst.	Final	Total	Remarks
3	CE655	Engineering Economics	20	80	8	32	12	A		
4	CT651	Object Oriented Analysis & Design	20	80	8	32	15	A		
5	CT652	Database Management System	20	80	8	32	14	A		
6	CT653	Artificial Intelligence	20	80	8	32	17	A		
7	CT655	Embedded System	20	80	8	32	16	A		
8	CT656	Operating System	20	80	8	32	16	A		
9										

Result and Analysis-[2]

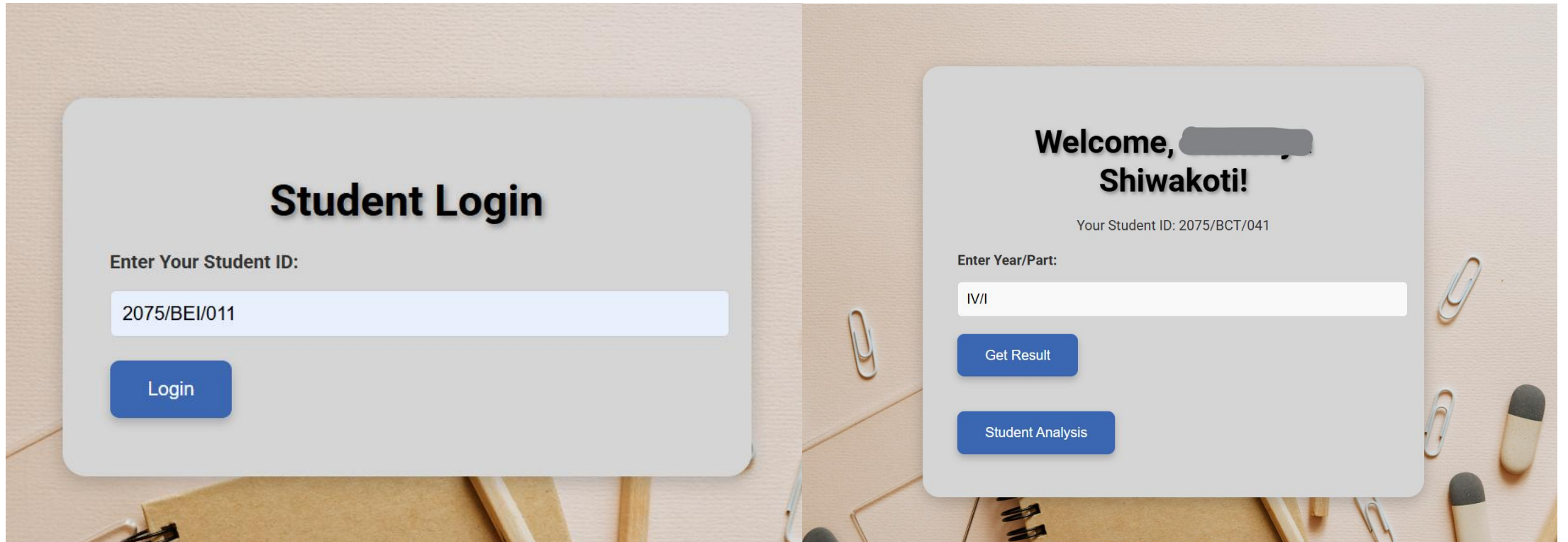
(Text Detection And Extraction)

- Extracted Text From TesseractOCR Formatted into Excel file

Name	Exam Roll No	Level	CRN	Campus	T.U. Regd. No	Year/Part	Programme
	72254	Bachelors in Engineering	2075/BCT/037	Thapathali Campus	3.2.26.404.2018	III/II	Computer Engineering

Result and Analysis-[3] (User Interface)

- Student Login Page and Dashboard Page



Result and Analysis-[3] (User Interface)

- Result Generated

Results for **Shiwakoti** (ID: 2075/BCT/041)

Programme: Computer Engineering
Year/Part: IV/I

Subject Title	Full Marks Ass.	Full Marks Final	Pass Marks Ass.	Pass Marks Final	Marks Ass.	Marks Final	Total Marks
Distributed System PRACTICAL	25	0	10	0	22.0		22.0
Digital Signal Analysis & Processing	20	80	8	32	17.0	67	84.0
Digital Signal Analysis & Processing PRACTICAL	25	0	10	0	24.0		24.0
Project I PRACTICAL	50	0	20	0	46.0		46.0
Energy Environment & Society	10	40	4	16	9.0	23	32.0
Organization & Management	20	80	8	32	19.0	61	80.0
Web Technologies and Applications (Elective I)	20	80	8	32	17.0	62	79.0
Web Technologies and Applications (Elective I) PRACTICAL	25	0	10	0	23.0		23.0

Download as PDF

Result and Analysis-[3] (User Interface)

- Result in PDF

File C:/Users/sugam/Downloads/result_2075_BCT_041_IV_I.pdf

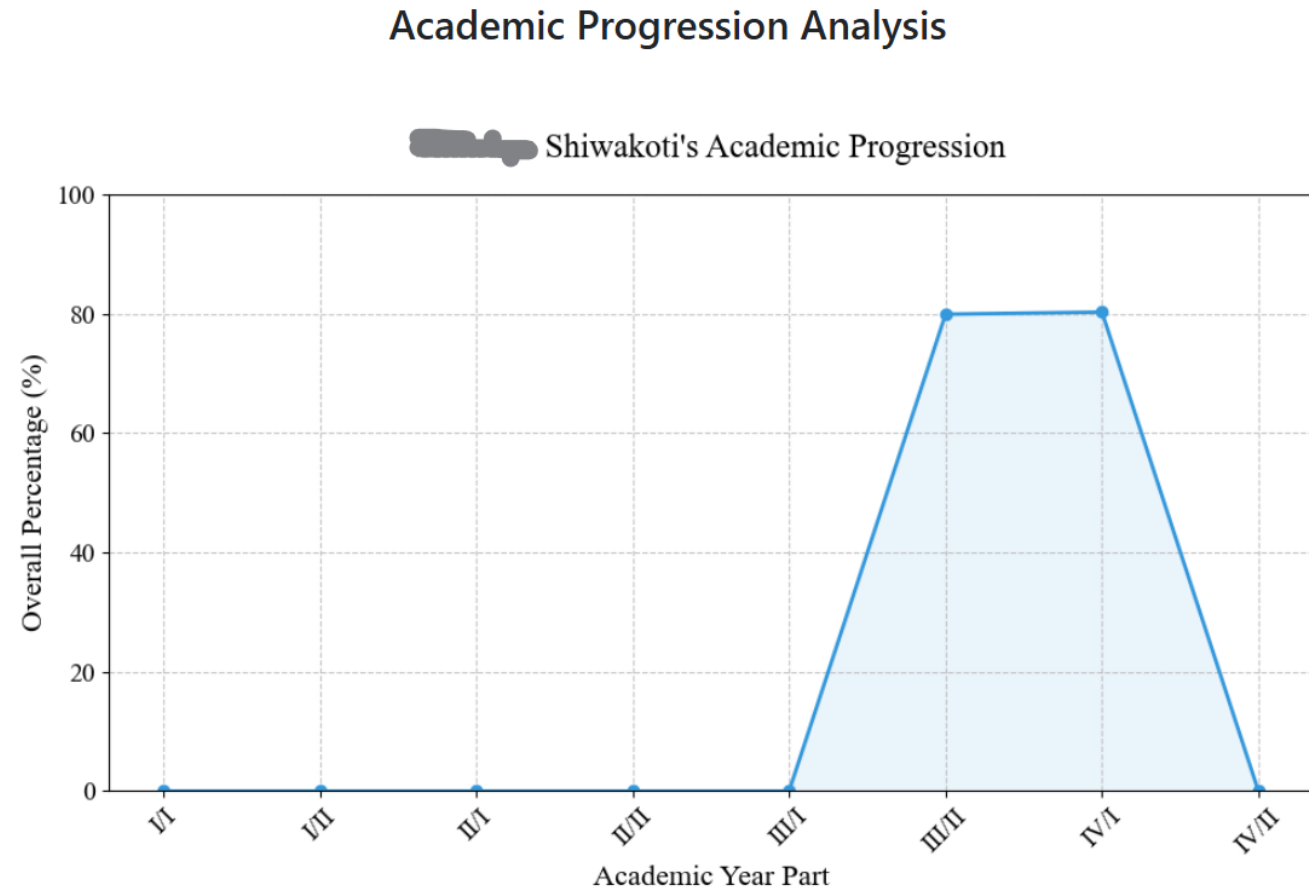
1 / 1 | 100% +

Institute of Engineering, Thapathali Campus
Student name: Shiwakoti
Student ID: 2075/BCT/041
Programme: Computer Engineering
Year/Part: IV/I

Subject Title	Full Marks Ass.	Full Marks Final	Pass Marks Ass.	Pass Marks Final	Marks Ass.	Marks Final	Total Marks
Project Management	20	80	8	32	19.0	65	84.0
Computer Network	20	80	8	32	19.0	58	77.0
Computer Network PRACTICAL	50	0	20	0	46.0	0	46.0
Distributed System	20	80	8	32	19.0	46	65.0
Distributed System PRACTICAL	25	0	10	0	22.0		22.0
Digital Signal Analysis & Processing	20	80	8	32	17.0	67	84.0
Digital Signal Analysis & Processing PRACTICAL	25	0	10	0	24.0		24.0
Project I PRACTICAL	50	0	20	0	46.0		46.0
Energy Environment & Society	10	40	4	16	9.0	23	32.0
Organization & Management	20	80	8	32	19.0	61	80.0
Web Technologies and Applications (Elective I)	20	80	8	32	17.0	62	79.0
Web Technologies and Applications (Elective I) PRACTICAL	25	0	10	0	23.0		23.0

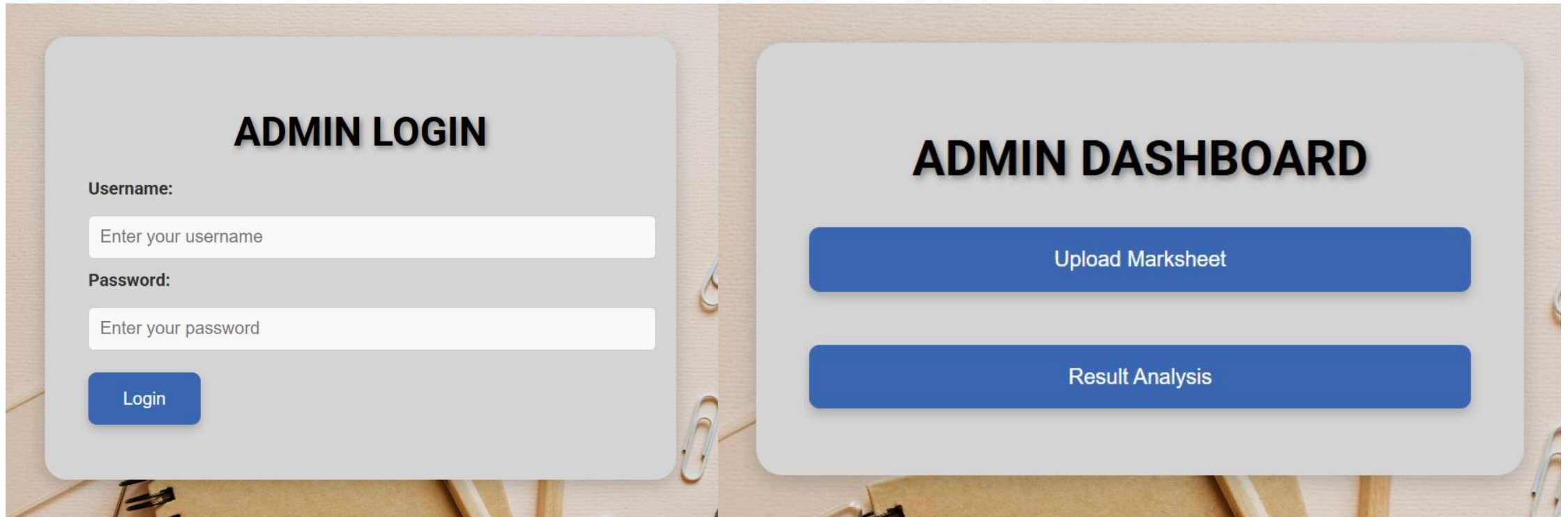
Result and Analysis-[3] (User Interface)

- Student Analysis



Result and Analysis-[3] (User Interface)

- Admin Login Page and Dashboard Page



Result and Analysis-[3] (User Interface)

- Marksheet upload page

UPLOAD MARKSHEET

Marksheet:

Choose File

Image (26).jpg

Upload

Marksheet

Name: **WILSON**

Exam Roll No: 58225

Level: Bachelors in Engineering

Campus: Thapathali Campus

Programme: Electronics, Communication & Information

Subject Code	Subject Title	Marks Assessment	Marks Final	Total Marks
EE460	Electric Circuits & Machines	17.0	A	-

Result and Analysis-[3] (User Interface)

- Result Analysis Page with type selection

Result Analysis Dashboard

Total Students: 85

Select Chart Type:

Enter Year:

Select Year Part:

Programme Pass/Fail

YYYY

-- Select Year Part --

Generate Chart

Reset Filters

Select Chart Type:

Programme Pass/Fail

Programme Pass/Fail

Subject Pass/Fail

Total Students Pass/Fail

Rank of Students

Pass Percentage Trend

Subject Average Marks

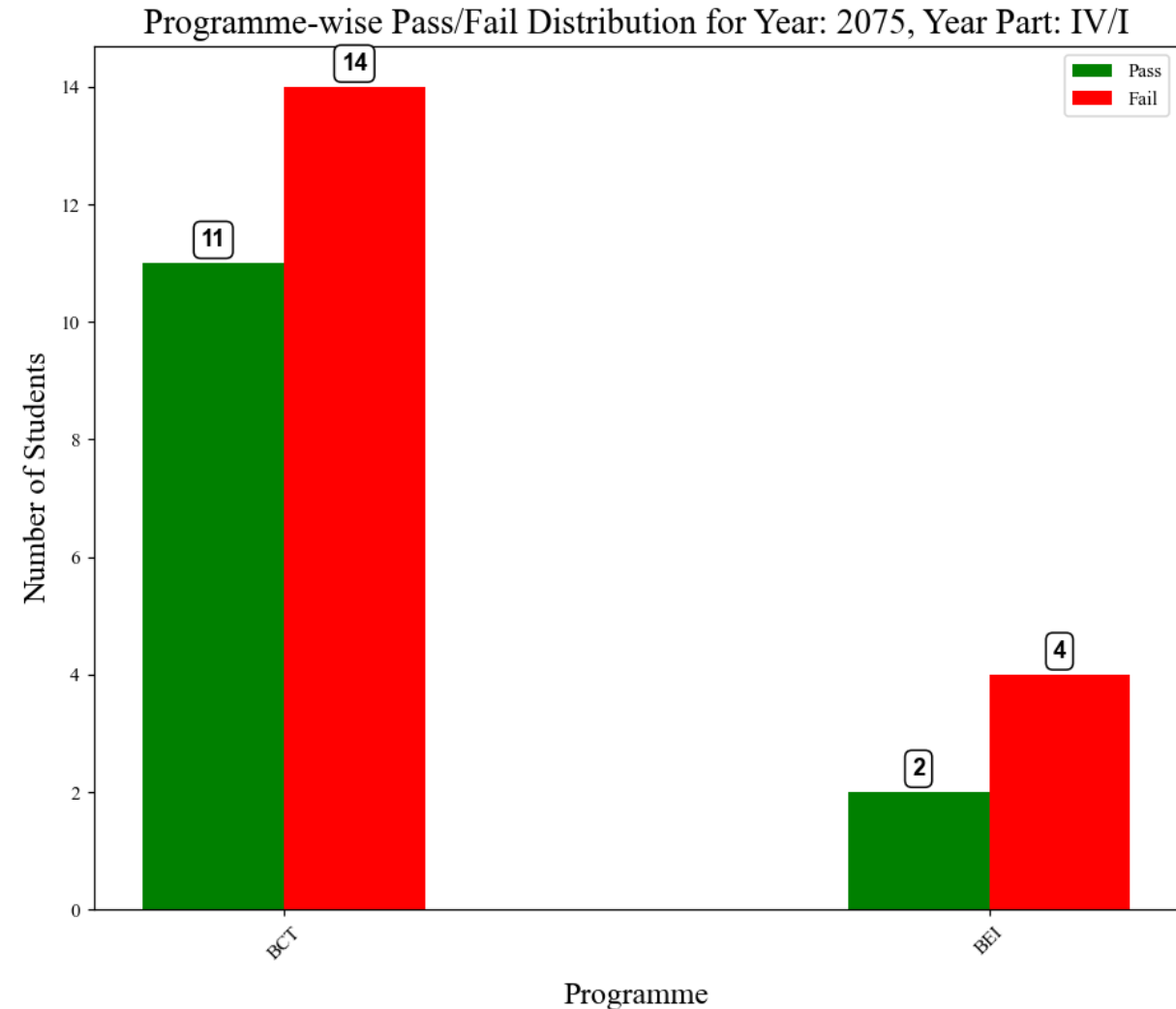
Grade Distribution Analysis

Assessment Vs Final Comparision

Subject Improvement

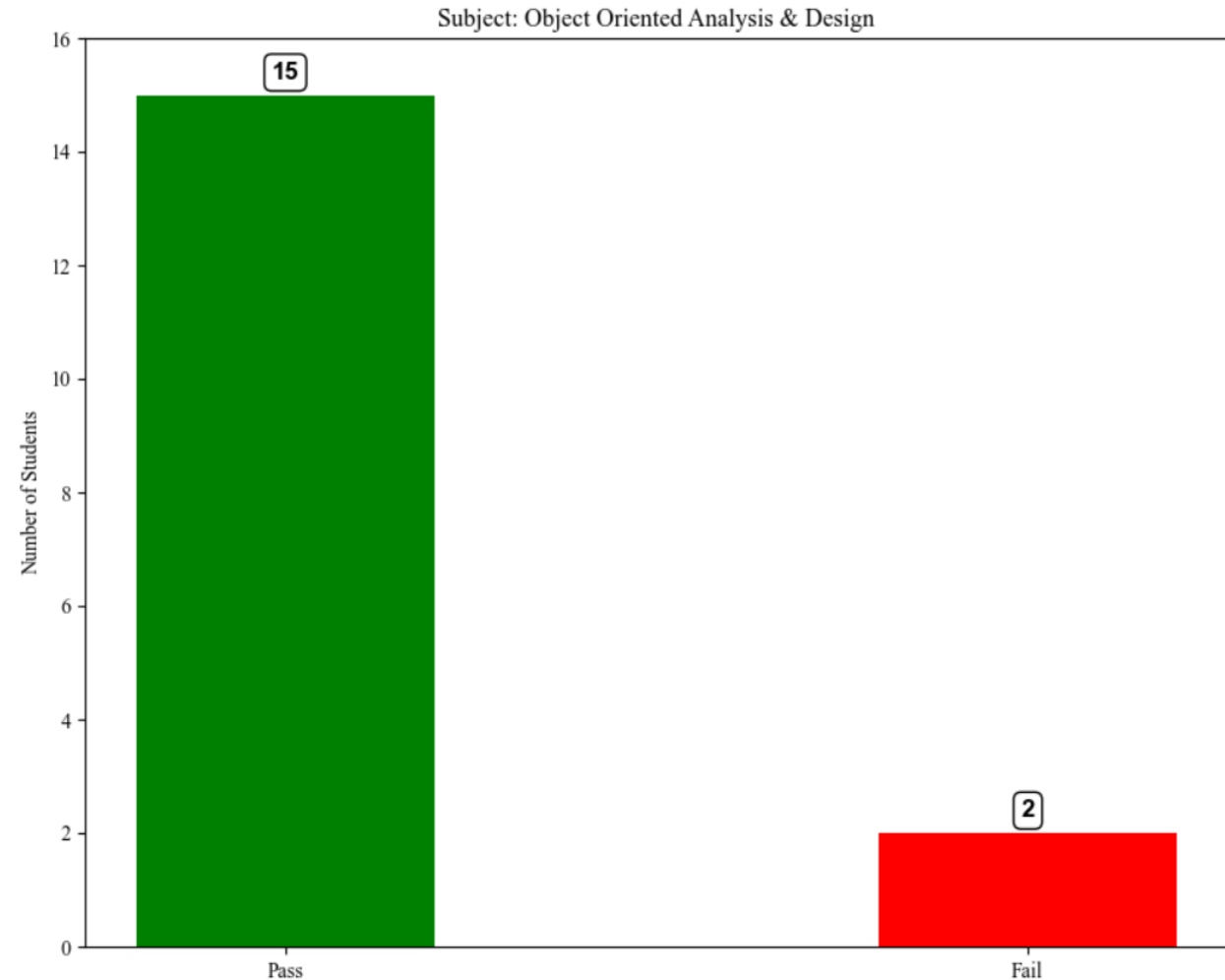
Result and Analysis-[3] (User Interface)

- Programme wise Pass/fail Graph



Result and Analysis-[3] (User Interface)

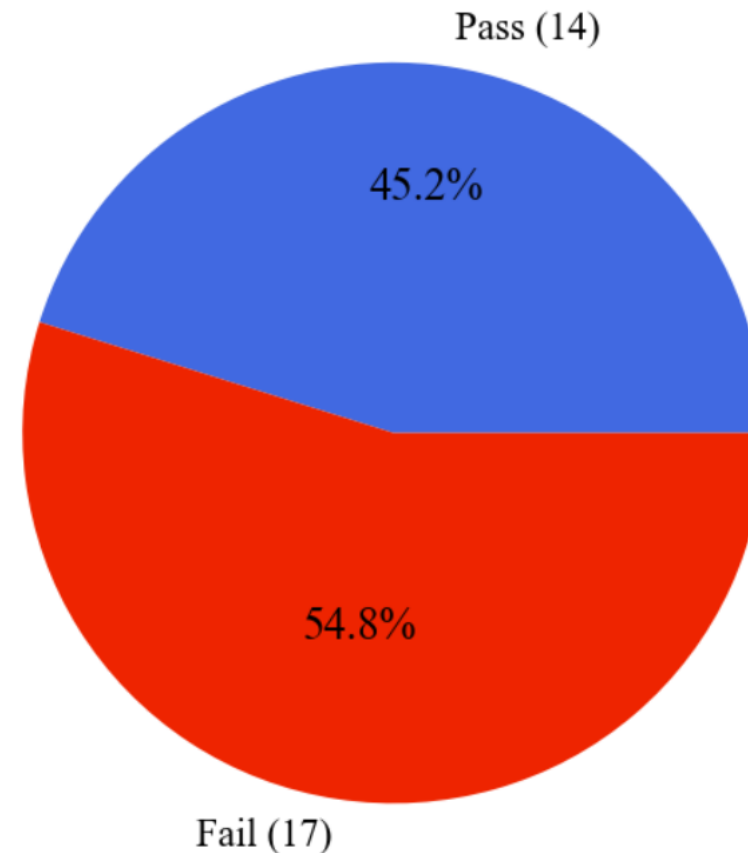
- Subject wise
Pass/fail Graph



Result and Analysis-[3] (User Interface)

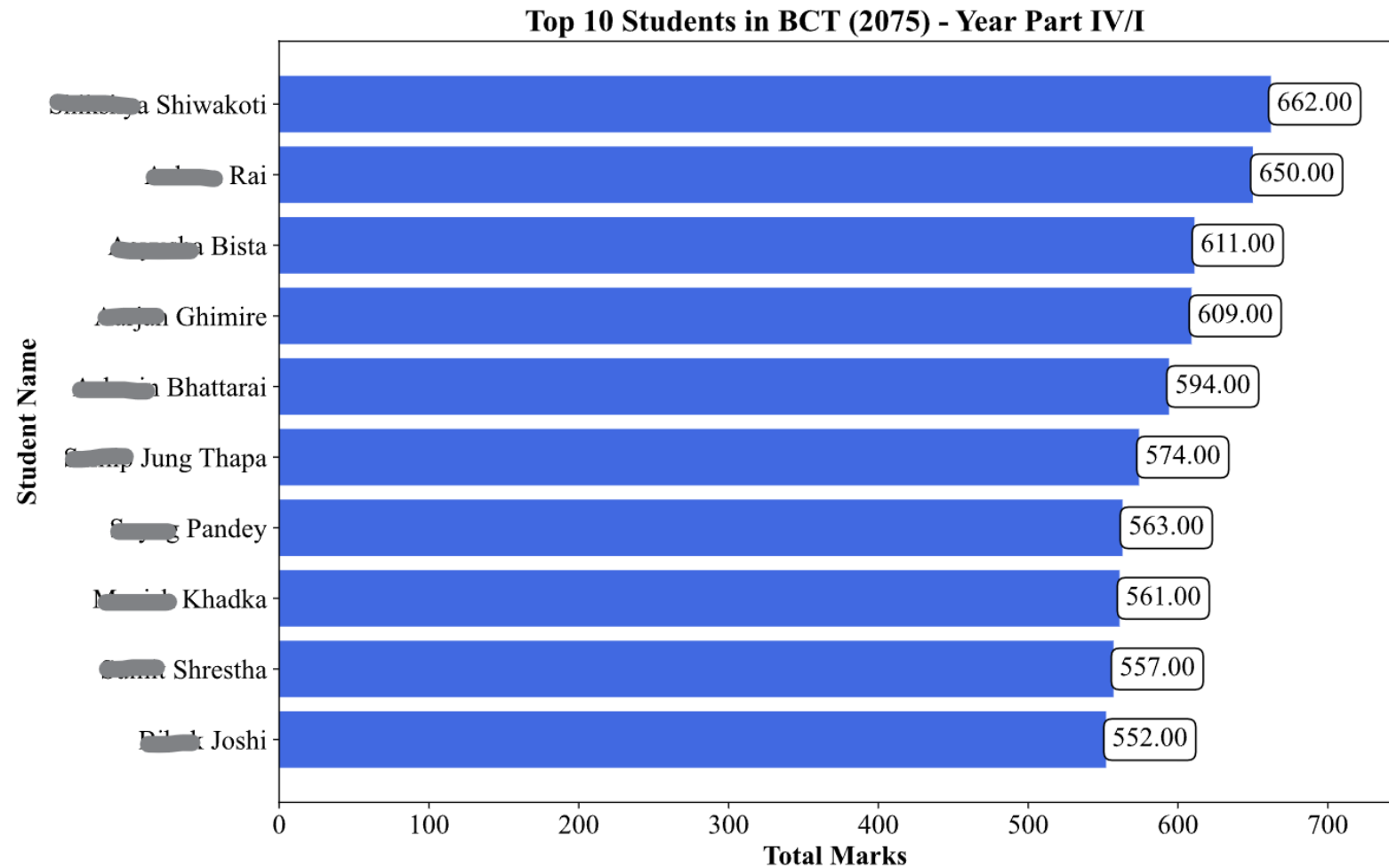
- Semester wise
Pass/fail Pie-chart

Total Students for Year 2075, Year/Part IV/I: 31



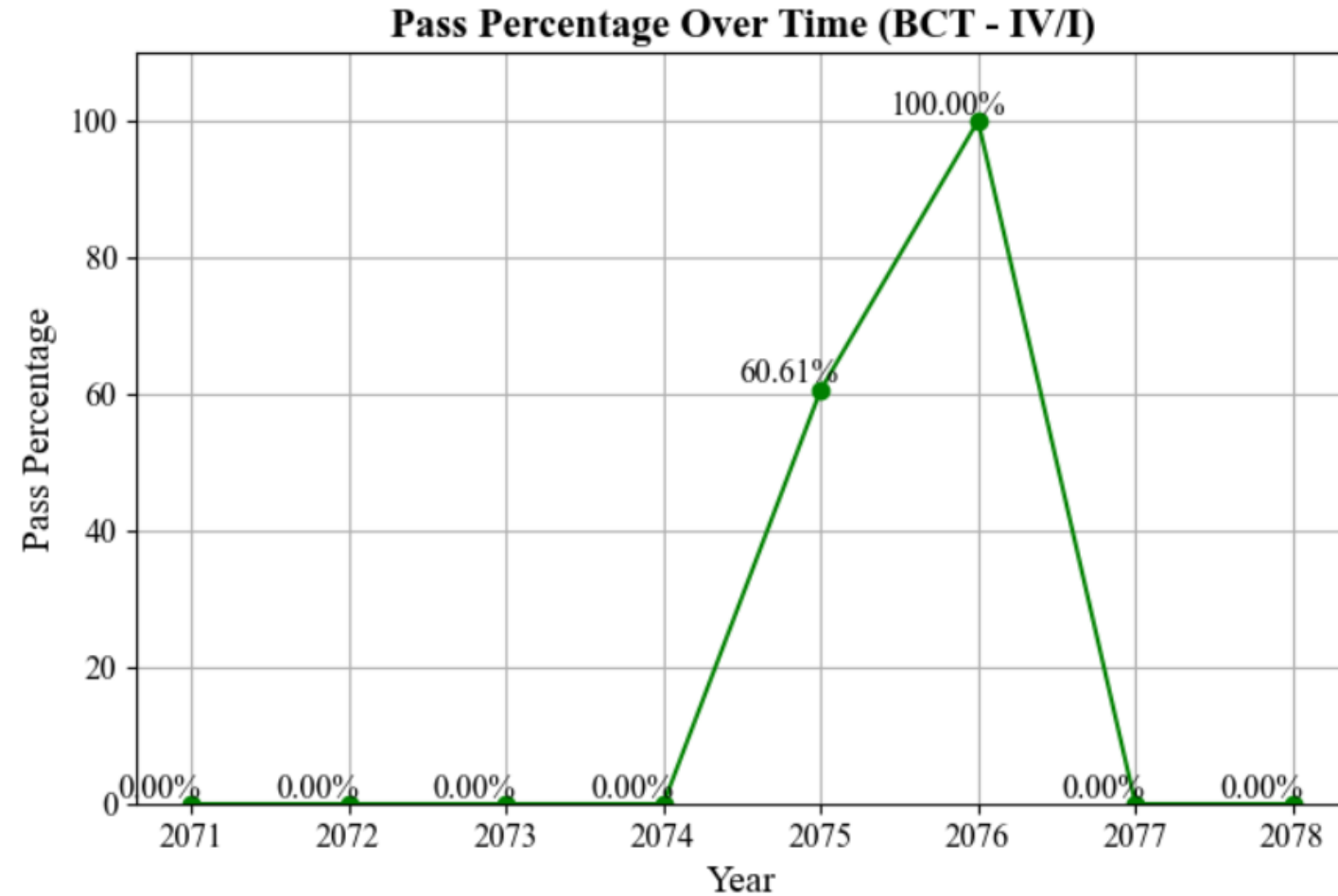
Result and Analysis-[3] (User Interface)

- Rank of Students



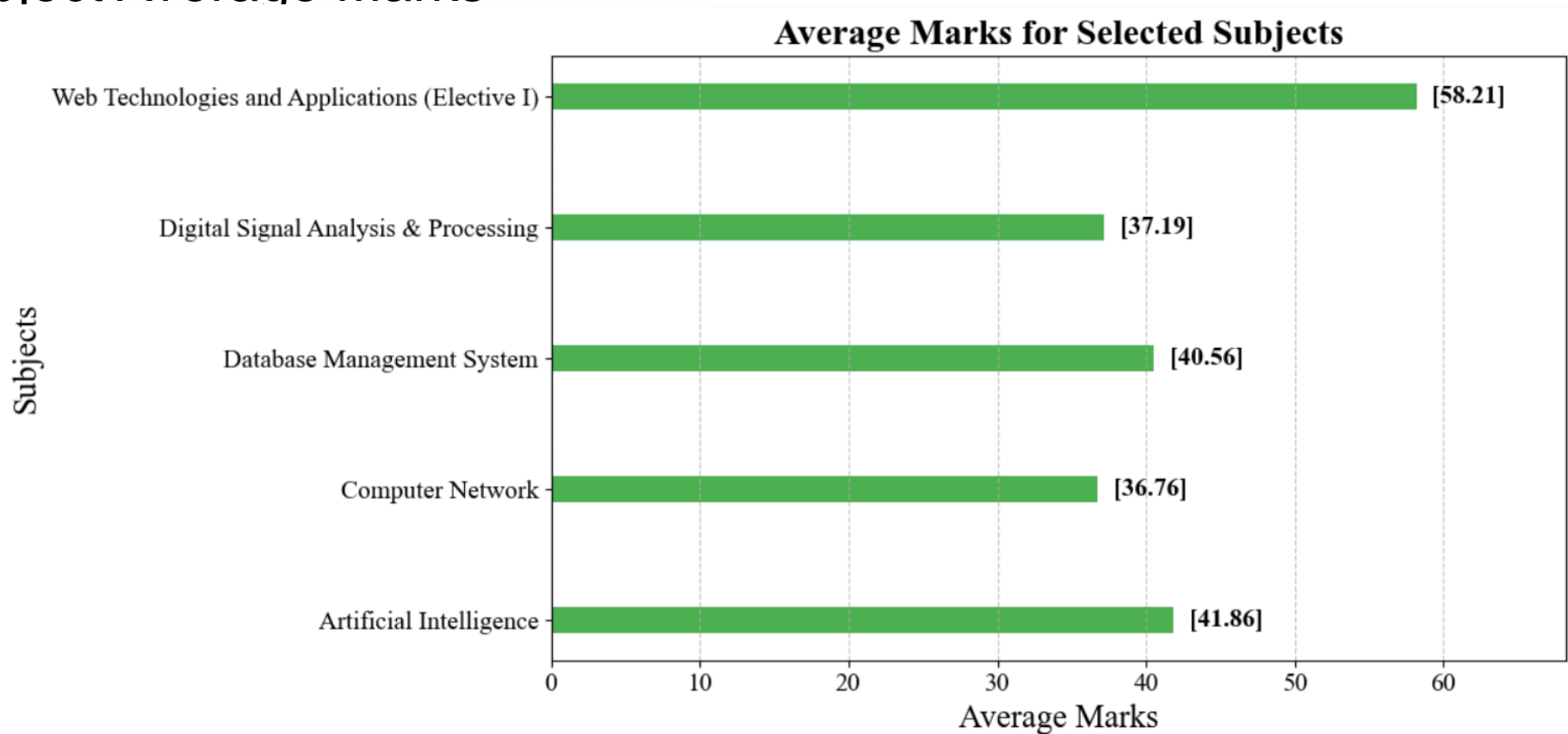
Result and Analysis-[3] (User Interface)

- Pass Percentage Trend



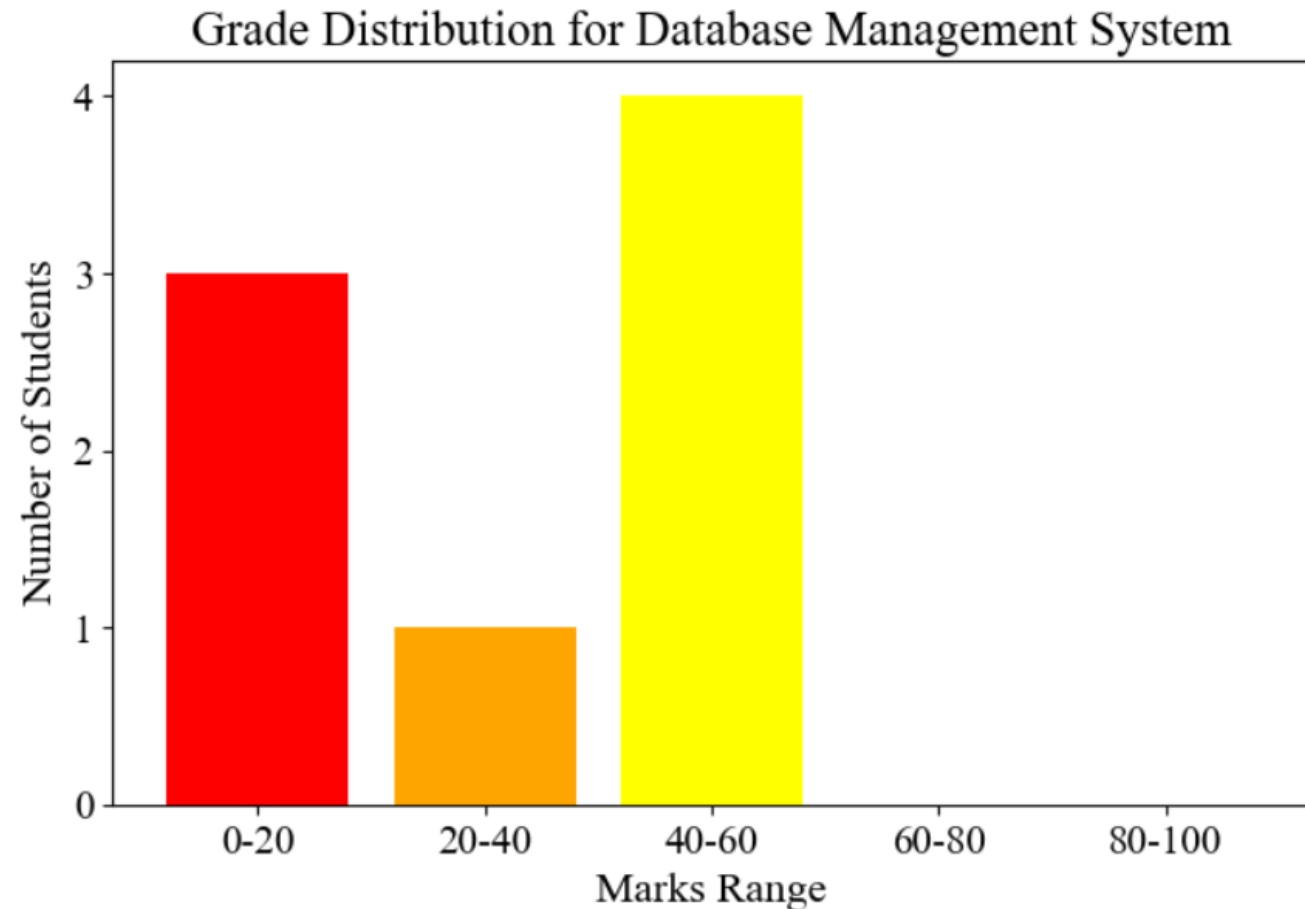
Result and Analysis-[3] (User Interface)

- Subject Average Marks



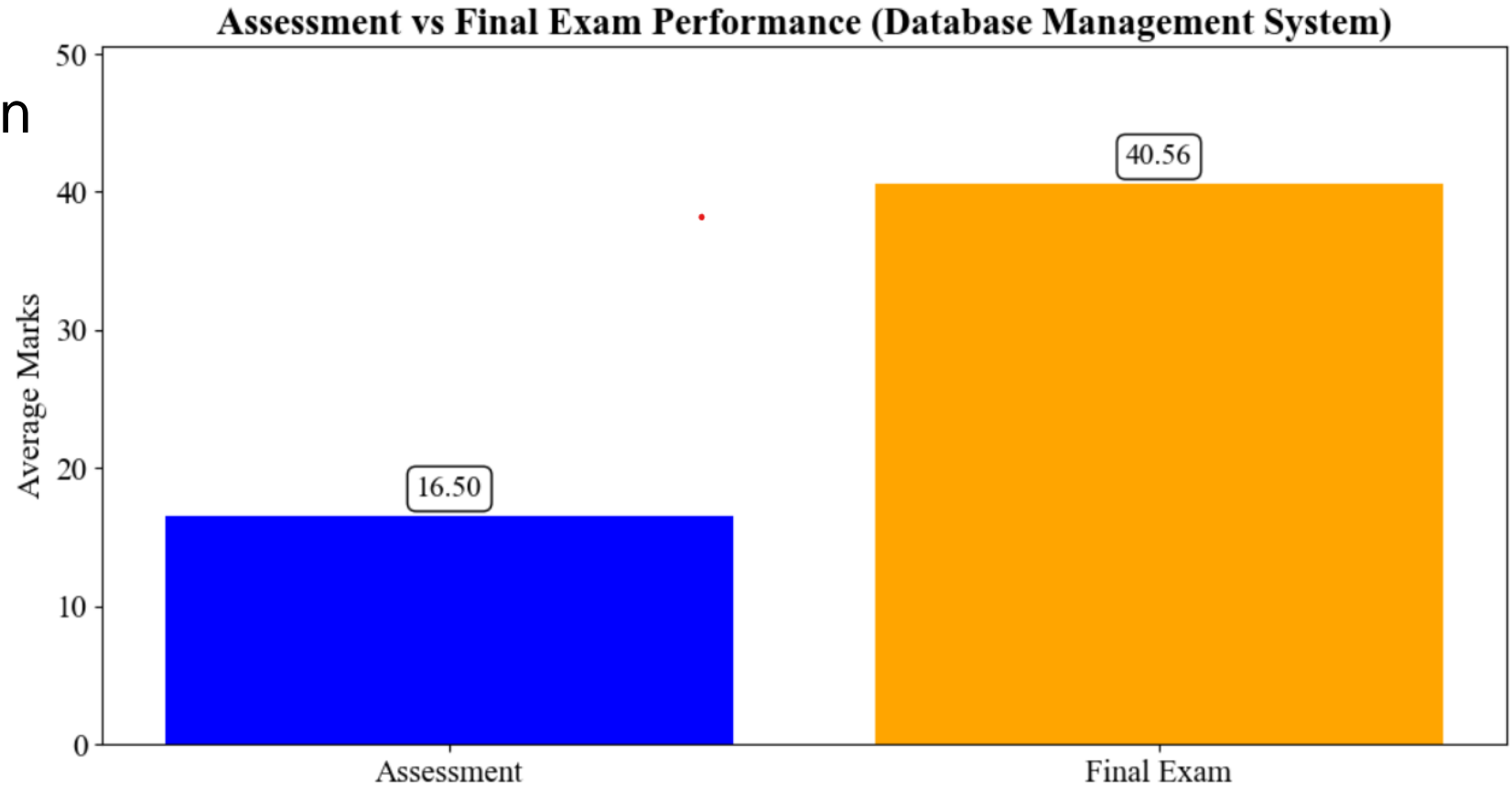
Result and Analysis-[3] (User Interface)

- Grade Distribution Analysis



Result and Analysis-[3] (User Interface)

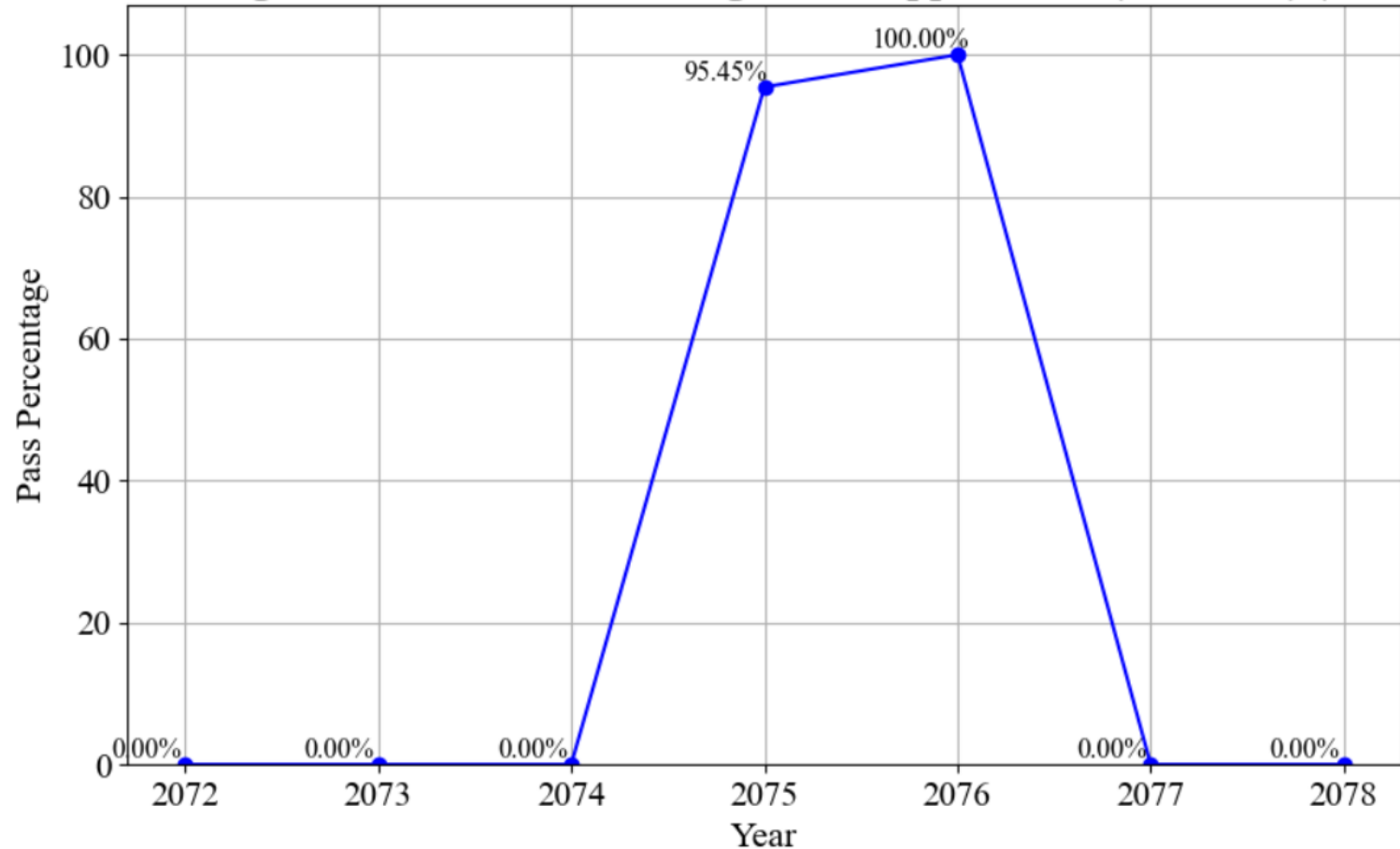
- Assessment vs Final Comparison



Result and Analysis-[3] (User Interface)

- Subject Improvement

Pass Percentage Trend for Web Technologies and Applications (Elective I) (2072-2078)



Result and Analysis-[4]

(Database Schema and Instances)

- Student information table

studen...	name	level	campus	exam_rol...	programme
Filter...	Filter...	Filter...	Filter...	Filter...	Filter...
2075/BCT/037	Samriddha Shrestha	Bachelors in Engineering	Thapathali Campus	72254	Computer Engineering
2075/BEV/009	Bhuwan Khatiwada	Bachelors in Engineering	Thapathali Campus	52208	Electronics Communication & Information
2072/BEX/310	Bibek Dhakal	Bachelors in Engineering	Thapathali Campus	70852	Electronics & Communication Engineering
2075/BE/006	Anjal Bam	Bachelors in Engineering	Thapathali Campus	52205	Electronics, Communication & Information
2078/BEI/001	Aayush Chhetri	Bachelors in Engineering	Thapathali Campus	58223	Electronics, Communication & Information
2078/BEI/015	Diwas Dahal	Bachelors in Engineering	Thapathali Campus	58225	Electronics, Communication & Information
2072/BEX/347	Sushmita Bhatt	Bachelors in Engineering	Thapathali Campus	70855	Electronics & Communication Engineering
2073/BEX/316	Kapalik Khanal	Bachelors in Engineering	Thapathali Campus	54852	Electronics & Communication Engineering
2074/BEX/036	Satkrit Raj Pandey	Bachelors in Engineering	Thapathali Campus	70862	Electronics & Communication Engineering
2075/BEL/009	Bhuwan Khatiwada	Bachelors in Engineering	Thapathali Campus	58201	Electronics Communication & Information
2077/BEI/009	Ashim Panthi	Bachelors in Engineering	Thapathali Campus	58215	Electronics Communication & Information
2075/BEI/003	Abhishek Poudel	Bachelors in Engineering	Thapathali Campus	73201	Electronics Communication & Information
2076/BEI/005	Anuka K.C.	Bachelors in Engineering	Thapathali Campus	73207	Electronics Communication & Information
2076/BEI/007	Ayush Acharya	Bachelors in Engineering	Thapathali Campus	73208	Electronics Communication & Information
2076/BEI/008	Anupam Bhattarai	Bachelors in Engineering	Thapathali Campus	73209	Electronics Communication & Information
2075/BCT/033	Prashant Bhusal	Bachelors in Engineering	Thapathali Campus	72253	Computer Engineering
2072/BEX/308	Anil Tamang	Bachelors in Engineering	Thapathali Campus	70851	Electronics & Communication Engineering
2073/BEX/311	Gokul Adhikari	Bachelors in Engineering	Thapathali Campus	70856	Electronics & Communication Engineering
2073/BEX/314	James Shrestha	Bachelors in Engineering	Thapathali Campus	70857	Electronics & Communication Engineering
2074/BEX/002	Abhinit Kumar Das	Bachelors in Engineering	Thapathali Campus	70858	Electronics & Communication Engineering
2074/BEX/025	Pukar Giri	Bachelors in Engineering	Thapathali Campus	70859	Electronics & Communication Engineering
2075/BCT/023	Kumar Tiwari	Bachelors in Engineering	Thapathali Campus	65804	Computer Engineering

Result and Analysis-[4]

(Database Schema and Instances)

- Marks obtained table

id	year_part	marks_as...	marks_final	studen...	title_id	total_mar...
54	I/II	12.0	A	2075/BCT/037	Engineering Economics	-
55	I/II	15.0	A	2075/BCT/037	Object Oriented Analysis & Design	-
56	I/II	14.0	A	2075/BCT/037	Database Management System	-
57	I/II	17.0	A	2075/BCT/037	Artificial Intelligence	-
58	I/II	16.0	A	2075/BCT/037	Embedded System	-
59	I/II	16.0	A	2075/BCT/037	Operating System	-
60	IVA	14.0	32	2075/BEV/009	Digital Signal Analysis & Processing	46.0
61	IVA	22.0	"	2075/BEV/009	Digital Signal Analysis & Processing PRACTICAL	22.0
62	IVA	15.0	41	2075/BEV/009	Artificial Intelligence	56.0
63	IVA	24.0	"	2075/BEV/009	Artificial Intelligence PRACTICAL	24.0
64	IVA	44.0	"	2075/BEV/009	Project I PRACTICAL	44.0
65	IVA	15.0	32	2075/BEV/009	RF& Microwave Engineering	47.0
66	IVA	24.0	"	2075/BEV/009	RF & Microwave Engineering PRACTICAL	24.0
67	IVA	18.0	45	2075/BEV/009	Organization & Management	63.0
68	IVA	21.0	"	2075/BEV/009	Web Technologies and Applications (Elective I PRACTICAL	21.0
69	II/I	15.0	A	2075/BCT/037	Object Oriented Programming	-
70	II/I	17.0	A	2075/BCT/037	Theory of Computation	-
71	II/I	12.0	A	2075/BCT/037	Electric Circuit Theory	-
72	II/I	15.0	A	2075/BCT/037	Electronic Devices & Circuits	-
73	II/I	13.0	A	2075/BCT/037	Digital Logic	-
74	II/I	18.0	A	2075/BCT/037	Electromagnetics	-
75	II/I	13.0	A	2075/BCT/037	Engineering Mathematics III	-

Result and Analysis-[4]

(Database Schema and Instances)

- Subject information table

subject_c...	title	full_mark...	full_mark...	pass_mar...	pass_mar...
Filter	Filter...	Filter	Filter	Filter	Filter
CE655	Engineering Economics	20	80	8	32
CT651	Object Oriented Analysis & Design	20	80	8	32
CT652	Database Management System	20	80	8	32
CT653	Artificial Intelligence	20	80	8	32
CT655	Embedded System	20	80	8	32
CT656	Operating System	20	80	8	32
CT704	Digital Signal Analysis & Processing	20	80	8	32
CT704	Digital Signal Analysis & Processing PRACTICAL	25	0	10	0
CT710	Artificial Intelligence PRACTICAL	25	0	10	0
EX707	Project I PRACTICAL	50	0	20	0
EX716	RF& Microwave Engineering	20	80	8	32
EX716	RF & Microwave Engineering PRACTICAL	25	0	10	0
ME708	Organization & Management	20	80	8	32
72505	Web Technologies and Applications (Elective I PRACTICAL	25	0	10	0
CT451	Object Oriented Programming	20	80	8	32
CT502	Theory of Computation	20	80	8	32
EE501	Electric Circuit Theory	20	80	8	32
X501	Electronic Devices & Circuits	20	80	8	32
EX502	Digital Logic	20	80	8	32
EX503	Electromagnetics	20	80	8	32
SH501	Engineering Mathematics III	20	80	8	32
EX601	Advanced Electronics	20	80	8	32

Result and Analysis-[5] (Comparison Between PaddleOCR and Tesseract)

Table Extracted by Paddle:

Subjects		Full Marks		Pass Marks		Marks Obtained		Total
Code	Title	Asst.	Final	Asst.	Final	Asst.	Final	
CT701	Project Management	20	80	8	32	18	46	64
CT702	Computer Network	20	80	8	32	12	32	44
CT702	Computer Network PRACTICAL	50		20		35		35
CT703	Distributed System	20	80	8	32	18	40	58
CT703	Distributed System PRACTICAL	25		10		24		24
CT704	Digital Signal Analysis & Processing	20	80	8	32	19	47	66
CT704	Digital Signal Analysis & Processing PRACTICAL	25		10		25		25
CT707	Project I PRACTICAL	50		20		45		45
EX701	Energy Environment & Society	10	40	4	16	6	24	30
ME708	Organization & Management	20	80	8	32	18	45	63
CT72502	Data Mining (Elective I)	20	80	8	32	14	45	59
CT72502	Data Mining (Elective I) PRACTICAL	25		10		20		20

Table Extracted by Tesseract:

Code	Title	Asst.	Final	Asst.	Final	Asst.	Final	Total	Remarks
CT701	Project Management	20	80	8	32	18	46	64	
Sia	Computer Network	20	80	8	32	12	32	44	
CT702	Computer Network PRACTICAL SO oes Ge ee tC								
CT703	Distributed System	20	80	8	32	18	40	58	
CT703	Distributed System PRACTICAL	25	—	10	2s	24	24		
CT704	Digital Signal Analysis & Processing	20	80	8	32	19	47	66	
CT704	Digital Signal Analysis & Processing	25	—	10	—	25	—	25	
PRACTICAL									
CT707	Project	1	PRACTICAL	50	—	20	—	45	— 45
i									
EX701	Energy Environment & Society	10	40	4	16	6	24	30	
ME708	Organization & Management	20	80	8	32	18	45	63	
CT72502	Data Mining Elective I	20	80	8	32	14	45	59	i
CT72502	Data Mining Elective I PRACTICAL	25	—	10	—	20	20		

Result and Analysis-[5]

(Character Accuracy Rate)

- **Dataset:** 10 marksheets from the 2075 batch, IV/I part
- **Comparison:** Extracted text vs. Ground truth
- **Error Detection:** Used Levenshtein distance to identify:
Substitutions, Deletions, Insertions
- **Average CAR: 94.62%** (High OCR accuracy)

Future Enhancement

- **Integration of Student Records for Enhanced Analysis in QAA Reports**
 - Enables multiparameter analysis of Exam Result
- **Enhancing Accuracy with Paid OCR Integration**
 - Improves text extraction accuracy
 - Better handling of low-quality images

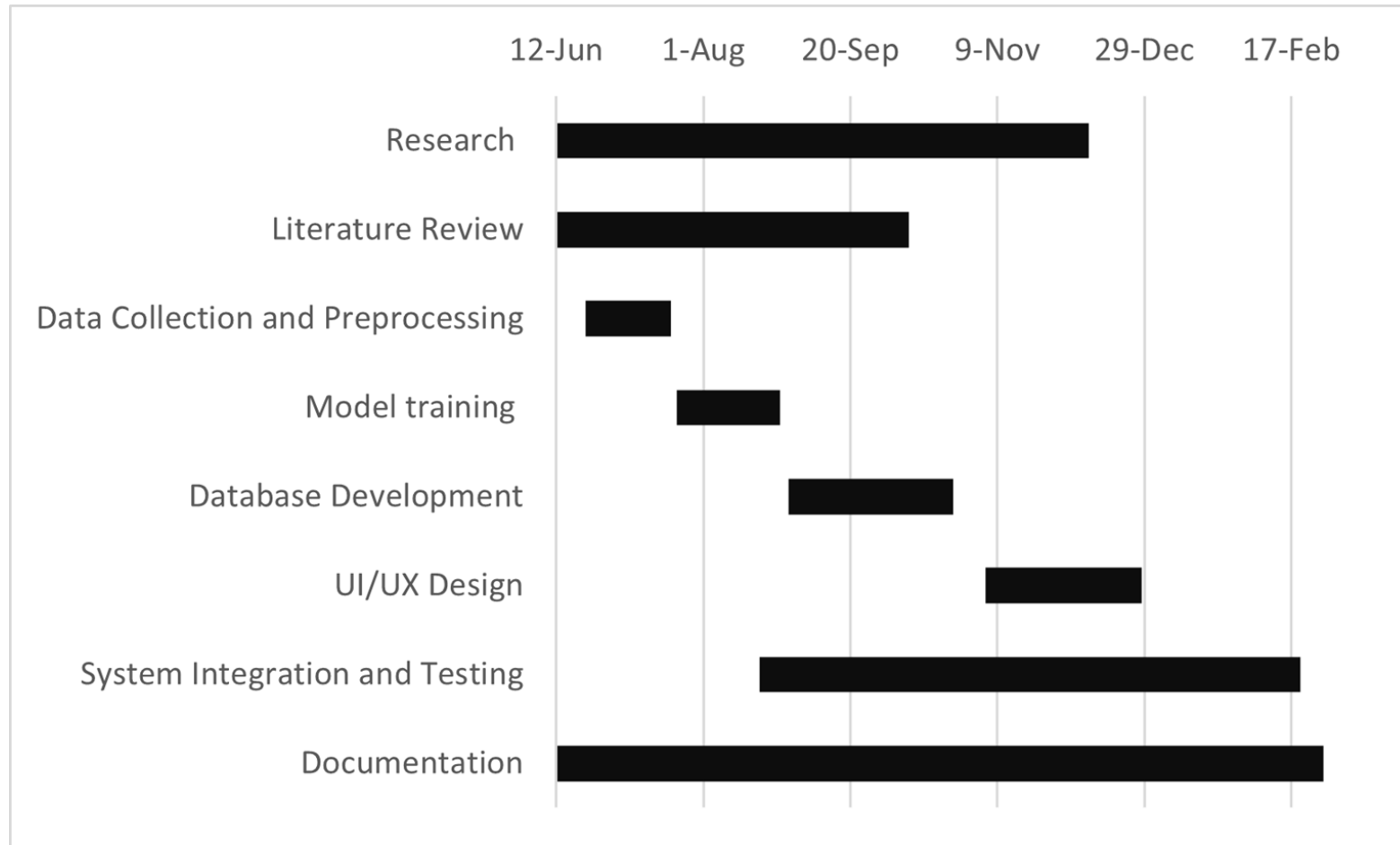
Application

- In sectors requiring efficient management of large volumes of printed data such as government agencies and private organizations
- In healthcare facilities to digitize patient records.
- In banking for processing financial documents.
- In local shops and stores for managing handwritten/digital bills

Conclusion

- The system successfully automates marksheet processing and report generation, improving efficiency and accuracy in academic record management.
- It enhances accessibility by allowing admins to manage student records and generate reports while enabling students to easily view and download their results.
- Future improvements, such as better OCR handling for noisy marksheets and advanced data visualization, will further refine the system's performance and insights.

Project Schedule



Project Expenses

Particulars	Price
Miscellaneous	10000
Total	10000

References

- P. Pyreddy and W. B. Croft , "TINTIN: A System for Retrieval in Text Tables," Proceedings of the second ACM international conference on Digital libraries, 1997.
- T. Kasar, P. Barlas, S. Adam, C. Chatelain and T. Paquet, "Learning to Detect Tables in Scanned Document Images Using Line Information," 2013 12th International Conference on Document Analysis and Recognition. IEEE, 2013.
- T. Ojala, M. Pietikäinen and T. & Mäenpää, "Gray scale and rotation invariant texture classification with local binary patterns," Springer Berlin Heidelberg, 2000.

References(Contd...)

- K. Wang, B. Babenko and S. Belongie, "End-to-end scene text recognition.," 2011 International conference on computer vision. IEEE, 2011.
- D. N. Tran, T. A. Tran, A. Oh, S. H. Kim and I. S. & Na, "Table detection from document image using vertical arrangement of text blocks," International Journal of Contents, 2015.
- M. Stonebraker, L. A. Rowe and M. Hirohama, "The Implementation Of Postgres,"1990.[Online].Available:
https://www.researchgate.net/publication/3296158_The_Implementation_Of_Postgres.

THANK YOU