The Report is Generated by DrillBit Plagiarism Detection Software

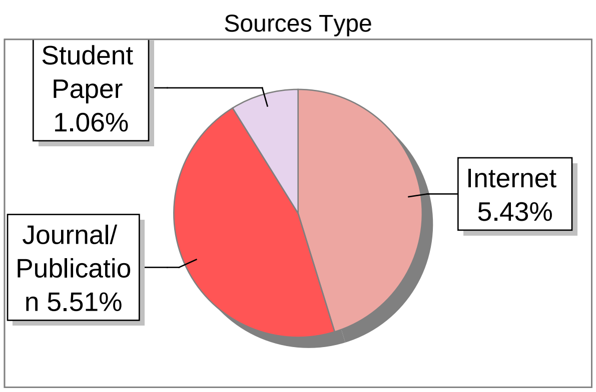
### *Submission Information*

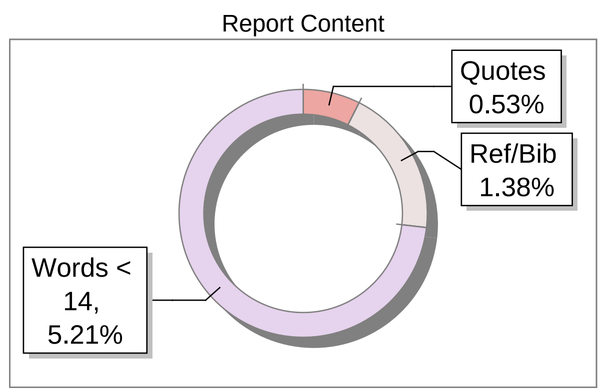
| | |
|---|---|
| Author Name | Saloni Navgire |
| Title | Automated Answar sheet evalution system |
| Paper/Submission ID | 5013820 |
| Submitted by | rajurkar_am@mgmcen.ac.in |
| Submission Date | 2025-12-19 16:06:56 |
| Total Pages, Total Words | 48, 11037 |
| Document type | Project Work |

### *Result Information*

Similarity **12 %**

1 10 20 30 40 50 60 70 80 90




### *Exclude Information*

| | |
|---|---|
| Quotes | Not Excluded |
| References/Bibliography | Not Excluded |
| Source: Excluded < 14 Words | Not Excluded |
| Excluded Source | **0 %** |
| Excluded Phrases | Not Excluded |

### *Database Selection*

| | |
|---|---|
| Language | English |
| Student Papers | Yes |
| Journals & publishers | Yes |
| Internet or Web | Yes |
| Institution Repository | Yes |

A Unique QR Code use to View/Download/Share Pdf File

DrillBit

**DrillBit Similarity Report**

| 12 | 84 | B | A-Satisfactory (0-10%)<br>B-Upgrade (11-40%)<br>C-Poor (41-60%)<br>D-Unacceptable (61-100%) |
|---|---|---|---|
| SIMILARITY % | MATCHED SOURCES | GRADE | |

| LOCATION | MATCHED DOMAIN | % | SOURCE TYPE |
|---|---|---|---|
| 1 | savedelicious.com | 1 | Internet Data |
| 2 | www.setproduct.com | <1 | Internet Data |
| 3 | gcet.edu.in | <1 | Publication |
| 4 | REPOSITORY - Submitted to Exam section VTU on 2024-07-31 16-28 951608 | <1 | Student Paper |
| 5 | ijarcce.com | <1 | Publication |
| 6 | mite.ac.in | <1 | Publication |
| 7 | ijarcce.com | <1 | Publication |
| 8 | REPOSITORY - Submitted to Exam section VTU on 2024-07-31 15-55 914974 | <1 | Student Paper |
| 9 | translate.google.com | <1 | Internet Data |
| 10 | ia800202.us.archive.org | <1 | Publication |
| 11 | journals.lww.com | <1 | Internet Data |
| 12 | trackpost.online | <1 | Internet Data |
| 13 | www.slideshare.net | <1 | Internet Data |

| | | | |
|---|---|---|---|
| 14 | Thesis Submitted to Shodhganga, shodhganga.inflibnet.ac.in | <1 | Publication |
| 15 | www.ijirset.com | <1 | Publication |
| 16 | Application support to product variety management by Forza-2007 | <1 | Publication |
| 17 | Student Archives Data | <1 | Student Paper |
| 18 | ACM Press the 7th International Conference- Ume229, Sweden (200, by Snchez, Ivn Riek- 2008 | <1 | Publication |
| 19 | businessdocbox.com | <1 | Internet Data |
| 20 | Thesis Submitted to Shodhganga Repository | <1 | Publication |
| 21 | abnews-wire.blogspot.com | <1 | Internet Data |
| 22 | www.svec.education | <1 | Publication |
| 23 | epj-conferences.org | <1 | Internet Data |
| 24 | pdfs.semanticscholar.org | <1 | Publication |
| 25 | Thesis Submitted to Shodhganga Repository | <1 | Publication |
| 26 | www.breadcrumbdigital.com.au | <1 | Internet Data |
| 27 | ijarcce.com | <1 | Publication |
| 28 | index-of.es | <1 | Publication |
| 29 | Usage of Probabilistic and General Regression Neural Network for Early, By Sharma, Neha Om, Hari, Yr-2015 | <1 | Publication |
| 30 | www.studocu.com | <1 | Internet Data |
| 31 | globallandscapesforum.org | <1 | Internet Data |
| 32 | Systems Engineering Building Successful Systems by Eisner-2011 | <1 | Publication |

| | | | |
|---|---|---|---|
| 33 | www.svce.ac.in | <1 | Publication |
| 34 | 618media.com | <1 | Internet Data |
| 35 | Comparison of automatic methods to detect sunspots in the Coimbra Observatory sp by Carvalho-2020 | <1 | Publication |
| 36 | www.fraud.com | <1 | Internet Data |
| 37 | www.linkedin.com | <1 | Internet Data |
| 38 | Comparison of Molecular Typing Methods for Characterization of Staphylococcus ep by Miragaia-2007 | <1 | Publication |
| 39 | documents.mx | <1 | Internet Data |
| 40 | figshare.com | <1 | Publication |
| 41 | files.eric.ed.gov | <1 | Publication |
| 42 | moldstud.com | <1 | Internet Data |
| 43 | REPOSITORY - Submitted to Exam section VTU on 2024-07-31 15-53 915096 | <1 | Student Paper |
| 44 | springeropen.com | <1 | Internet Data |
| 45 | www.ijraset.com | <1 | Publication |
| 46 | www.indeed.com | <1 | Internet Data |
| 47 | www.ncbi.nlm.nih.gov | <1 | Internet Data |
| 48 | 1library.co | <1 | Internet Data |
| 49 | Are biofuels good for African development An analytical framework with evidence by Arndt-2011 | <1 | Publication |
| 50 | article.sciencepg.net | <1 | Publication |

| | | | |
|---|---|---|---|
| 51 | arxiv.org | <1 | Internet Data |
| 52 | arxiv.org | <1 | Publication |
| 53 | A Windows application for computing standardized mortality ratios and by Geiss-213 | <1 | Publication |
| 54 | biomedcentral.com | <1 | Internet Data |
| 55 | bip.uni.wroc.pl | <1 | Publication |
| 56 | carmencloud.com | <1 | Internet Data |
| 57 | Chemical-Looping Combustion and Chemical-Looping Reforming in a Circulating Flui by Ryden-2008 | <1 | Publication |
| 58 | codeburst.io | <1 | Internet Data |
| 59 | COMPARATIVE ANALYSIS OF EVALUATION CRITERIA FOR UNPAVED RURAL ROADS A CASE STU By Carlos Javier Vasquez Montero, Yr-2024,3 | <1 | Publication |
| 60 | core.ac.uk | <1 | Internet Data |
| 61 | docs.horizon3.ai | <1 | Publication |
| 62 | documents.mx | <1 | Internet Data |
| 63 | eduhk.hk | <1 | Internet Data |
| 64 | eprints.umm.ac.id | <1 | Internet Data |
| 65 | ftb.com.hr | <1 | Internet Data |
| 66 | ijarcce.com | <1 | Publication |
| 67 | informatics.nic.in | <1 | Publication |
| 68 | journal.penerbitjurnal.com | <1 | Publication |

| | | | |
|---|---|---|---|
| 69 | medium.com | <1 | Internet Data |
| 70 | moam.info | <1 | Internet Data |
| 71 | pmc.ncbi.nlm.nih.gov | <1 | Internet Data |
| 72 | Submitted to U-Next Learning on 2025-01-31 09-05 3198640 | <1 | Student Paper |
| 73 | Thesis Submitted to Shodhganga Repository | <1 | Publication |
| 74 | Thesis Submitted to Shodhganga Repository | <1 | Publication |
| 75 | translate.google.com | <1 | Internet Data |
| 76 | translate.google.com | <1 | Internet Data |
| 77 | www.edu.su.se | <1 | Publication |
| 78 | www.fraud.com | <1 | Internet Data |
| 79 | www.ijsat.org | <1 | Publication |
| 80 | www.libraryassessment.org | <1 | Publication |
| 81 | www.mdpi.com | <1 | Internet Data |
| 82 | www.planningmalaysia.org | <1 | Publication |
| 83 | www.scaler.com | <1 | Internet Data |
| 84 | IEEE 2019 2nd International Conference on new Trends in Computing Sc | <1 | Publication |

*Chapter 1*

# INTRODUCTION

Education plays a fundamental role in shaping the intellectual and professional development of individuals. One of the most important components of the educational process is the evaluation of student performance. For decades, educational institutions have relied heavily on the conventional method of manual answer sheet checking. In this process, teachers carefully read every answer written by the student, interpret the content, compare it with the expected solution, and assign marks accordingly. Although this method has been followed for generations and is widely accepted, it presents numerous problems in the modern context where academic pressure, student population, and examination frequency continue to increase.

Manual evaluation is slow and requires extensive human labor. A single teacher may need to check hundreds of answer sheets within a limited timeframe, especially during final examinations. This repetitive and exhaustive process causes fatigue, which naturally affects consistency and judgment. Even the most experienced teachers may unintentionally overlook key points or fail to maintain uniformity in marking criteria. These issues highlight an urgent need for improvement, especially considering the growing scale of educational systems.

In recent years, the digitization of education has rapidly increased. Institutions are adopting online classes, digital notes, computer-based exams, and automated administrative systems. In this evolution, one area that still remains largely manual is the evaluation of handwritten answer sheets. The lack of automation in this area slows down result generation, affects accuracy, and places a heavy burden on teachers. With advancements in technology such as Optical Character Recognition (OCR), machine learning, image processing, and automated text extraction, a new opportunity has emerged to modernize the assessment process. Technologies like Tesseract OCR, PyMuPDF, OpenCV, and intelligent web-based platforms enable computers to read, understand, and process scanned answer sheets in a more efficient manner.

The project titled "Digital Marking System for Scanning Answer Sheets" is designed as a step toward bridging the gap between traditional manual assessment and modern digital processing. The aim of this project is not to replace teachers but to assist them by automating repetitive tasks, such as reading handwritten text, extracting important details, identifying answer types, and preparing structured content for evaluation. By reducing manual effort and increasing accuracy, the system enhances the efficiency of the academic evaluation process. Digital marking allows teachers to focus more on teaching and feedback, rather than spending the majority of their time manually reviewing papers.

This system introduces a semi-automated method of evaluation where scanned answer sheets are uploaded into the platform. The system then processes the content using OCR and image processing techniques to extract the handwritten responses. After extraction, the content is analyzed, categorized, and prepared for marking. Teachers can then review the extracted answers and allot marks based on predefined rules or marking schemes. The digital nature of this system brings transparency, speed, consistency, and accuracy to the evaluation process, making it a valuable tool for educational institutions.

## 1.1 Background

The background of this study lies in understanding how educational evaluation has evolved over the years and the role of technology in transforming academic processes. Traditionally, handwritten examinations conducted on paper were considered the most authentic and fair method for assessing a student's understanding. Even today, most schools, colleges, and universities rely heavily on pen-and-paper examinations. While these examinations are effective in capturing student knowledge, the evaluation that follows is often challenging. Manual checking of handwritten answer sheets consumes a significant amount of time. Teachers must carefully analyze each response, interpret unclear handwriting, and identify important keywords or steps. When student numbers are large, this process becomes overwhelming. In many institutions, teachers spend several days or even weeks checking answer sheets, causing delays in result declaration. Delayed results not only affect student morale but also hinder academic planning, admissions, and progression.

With the expansion of education systems and the increase in student enrollment, the limitations of manual evaluation became more visible. Many universities began exploring ways to digitize examinations. Computer-based tests (CBT) became popular for objective-type questions, but subjective handwritten answers still required manual checking. Even when answer sheets were scanned and stored digitally, they were still evaluated manually on a screen. Thus, digital storage did not solve the fundamental problem of slow and inconsistent evaluation.

Advancements in OCR technology presented a major breakthrough. OCR systems allow computers to read printed and handwritten text from images and scanned documents. This technology has been used in several fields such as banking (cheque processing), postal services (address recognition), and government (digitization of documents). Over time, OCR tools improved significantly, enabling better recognition of handwritten text. This progress created an opportunity to apply OCR to educational answer sheets.

At the same time, image processing techniques became more powerful, providing tools to enhance image clarity, remove noise, detect boundaries, and segment specific parts of a scanned document. These advancements made it possible to design a system that not only scans answer sheets but also processes them intelligently. The emergence of deep learning further improved text understanding, mathematical expression recognition, and diagram interpretation.

## 1.2 Problem Statement

The manual evaluation of handwritten answer sheets presents several challenges that directly affect the efficiency and accuracy of the assessment process. One of the biggest challenges is the significant amount of time required to check a large number of papers. A teacher may be responsible for evaluating hundreds of answer sheets within a tight timeframe. This results in fatigue and increases the chances of missing key points, misinterpreting answers, or awarding inconsistent marks.

Human evaluation is subjective. Two different teachers may interpret the same answer differently, leading to inconsistency. Even the same teacher may evaluate differently depending on the time of day, workload, or mental fatigue. Such inconsistencies can lead to unfair results and dissatisfaction among students.

Another problem arises in large-scale examinations where thousands of answer sheets must be evaluated. The process becomes inefficient and difficult to manage manually.

Storing, organizing, and retrieving physical answer sheets also poses logistical challenges and risks such as damage or misplacement. The primary objective of this project is to design and develop a digital system that scans, extracts, and evaluates handwritten answer sheets using advanced technologies. The goal is to reduce the time and effort required for manual checking while improving accuracy and consistency. The system aims to create a supportive environment for teachers by automating repetitive tasks and presenting information in an organized format.

One of the major objectives is to use OCR technology to extract text from scanned answer sheets. This enables the system to convert handwritten answers into machine-readable form. Another objective is to classify the type of answer, whether it is descriptive, mathematical, or diagram-based. This classification helps apply suitable evaluation techniques for each category.

The system aims to assist teachers by highlighting important keywords, analyzing the structure of the answer, and providing computational support in marking. This reduces the workload and speeds up the evaluation process. A user-friendly interface is also an objective, ensuring that teachers can easily upload answer sheets, review extracted content, and generate marks.

## 1.3 Significance of the Project:

The Digimark project is designed to transform how educational institutions evaluate handwritten answer sheets. Traditionally, grading by hand is slow, exhausting, and prone to errors, especially when there are many students. Teachers often experience fatigue, which can lead to inconsistencies or mistakes in marking. This not only affects fairness but can also delay results, leaving students waiting for feedback and slowing down their learning process. Digimark addresses these problems by using advanced technologies like Optical Character Recognition (OCR), Natural Language Processing (NLP), and AI-based evaluation. These tools allow the system to automatically read and analyze handwritten answers, providing faster, more accurate, and consistent grading.

For teachers and administrators, Digimark is a game-changer. It significantly reduces the time and effort required to evaluate papers, freeing educators to focus on teaching and guiding students. The system automatically evaluates answers per question, ensuring that grading standards are uniform and free from personal bias or human fatigue. Features such as batch processing, handling multi-page PDFs, and instant result generation make it ideal for large classes or institutional exams. The dashboard gives a clear overview of student performance, completed evaluations, and pending tasks, helping administrators and teachers manage the evaluation process efficiently.

For students, Digimark brings fairness and transparency. Every answer sheet is evaluated according to the same criteria, so no one is disadvantaged by inconsistent grading. Students receive detailed feedback and per-question scores, which helps them understand where they did well and where they need improvement. Quick access to results also reduces anxiety and allows them to learn from mistakes in a timely manner. Additionally, the system securely handles student data, protecting privacy and ensuring that records are kept safe and accurate.

Beyond individual users, Digimark benefits the entire educational ecosystem. Schools, colleges, and universities can use it for multiple subjects, semesters, and courses without needing extra resources. Its web-based design ensures it works across desktops, laptops, and mobile devices, so both teachers and students can access the system conveniently from anywhere. This flexibility makes Digimark a modern and adaptable solution for today's digital classrooms.

Digimark is much more than a technological tool—it is a transformative solution. It makes the grading process faster, more accurate, and fairer, while also reducing administrative work for educators. By delivering consistent assessments and timely feedback, it helps students improve their learning outcomes and supports better academic management. Digimark provides a practical, reliable, and scalable approach to automating handwritten answer sheet evaluation, paving the way for smarter and more efficient education systems.

## 1.4 Objectives of the Project:

The primary objective of the Digimark project is to automate the evaluation of handwritten answer sheets, minimizing reliance on manual grading and significantly reducing the time and effort required by educators. Traditional grading is slow, labor-intensive, and prone to human errors, especially during large-scale examinations. Teachers often face fatigue, which can lead to inconsistent marking, delays in result declaration, and unfair evaluation outcomes. By leveraging advanced technologies such as Optical Character Recognition (OCR) for text extraction, Natural Language Processing (NLP) for semantic understanding, and AI-based scoring algorithms, Digimark provides a fast, accurate, and reliable alternative to conventional evaluation practices. The system is designed to handle large volumes of answer sheets efficiently, improving overall academic workflow.

A key objective is to ensure consistency, fairness, and transparency in grading. Manual evaluation is often influenced by subjective judgment, fatigue, or varying interpretations of answers. Digimark addresses this by using standardized evaluation criteria, semantic analysis, and keyword-based assessment, ensuring that every student is graded according to the same rules. The system generates per-question marks, cumulative scores, and detailed feedback, helping students understand their strengths and weaknesses. This level of transparency also allows educators to provide more targeted support and guidance, contributing to improved learning outcomes.

Another significant objective is to reduce grading time and increase efficiency. Digimark allows batch processing of multiple answer sheets and supports multi-page PDF uploads, automatically extracting handwritten text, evaluating answers, and generating results in real time. This functionality is particularly valuable for institutions conducting large-scale examinations, as it ensures timely result declaration while reducing administrative workload. Features such as instant PDF report generation, automatic evaluation summaries, and customizable scoring criteria further enhance efficiency and convenience.

The project also emphasizes secure academic record management and data integrity. All uploaded answer sheets, evaluation results, and generated reports are stored in a centralized and encrypted database. This enables teachers and administrators to access historical records easily, track student performance trends over time, and maintain the confidentiality of sensitive academic data. Secure data handling ensures compliance with data privacy standards, preventing unauthorized access or data breaches.

Additionally, Digimark aims to be a scalable, accessible, and user-friendly platform. The system is web-based and compatible with desktops, laptops, and mobile devices, providing flexibility for educators and students to use it anytime and anywhere. Its modular architecture allows for future enhancements, such as multi-language support, AI-driven semantic evaluation, integration of diagrammatic or formula-based answers, and support for multiple question types. The system is designed to evolve with educational needs, making it adaptable for different subjects, courses, and institutional requirements.

The project strives to enhance the overall quality of education. By automating routine grading tasks, Digimark frees up teachers' time, allowing them to focus on interactive teaching and personalized student support. Students benefit from timely feedback, objective evaluation, and detailed performance insights, which helps them improve academically. Overall, Digimark represents a practical, scalable, and transformative solution that modernizes the evaluation process, improves academic efficiency, and ensures fair, accurate, and transparent assessment in educational institutions.

*Chapter 2*

# LITERATURE SURVEY

A Literature Survey is an essential part of any academic or engineering project. It provides a comprehensive understanding of the technologies, previous research works, and existing systems related to digital marking and automated assessment. In the context of this project Digital Marking System using OCR the literature review examines major developments in digital evaluation tools, optical character recognition methods, machine learning algorithms used for document processing, and educational technologies that aim to simplify assessment processes.

Evaluation is a core activity in the educational sector. Traditional evaluation methods are manual, time-consuming, and prone to human error, especially when large numbers of answer scripts must be checked. With the rapid growth of digital technologies, several tools and systems have been developed to support automated or semi-automated evaluation. This literature survey aims to analyze these developments and understand their strengths, limitations, and relevance to the proposed project.

## 2.1 Ramesh, D. et al. — "Systematic Literature Review on Automated Essay Scoring Systems"

**Convenient Name:** Thomas (2003) – Electronic Marking Evaluation
**Source:** ACM SIGCSE Bulletin (ACM Digital Library)

This foundational study investigates electronic marking systems (EMS) and evaluates how automatically generated scores compare with traditional human marking. The research involved controlled experiments where the automatically assigned scores were statistically compared to marks given by human evaluators. The results showed a high correlation between automated and human scores (0.86–0.95), indicating that electronic marking could achieve reliable outcomes. However, the study also noted that automatic scores tend to be slightly lower than human scores. With appropriate calibration and occasional manual intervention, electronic marking systems could closely match human evaluation, demonstrating their practical utility in academic assessment.

The study emphasizes that Automated Essay Scoring (AES) 69 is a key component of digital marking systems, especially for subjective answers where manual evaluation is time-

consuming and prone to inconsistencies. AES systems rely on a combination of linguistic feature analysis, semantic understanding, and increasingly, neural network approaches to assess essays. The research highlights that essay evaluation is inherently complex, as it involves interpreting meaning, assessing argument structure, and understanding context—tasks that require more than just keyword matching. Consequently, hybrid systems that combine automated analysis with human oversight are recommended to ensure both accuracy and fairness in marking.

Key Points from the Study:

- AES 9 plays a critical role in digital marking systems for descriptive answers**,** providing a foundation for automating subjective evaluations.
- The study reviews feature-based and neural approaches for scoring essays, noting current trends, limitations, and challenges in the field.
- Essay evaluation is complex, often requiring hybrid systems that combine linguistic feature extraction, semantic analysis, and manual calibration to achieve accurate and reliable scoring.
- The findings provide quantitative evidence that automation in marking can significantly reduce manual effort while maintaining a high level of reliability.
- The research suggests that digital marking systems can be scaled for large academic settings, particularly when combined with human oversight to handle ambiguous or nuanced responses.

## 2.2 Messer, M. et al. — "Automated Grading and Feedback Tools for Programming Education: A Systematic Review"

**Convenient Name:** Messer et al. (2023) – Automated Grading Tools for Programming
**Source:** ACM Transactions on Computing Education / arXiv Preprint

This systematic review examines automatic grading tools for programming assignments, analyzing studies and systems developed between 2017 and 2021. The review identifies common techniques used in automated programming assessment, including unit testing, static code analysis, and basic style checks, which allow the systems to evaluate correctness, performance, and fundamental quality metrics of student code. These automated tools provide rapid feedback, helping students learn and improve their programming skills in real time, while significantly reducing the workload of instructors.

Although the study focuses on structured code assessments rather than handwritten or essay answers, it highlights the broader impact of automation in educational evaluation. Automated grading tools enhance throughput and consistency, ensuring that all student submissions are evaluated using standardized criteria. They also allow educators to focus on more complex or subjective aspects of teaching, such as mentoring, problem-solving guidance, and advanced feedback on algorithm design.

The review also notes some limitations of these systems. While they can fully automate evaluation of code correctness and basic quality attributes, they often struggle to assess maintainability, readability, or deeper semantic understanding of code. This limitation reflects a broader challenge in automated assessment systems: while structured, rule-based automation can handle straightforward tasks efficiently, more nuanced evaluation often requires human judgment or advanced AI techniques.

Key Points from the Study:

- Demonstrates how automation supports structured assessments, such as programming assignments, by handling large volumes of submissions efficiently.

- Highlights the improvement of grading speed, consistency, and standardization, which reduces manual effort and ensures fair evaluation.
- Serves as a contrast to essay scoring and handwritten text evaluation, showing that different types of assessment require tailored automated approaches.
- Provides insights into the limitations of automation, emphasizing the need for hybrid systems combining automated evaluation and human oversight.

### 2.3 Thomas, P. G. — "The Evaluation of Electronic Marking of Examinations"

**Convenient Name:** Thomas (2003) – Electronic Marking Evaluation
**Source:** ACM SIGCSE Bulletin (ACM Digital Library)

This seminal work investigates the effectiveness of electronic marking systems (EMS) and evaluates how automatically generated results compare with traditional human evaluation. The study conducted controlled experiments in which automatic scores were statistically compared with scores assigned by human markers. The results showed a high correlation between human and automated scores (0.86–0.95), demonstrating that electronic marking systems could produce reliable results even in the early stages of their development. While automatic scores were occasionally slightly lower than human-assigned marks, the study emphasized that appropriate calibration and minimal manual intervention could align automated scores closely with human evaluation.

The research highlighted that electronic marking can improve efficiency and consistency in academic assessments. By using EMS, institutions can reduce examiner workload and minimize errors due to fatigue or subjective bias, ensuring fairer and more standardized grading across large numbers of submissions. The study also provided quantitative evidence supporting the reliability of automated marking systems, which helped to build confidence in the adoption of digital marking technologies in educational environments.

Key Points from the Study:

- Provides an early empirical assessment comparing human and automated marking systems.
- Demonstrates a high correlation between human and automated marking, validating the reliability of digital marking.

- Suggests that hybrid systems, combining automation with manual oversight, provide optimal results by maintaining accuracy while improving efficiency.
- Lays the foundation for subsequent digital marking systems, including hybrid approaches that integrate AI, OCR, and NLP for more complex evaluation tasks.

### 2.4Thakur et al. — "TrueGradeAI: Retrieval-Augmented and Bias-Resistant AI for Transparent and Explainable Digital Assessments"

**Convenient name:** Thakur et al. (2025) – TrueGradeAI

**Source:** arXiv preprint (2025)

TrueGradeAI is a recent framework combining handwriting preservation (tablet stylus capture) with advanced OCR and large language models for explainable automated assessment. The system focuses on transparent decision-making and bias mitigation, using evidence links to score student responses, bridging human and automated evaluation.

Key points you can include:

- Introduces advanced end-to-end AI frameworks for grading both handwritten and typed submissions.
- Focuses on transparency, fairness, and explainability, addressing key concerns in contemporary automated assessment.
- Integrates retrieval-based techniques and LLMs to evaluate student answers in context, comparing them against model solutions for accurate scoring.
- Provides bias-resistant scoring mechanisms to reduce unfair penalization and ensure equitable evaluation across diverse student responses.
- Demonstrates the potential of combining handwriting capture, OCR, and AI for scalable, reliable, and interpretable digital marking.

### 2.5 Ghadekar et al. - "Automating the Grading of Handwritten Examinations through OCR and ML"

**Convenient Name:** Ghadekar et al. (2025) – Handwritten OCR + ML Grading

**Source:** Journal of Integrated Science & Technology (ResearchGate)

Ghadekar et al. (2025) presented a digital grading system designed to automate the evaluation of handwritten examination answer sheets. The system first digitizes scanned answer sheets using Optical Character Recognition (OCR) to extract handwritten content, including text, block diagrams, and mathematical expressions. After digitization, machine learning (ML) algorithms are applied to evaluate student responses, assessing both correctness and contextual relevance. This approach is particularly valuable in traditional academic environments where manual handwriting is still prevalent, and there is a need for efficient, automated grading systems.

Key Points from the Study:

- Focused on real-world exam scripts, making the system applicable in practical educational settings.
- Combines digitization (OCR) with context-aware machine learning scoring, allowing nuanced assessment of student answers.
- Capable of interpreting block diagrams, formulas, and textual answers, extending automated grading beyond simple text.
- Demonstrates how OCR + NLP/ML can streamline grading while maintaining accuracy comparable to traditional human marking.
- Provides insights for designing scalable, automated systems capable of handling multi-page and large-scale examination papers.

### 2.6 Survey of Digital Watermarking Techniques and Its Applications :

**Convenient Name:** Saini & Shrivastava (2014) – Digital Watermarking Survey
**Source:** arXiv / ResearchGate

Although digital watermarking is not a grading system, it is highly relevant to security, authenticity, and integrity in digital assessment systems. The survey explains the principles, techniques, and applications of digital watermarking, which can be used to protect scanned answer sheets, PDFs, and other assessment documents from tampering or unauthorized duplication. By embedding hidden codes or signals within digital content, watermarking ensures that any modification of the document can be detected, thereby safeguarding the integrity of student submissions and institutional records.

The study categorizes watermarking techniques into visible and invisible watermarks, robust and fragile watermarks, and different embedding domains such as spatial and frequency domains. Applications include copyright protection, document authentication, tamper detection, and secure transmission of sensitive data. In the context of digital marking systems, watermarking can be employed to secure uploaded answer sheets, generated reports, and other digital assets, ensuring that the documents remain unaltered and trustworthy throughout the evaluation process.

Key Points from the Study:

- Watermarking embeds hidden authentication codes to protect digital content from tampering.
- Supports document integrity and authenticity, crucial for high-stakes assessments.
- Can be applied to exam answer sheets, scanned submissions, and reports to prevent unauthorized editing or duplication.
- Enhances the security and trustworthiness of digital marking platforms by ensuring all assessment documents are verifiable.
- Provides a foundation for secure and reliable digital assessment systems, complementing technologies like OCR, NLP, and AI-based evaluation.

Overall, Saini & Shrivastava (2014) highlights the importance of digital watermarking as a security layer in automated marking systems. By integrating watermarking techniques, digital evaluation platforms like Digimark can ensure that all scanned answer sheets and reports remain secure, authentic, and tamper-proof, adding a critical level of reliability to the automated assessment process.

*Chapter 3*

# METHODOLOGY

The methodology for developing the digital marking system for scanning answer sheets is based on integrating multiple software components, programming technologies, and algorithmic processes. Each technology plays a specific role in ensuring that scanned answer sheets are processed accurately, textual content is extracted reliably, and teachers receive clear, structured information for evaluation. This chapter explains in detail how each major component—Python, Flask, React, Tesseract OCR, PyMuPDF, pdf2image, and OpenCV—contributes to the overall workflow of the system.

## 3.1 Python

Python is the central programming language used for building the backend logic, integrating OCR operations, and performing computational tasks. The choice of Python is rooted in its simplicity, readability, and strong support for scientific and document-processing libraries.

Python’s dynamic typing and clean syntax make it ideal for rapid development, especially for academic projects that require flexibility and experimentation. It allows developers to write efficient code for file handling, image manipulation, text extraction, and server-side logic.

In this project, Python functions as the “brain” that controls the entire processing pipeline. When a scanned answer sheet is uploaded, Python orchestrates the operations in the following manner:

- It first identifies the nature of the file—PDF or image—and routes it to the appropriate processing module.

- It uses libraries such as PyMuPDF and pdf2image to extract pages and convert them into image formats suitable for OCR.

- It passes the image data to OpenCV for preprocessing, where several enhancements are applied.

- It triggers Tesseract OCR to extract text and then processes the textual output to remove errors and structure the content.

- Finally, it packages the cleaned text into JSON format and sends it to the frontend through Flask APIs.

Python's large ecosystem is one of its biggest advantages. It offers numerous libraries that are highly compatible with each other, enabling smooth integration between document-processing, machine learning, and image-processing modules. Its extensibility ensures that future improvements—such as deep learning–based handwriting recognition or automated scoring—can be built without switching technologies.

### 3.2 Flask Backend

The backend of the digital marking system is developed using Flask, a lightweight and versatile Python web framework that provides the necessary infrastructure for handling server-side operations. Flask is particularly well-suited for this project due to its simplicity, flexibility, and the ease with which it can integrate with various Python libraries for document processing, image manipulation, and OCR. In the context of the digital marking system, the backend is responsible for managing all operations related to file handling, processing scanned answer sheets, extracting textual content, and coordinating communication with the frontend interface.

The Flask framework enables the development of RESTful APIs that act as the communication bridge between the user interface and the processing modules. When a teacher uploads an answer sheet through the frontend, the backend receives the file and determines its format, whether it is a PDF or an image. Once the file type is identified, Flask routes the document to the appropriate processing modules, such as PDF extraction, PDF-to-image conversion, image preprocessing, and OCR text extraction. This modular routing ensures that the system can handle different file formats efficiently and process them in a standardized manner. Flask also provides robust capabilities for managing file storage and temporary data during processing. Uploaded documents are securely stored on the server while being processed, and the system ensures that these files are removed once the extraction and evaluation steps are complete. This helps maintain server efficiency and prevents unnecessary storage accumulation. Additionally, Flask allows for error handling and validation at multiple levels. For instance, if a file is corrupted, incompatible, or incomplete, the backend can detect the issue and send informative feedback to the frontend, thereby preventing disruptions in the evaluation process.

Another significant role of the Flask backend is the orchestration of the processing pipeline. It coordinates the sequential execution of various modules, such as converting PDF pages into images, applying image preprocessing techniques to enhance text clarity, and invoking Tesseract OCR for text extraction. The backend ensures that the output from one module is properly formatted and passed to the next module in the pipeline, maintaining data consistency and accuracy. After OCR extraction, Flask also performs additional post-processing to structure the extracted text according to questions or answer sections, making it easier for teachers to review and assign marks.

Flask's lightweight architecture offers scalability, allowing the system to handle multiple requests simultaneously. This is particularly important when evaluating large batches of answer sheets or when multiple teachers are using the system concurrently. The framework supports asynchronous operations, enabling the backend to process one document while accepting new uploads from other users without interruption. Moreover, Flask can be easily extended to integrate with databases for storing extracted text, marks, and student performance records, providing a foundation for future development of more comprehensive digital evaluation systems.

Security and accessibility are also addressed through Flask's framework. It allows for the implementation of secure file transfer protocols, authentication mechanisms, and controlled access to resources, ensuring that only authorized users can upload and access answer sheets. The backend can also provide real-time responses to the frontend, indicating processing status, progress updates, and final results, which enhances the interactivity and responsiveness of the system.

### 3.3 React Frontend

The frontend of the digital marking system is developed using React, a modern JavaScript library that is widely recognized for building dynamic and interactive web applications. In this project, React plays a critical role in providing a seamless and user-friendly interface through which teachers can interact with the system. Unlike static web pages, the React-based frontend allows the application to respond instantly to user actions, efficiently display extracted content, and manage large volumes of data without compromising performance. React's component-based architecture enables the creation of independent, reusable elements, each responsible for a specific part of the user interface, such as uploading files, displaying OCR results, navigating between pages, or entering marks. This modularity not only simplifies the development process but also allows for easy scalability and the addition of new features in the future, such as dashboards, automated suggestions, or summary generation.

A key feature of React is its virtual Document Object Model (DOM), which improves the performance of the application by minimizing the need to reload entire web pages. When a teacher uploads an answer sheet, the backend processes the file and returns the extracted text. React updates only the relevant components displaying the results, rather than refreshing the entire page. This makes the system faster and more responsive, ensuring that teachers experience minimal delays even when dealing with large or multi-page answer sheets. The use of virtual DOM is particularly beneficial in scenarios where multiple updates occur simultaneously, such as highlighting keywords, showing marking suggestions, or navigating between pages of a scanned PDF.

The design of the React interface emphasizes usability and clarity. Since teachers may have limited time to evaluate large numbers of answer sheets, the layout is structured to display extracted answers in a clear and organized manner. Pages are presented sequentially, with answers segmented according to question numbers, allowing for quick comparison with model answers. Loading indicators and status messages guide the user throughout the evaluation process, while error handling mechanisms inform teachers of any issues such as corrupted files or low-quality scans. This ensures that the system remains stable and reliable under various operating conditions.

React also facilitates the integration of frontend components with the backend services through API calls. When a teacher uploads an answer sheet, the React frontend sends the file to the Flask backend, which processes it and returns the OCR results. The frontend then renders these results in a readable format, highlighting important information and maintaining consistent formatting. This seamless interaction between frontend and backend ensures that teachers can complete the evaluation process efficiently without encountering delays or interface issues.

The React frontend serves as the interactive and visual layer of the digital marking system. By leveraging its component-based structure, virtual DOM rendering, state management capabilities, and integration with backend APIs, React provides a fast, responsive, and intuitive environment for teachers. It transforms the extracted data into a clear and organized interface, enabling efficient review and marking of student answers. The frontend thus plays a crucial role in making the semi-automated digital marking system practical, reliable, and user-friendly for educational assessment.

## 3.4 OCR (Tesseract)

Optical Character Recognition (OCR) is a critical technology in the development of the digital marking system, as it allows the conversion of handwritten or printed text in scanned answer sheets into machine-readable digital text. Among the available OCR engines, Tesseract has been selected for this project due to its open-source nature, high accuracy, and versatility in handling various document formats. The integration of Tesseract in the system ensures that large volumes of handwritten answer sheets can be processed efficiently, reducing the manual effort required for marking and evaluation.

Tesseract operates by analyzing pixel patterns in images to identify text. When a scanned answer sheet is provided as input, the engine processes each image at the pixel level, recognizing shapes and strokes that correspond to individual characters. It then compares these shapes with its internal trained datasets to determine the most likely characters and words. The OCR output is typically in plain text format, which is then further processed to structure answers according to questions and sections. This structured output enables teachers to view extracted content in a readable and organized manner, significantly speeding up the evaluation process.

One of the challenges in applying OCR for handwritten answer sheets is the variability in handwriting styles. Students may write with different slants, sizes, spacing, and cursive forms, all of which impact recognition accuracy. To overcome this, the system uses image preprocessing techniques with OpenCV before passing the input to

Tesseract. Preprocessing steps such as noise removal, thresholding, skew correction, and contrast enhancement ensure that the handwritten text is more clearly defined, which improves the accuracy of OCR results. By preparing the images carefully, Tesseract is able to recognize characters and words with higher precision, even in documents where the handwriting is not perfectly legible.

Tesseract is also capable of handling multi-page documents and complex layouts. Many scanned answer sheets contain multiple pages with varying structures, such as tables, diagrams, and mixed text formats. The engine, when combined with PDF-to-image conversion and page segmentation modules, allows each page to be processed individually, ensuring that no content is lost or misread. This is particularly important in educational assessments where even a small missed answer can affect the evaluation process. Furthermore, Tesseract supports multiple languages and fonts, which provides a foundation for future improvements, such as handling multilingual answer sheets or regional scripts.

In the digital marking system, Tesseract's role extends beyond basic text recognition. After extracting text, the system analyzes it to identify the structure of the answers. For example, it distinguishes between descriptive answers, short answers, and numerical or formula-based responses. The extracted text is then formatted and organized so that teachers can easily review it alongside the original scanned images. Although Tesseract does not inherently interpret diagrams, equations, or complex symbols, its text extraction capabilities are sufficient for a semi-automated marking system, where human evaluators review and confirm the marks.

Despite its advantages, Tesseract has certain limitations. Its performance can be affected by poor-quality scans, low resolution, inconsistent lighting, and heavily cursive handwriting. In such cases, additional preprocessing and enhancement techniques become critical to improve recognition. Moreover, Tesseract alone cannot evaluate the correctness of answers; it only converts handwritten input into text. The evaluation logic, which compares extracted answers with model answers or marking schemes, is implemented separately in the Python backend. Nonetheless, Tesseract remains a pivotal component, as it transforms unstructured handwritten data into a format that can be further analyzed and evaluated efficiently.

## 3.5 PyMuPDF (fitz) for PDF Extraction

In the development of a digital marking system, PyMuPDF, accessed through the Python fitz library, serves as a fundamental tool for processing PDF documents. Since educational institutions commonly scan answer sheets and store them as PDF files, an efficient method to extract, read, and analyze these files is crucial. PyMuPDF is particularly well-suited for this task because it provides a high-performance interface for opening, inspecting, and manipulating PDF content while maintaining the quality and integrity of scanned images.

The primary function of PyMuPDF in this system is to extract individual pages from multi-page PDFs. Each uploaded answer sheet, whether single or multi-page, is processed in a sequential manner. PyMuPDF allows the system to access each page individually and retrieve its contents, including embedded text when available. This capability is essential because some answer sheets may contain typed or printed instructions in addition to handwritten answers, and accessing all textual content ensures no information is lost during processing.

Beyond simple page extraction, PyMuPDF also facilitates metadata detection and page orientation analysis. Many scanned PDFs may be rotated or misaligned during scanning, and identifying these orientation issues at an early stage allows for correction before OCR processing. This step is particularly important for handwritten answer sheets, as even minor tilts can significantly reduce OCR accuracy. By integrating PyMuPDF, the system ensures that each page is prepared correctly for subsequent conversion and text extraction.

After extracting individual pages, PyMuPDF allows conversion of these pages into images. This conversion is necessary because OCR engines such as Tesseract operate primarily on images rather than PDF files. The library preserves the high resolution of the original scan, ensuring that the resulting images retain clarity, sharpness, and readability.High-resolution images are crucial for accurate OCR processing, particularly when dealing with handwritten content, which can be more challenging to recognize than printed text.

The workflow begins with a teacher uploading a scanned PDF of an answer sheet. PyMuPDF accesses the file, extracts each page, and verifies its orientation and resolution. Each page is then converted into a high-quality image format, which is subsequently passed to the OCR module for text extraction. By using PyMuPDF, the system can efficiently handle both single-page and multi-page answer sheets, ensuring that all content is preserved and processed accurately. PyMuPDF acts as a reliable intermediary between raw PDF input and image-based OCR processing. Its speed, accuracy, and ability to handle complex PDFs make it an indispensable component of the digital marking system, particularly when large volumes of scanned answer sheets need to be processed in an academic environment.

### 3.6 PDF-to-Image Conversion (pdf2image)

Once PDF pages are extracted using PyMuPDF, they must be converted into image formats compatible with OCR engines. The pdf2image library performs this conversion efficiently, transforming PDF pages into high-resolution images such as JPG or PNG. This conversion is a critical step because OCR engines, including Tesseract, analyze pixel data to recognize characters, and they cannot directly interpret PDF content in its original format.

The main role of pdf2image in the digital marking system is to ensure that each PDF page is converted into an image with sufficient clarity and resolution to support accurate text extraction. High-resolution images preserve fine details, including small handwriting strokes, variations in pen pressure, and other nuances present in scanned answer sheets. Without high-quality conversion, OCR accuracy would be significantly reduced, potentially resulting in missed words or incorrect text recognition. Another important function of pdf2image is handling page rotation and orientation issues. Scanned PDFs often contain pages that are skewed or rotated, either due to scanner settings or physical misalignment during scanning. The library can adjust the orientation of each page during conversion, ensuring that the resulting images are correctly aligned for OCR processing. This adjustment minimizes errors caused by misaligned text and enhances the recognition of handwritten content.

pdf2image also supports batch conversion, making it suitable for multi-page answer sheets. Each page of a PDF is converted into a separate image file, maintaining consistent quality across the entire document. These individual images are then passed sequentially to the OCR module, allowing for structured extraction of answers on a page-by-page basis. By maintaining uniform quality and resolution, the system ensures that all answers, whether descriptive or numerical, are captured accurately. In addition to improving OCR performance, the conversion step enables further image preprocessing operations. After the PDF is converted into an image, libraries such as OpenCV can be applied to enhance contrast, remove background noise, adjust brightness, and correct skew. These preprocessing steps further increase the accuracy of Tesseract OCR, resulting in reliable text extraction for semi-automated marking.

*Chapter 4*

# SYSYEM ANALYSIS

In many educational institutions, answer sheets are still checked manually by teachers. This process takes a lot of time and effort, especially when there are many students. Teachers may feel tired while checking papers, which can sometimes lead to mistakes or unequal marking. Results are also delayed, and students do not get feedback quickly. Managing piles of physical answer sheets and maintaining records is another big challenge.

The Digimark system is designed to solve these problems by making the evaluation process automatic and digital. Instead of checking papers manually, teachers can upload scanned handwritten answer sheets in PDF format. The system reads the handwriting using OCR technology and understands the answers using AI and language processing techniques. It then compares student answers with model answers and gives marks automatically. This saves time and makes the marking process more accurate and fair.

Digimark also helps in organizing academic data properly. All answer sheets, marks, and reports are stored securely in digital form. Teachers can easily check previous records, monitor student performance, and download result reports whenever needed. This reduces paperwork and makes record keeping simple and efficient. The system is easy to use and works through a web browser, so teachers and students can access it from laptops or mobile devices. It supports checking multiple answer sheets at the same time, which is very useful during exams. Overall, Digimark makes the evaluation process faster, fairer, and more reliable, helping both teachers and students in the education system.

## 4.1 Functional Requirements:

Functional requirements describe what the system should do. The Digimark system is designed to automate handwritten answer sheet evaluation and support teachers, students, and administrators efficiently.

**1. User Registration and Login**

- The system shall allow users (teachers, students, and administrators) to register using valid details.
- The system shall authenticate users using email and password.
- The system shall restrict access to authorized users only.
- The system shall provide role-based access (teacher, student, admin).
- The system shall display error messages for invalid login credentials.

**2. Answer Sheet Upload**

- The system shall allow teachers to upload scanned handwritten answer sheets.
- The system shall accept only valid file formats (PDF).
- The system shall support multi-page PDF uploads.
- The system shall validate file size and format before processing.
- The system shall prevent uploading of invalid or corrupted files.

**3. OCR (Handwritten Text Recognition)**

- The system shall extract handwritten text from uploaded answer sheets using OCR.
- The system shall handle different handwriting styles.
- The system shall process scanned images with acceptable accuracy.
- The system shall convert handwritten content into machine-readable text.

**4. Answer Evaluation**

- The system shall compare extracted text with predefined model answers.
- The system shall evaluate answers using NLP and AI-based techniques.
- The system shall assign marks for each question automatically.
- The system shall support keyword matching and semantic similarity analysis.
- The system shall ensure consistent and unbiased evaluation.

**5. Roll Number and Student Identification**

- The system shall detect and extract roll numbers from answer sheets.
- The system shall associate evaluated answers with the correct student.
- The system shall validate roll numbers before storing results.

**6. Result Generation**

- The system shall calculate total marks automatically.
- The system shall generate detailed evaluation reports.

- The system shall show per-question marks and overall score.
- The system shall generate downloadable result PDFs.

**7. Feedback System**

- The system shall provide feedback for each question.
- The system shall highlight missing or partially correct answers.
- The system shall help students understand mistakes and improve learning.

**8. Dashboard and Visualization**

- The system shall provide a dashboard for teachers and administrators.
- The dashboard shall display number of evaluated answer sheets.
- The system shall show average marks and performance statistics.
- The system shall display pending and completed evaluations.

**9. Batch Processing**

- The system shall allow evaluation of multiple answer sheets at once.
- The system shall reduce manual effort during large-scale examinations.
- The system shall process answer sheets efficiently without system failure.

**10. Data Storage and Record Management**

- The system shall securely store answer sheets and evaluation results.
- The system shall maintain historical academic records.
- The system shall allow easy retrieval of past evaluations.
- The system shall ensure data integrity and privacy.

**11. Report Download and Export**

- The system shall allow users to download result reports.
- The system shall support PDF report generation.
- The system shall allow printing of evaluation reports.

**12. Error Handling and Validation**

- The system shall display clear error messages for invalid inputs.
- The system shall handle OCR failures gracefully.
- The system shall prevent system crashes during processing.

**13. Security and Access Control**

- The system shall protect user data from unauthorized access.

- The system shall ensure secure login sessions.
- The system shall restrict students from modifying marks.

**14. Admin Management**

- The system shall allow administrators to manage users.
- The system shall allow administrators to monitor system activity.
- The system shall support system configuration and maintenance.

**15. System Logout**

- The system shall allow users to log out securely.
- The system shall terminate sessions after inactivity.

## 4.2 Non-Functional Requirements:

Non-functional requirements describe how well the system should work rather than what it does. They focus on quality, performance, security, usability, and reliability of the Digimark system.

**1. Performance**

- The system should process and evaluate answer sheets within a reasonable time.
- OCR and evaluation should complete quickly even for multi-page PDFs.
- Batch processing of multiple answer sheets should not significantly slow down the system.
- The dashboard should load results and reports without noticeable delay.

**2. Accuracy**

- The OCR system should accurately recognize handwritten text.
- The evaluation logic should produce consistent and reliable marks.
- The system should minimize errors in roll number detection and answer extraction.
- Results should closely match human evaluation standards.

**3. Reliability**

- The system should work correctly without frequent crashes or failures.
- Uploaded answer sheets and results should not be lost during processing.
- The system should recover safely from errors such as OCR failure or network interruption.
- Stored data should remain available and intact at all times.

**4. Usability**

- The system should be easy to use for teachers, students, and administrators.
- The interface should be clean, simple, and understandable even for non-technical users.
- Uploading answer sheets and viewing results should require minimal steps.
- Error messages should be clear and helpful.

**5. Security**

- The system should protect student and teacher data from unauthorized access.
- User login information should be securely stored and encrypted.
- Only authorized users should be able to view or modify evaluation data.
- Answer sheets and result reports should be protected from tampering.

**6. Privacy**

- Student academic records should remain confidential.
- Personal data should not be shared without authorization.
- The system should follow basic data protection and privacy standards.
- Access to sensitive information should be role-based.

**7. Scalability**

- The system should support a growing number of users and answer sheets.
- It should handle large-scale exams without performance degradation.
- New features or subjects should be added without major system changes.

**8. Compatibility**

- The system should work on different devices such as desktops, laptops, and mobile phones.
- It should support major web browsers.
- Uploaded PDF files should be processed consistently across platforms.

**9. Maintainability**

- The system should be easy to update and maintain.
- Bug fixes and feature upgrades should be implemented without disrupting users.
- The modular design should allow improvements in OCR or AI models.

**10. Availability**

- The system should be available whenever users need to access it.
- Downtime should be minimal and preferably scheduled.
- Users should be able to view previously generated results even if new uploads are paused.

**11. Data Storage and Backup**

- The system should store data securely in a centralized database.
- Regular backups should be maintained to prevent data loss.
- Old records should remain accessible for academic reference.

**12. Response Time**

- The system should respond quickly to user actions like login, upload, and result viewing.
- Dashboard statistics should update without long waiting times.

**13. Fault Tolerance**

- The system should handle invalid files or poor-quality scans gracefully.
- It should not crash due to unexpected user inputs.
- Meaningful error messages should guide users to correct actions.

## 4.3 System Constraints:

System constraints are the limitations and restrictions under which the Digimark system must operate. These constraints arise due to technical, operational, legal, and practical factors. Understanding them helps in setting realistic expectations and planning future improvements.

**1. Handwriting Recognition Limitations**

One of the major constraints of the Digimark system is the accuracy of handwritten text recognition. OCR performance depends heavily on the quality and clarity of handwriting. Poor handwriting, overlapping text, uneven spacing, or very light ink can reduce text extraction accuracy. Although image preprocessing improves results, OCR may still struggle with highly cursive or unclear handwriting. This limitation can affect the evaluation accuracy of descriptive answers.

**2. Quality of Scanned Documents**

The system relies on scanned answer sheets in PDF format. If the scanned documents are blurred, tilted, low-resolution, or contain shadows, the OCR engine may fail to extract text correctly. Improper scanning methods or low-quality scanners can therefore impact overall system performance. The system works best with clean, well-aligned scans.

**3. Limited Understanding of Complex Answers**

While Digimark uses NLP and keyword-based semantic matching, it has limited ability to understand deep logical reasoning, creativity, or unique answer structures. Answers that are conceptually correct but expressed in an uncommon way may receive lower marks. This constraint highlights the need for human oversight in highly subjective evaluations.

**4. Dependency on Predefined Model Answers**

The evaluation process depends on predefined model answers and keywords provided by educators. If the model answer is incomplete, poorly structured, or missing important variations, the system may not correctly evaluate student responses. This constraint requires careful preparation of model answers to ensure fair assessment.

**5. File Format Restrictions**

Currently, the system accepts only PDF files for question papers, model answers, and student answer sheets. Other formats such as images (JPG, PNG) or Word documents are not supported. This restriction ensures consistency and security but limits flexibility for users.

**6. Processing Time for Large Batches**

Although Digimark significantly reduces grading time, processing a very large number of answer sheets simultaneously can increase server load and response time. OCR and AI evaluation are computationally intensive, and system performance may slow down during peak usage without adequate server resources.

**7. Internet Dependency**

The system is web-based and requires a stable internet connection. In areas with slow or unreliable internet access, uploading large PDF files or viewing results may be delayed. This can affect usability, especially in remote or rural educational institutions.

**8. Data Privacy and Legal Constraints**

The system must comply with data protection and privacy regulations. Handling student data requires strict access control and secure storage. These legal constraints limit how data can be stored, shared, or analyzed and require continuous updates to security measures.

**9. Limited Support for Non-Textual Content**

Currently, the system focuses mainly on text-based answers. Evaluation of diagrams, graphs, mathematical expressions, and equations is limited. Although some basic recognition is possible, full automation of such content remains a constraint and may require advanced AI models in the future.

**10. Requirement of Human Supervision**

Despite automation, Digimark cannot fully replace human evaluators. For borderline cases, unclear answers, or dispute resolution, manual review is necessary. This hybrid dependency ensures accuracy but limits full automation.

*Chapter 5*

# IMPLEMENTATION AND RSEULT

The implementation of the Digital Marking and Automated Answer Sheet Evaluation System (DMEF) involves a structured and systematic development process that integrates image processing, OCR, AI-based scoring, and cloud-enabled backend services. The implementation focuses on ensuring accuracy, scalability, and reliability while maintaining a user-friendly interface for examiners and administrators. The key components of the implementation are described below.

## 5.1 Home page :

The Home Page of the Digital Marking for Scanned Answer Sheets system serves as the central entry point for users, offering a clean, intuitive, and user-friendly interface. It provides quick access to all major functionalities such as uploading answer sheets, viewing results, accessing the dashboard, and managing user accounts. The layout is designed with simplicity and clarity in mind, ensuring that teachers, evaluators, and administrators can easily navigate the system without technical difficulty. The Home Page prominently features a secure login section that validates the user's role teacher, admin, or evaluator before granting access to the core functionalities. A brief overview of the system is displayed to introduce users to key features like automated OCR extraction, AI-based evaluation, and result generation.

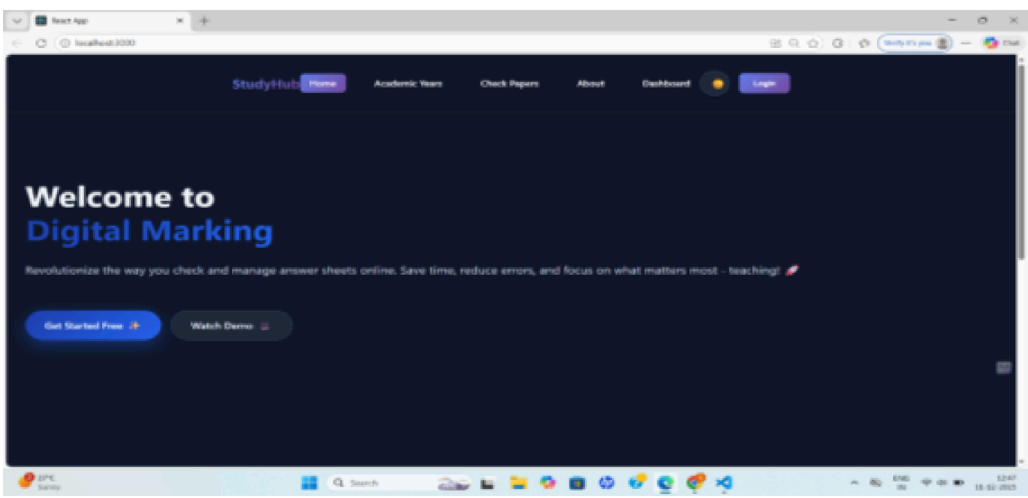



**Fig: 5.1 Home Page**

## 5.2 Academic Years Page:

The Academic Years Page serves as an organized and user-friendly interface that allows users to navigate through different academic years in the digital marking system. As soon as the user opens this page, they are presented with a clean and visually appealing layout featuring four distinct cards—1st Year, 2nd Year, 3rd Year, and Final Year. Each card is designed with a modern UI style, using rounded corners, soft gradients, and a highlighted numeric badge that visually represents the academic year. Under each card, a brief description prompts the user to click and explore the respective semesters. Additionally, the page provides quick information about the number of subjects available in both odd and even semesters, ensuring that users have essential academic details at a glance. The color theme, typography, and spacing contribute to a professional and student-friendly experience. This page functions as a critical navigation point, helping evaluators or students easily select their desired academic year before proceeding to semester-specific subjects and answer sheet evaluation tasks.

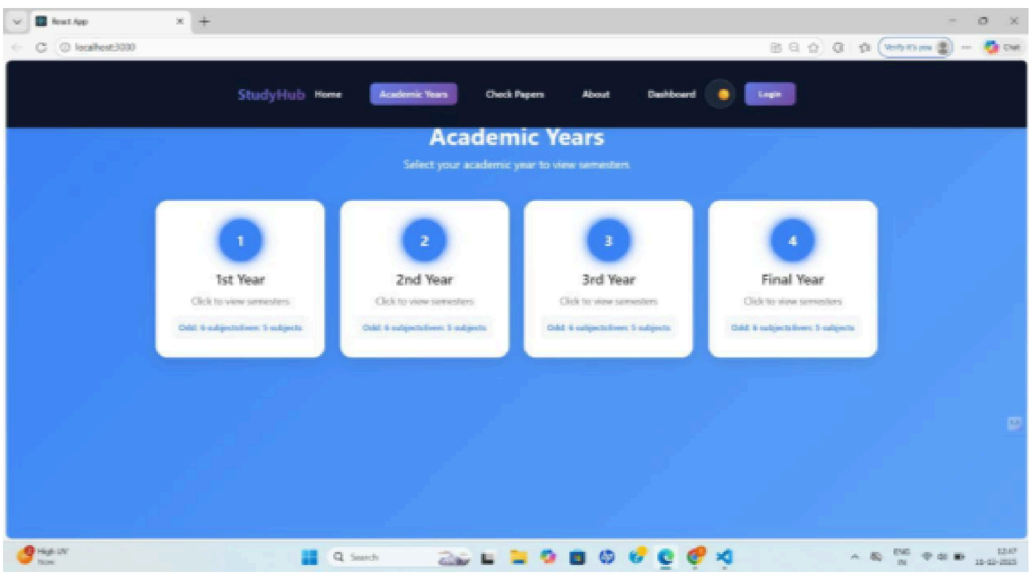



**Fig: 5.2 Academic Years Page**

## 5.3 : Check paper Page:

The Check Papers Page is one of the core modules of the system, designed to facilitate automated answer sheet evaluation using AI and OCR technologies. The page presents a structured and intuitive interface where users can input all necessary student and exam-related details before processing the evaluation. At the top, the page allows the user to enter Roll Number, Academic Year, and Class, ensuring the system can uniquely identify and categorize each student's submission. The Answer Type Configuration section provides selectable options such as descriptive answers, mathematical answers, and diagrammatic responses, enabling the evaluator to define which types of content the AI should analyze. This flexibility ensures that the evaluation process is tailored to the specific subject requirements.

The page includes a Document Upload module, where users can upload the Question Paper PDF, Model Answer PDF, and Student Answer PDF. Each upload field is designed with a clean UI containing buttons for selecting files, ensuring ease of use and preventing incorrect file submissions. Once all documents and parameters are uploaded, users can choose to either reset the form or proceed by clicking the Evaluate Answers button, which triggers the automated evaluation process

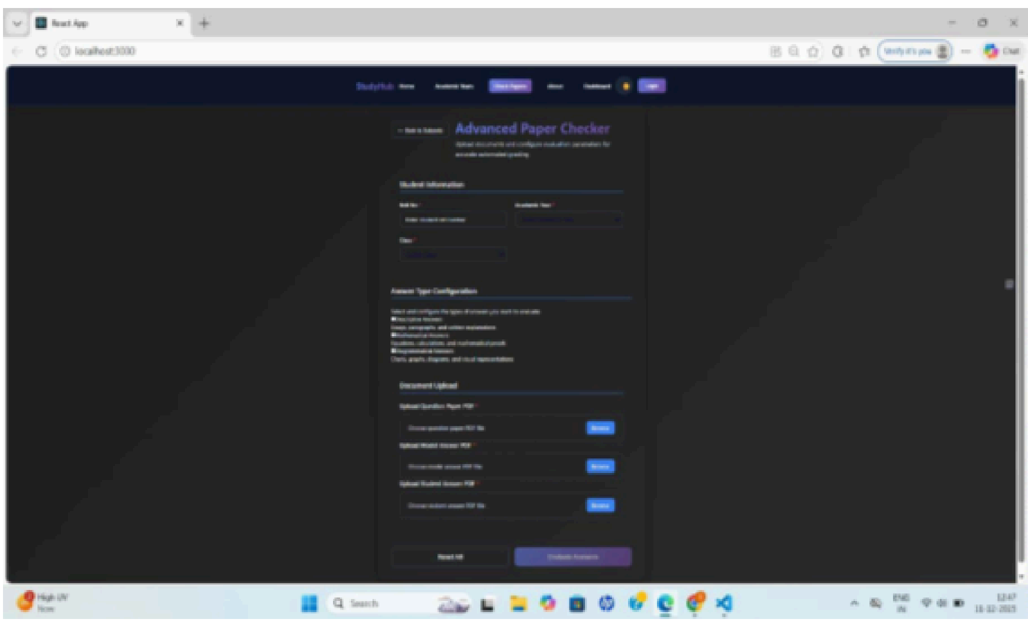



**Fig: 5.3 Check Paper Page**

## 5.4 About Page :

The About Page highlights the major benefits of using the digital paper-checking system and explains why it is better than traditional manual evaluation. The page is designed in a clean, modern layout and presents each feature in separate cards with icons, making it easy for users to understand the system’s advantages at a glance.

The first feature, Lightning Fast, explains that the system can check answer sheets within seconds because it uses optimized AI and automation. This saves teachers a lot of time compared to manual checking. The next feature, Highly Accurate, assures users that the system evaluates answers with almost perfect accuracy using pattern-recognition techniques, reducing human errors and ensuring fair results.

The page also highlights Detailed Analytics, which means teachers and administrators get deeper insights like performance reports, weak topics, and student trends. Another important benefit is Secure & Private, where all uploaded data is encrypted and protected with enterprise-level security, making sure student information stays safe.The AI Powered feature shows that the system becomes smarter with continuous use, learning from past evaluations to improve future accuracy. Finally, it mentions Mobile Friendly, meaning users can access the dashboard, upload papers, or check results from any device—mobile, tablet, or desktop—giving complete flexibility.

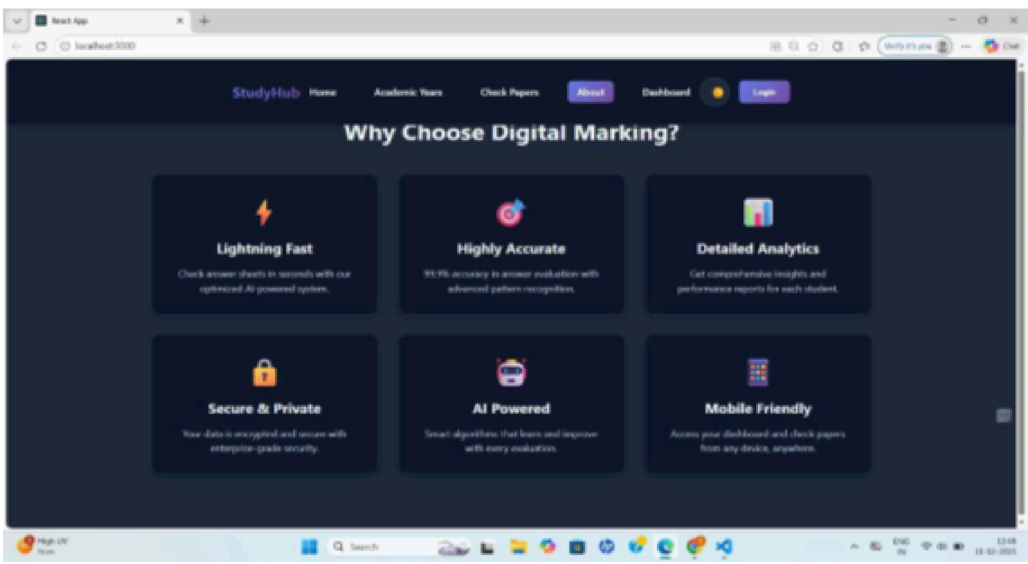



**Fig: 5.4 About Page**

## 5.5 Dashbaord Page:

The Dashboard Page acts as the main control center where users can view all important information related to answer sheet evaluations in a single place. It presents a clear and organized overview of how many papers have been checked, the average performance of students, and how many filters are currently applied. The design of the page is clean and user-friendly, allowing users to easily find the details they need without confusion.

At the top of the dashboard, users are greeted with three summary cards that provide quick insights. These cards show the total number of evaluations completed, the overall average percentage of the evaluated papers, and the number of active filters applied while browsing the records. This immediate overview helps users understand the current progress of evaluations at a glance.

Below the summary section, the dashboard includes a filtering area where users can refine the displayed records based on academic year, class, or subject. These filters make it easier to locate specific evaluation data without scrolling through all entries. A "Clear Filters" option is also available, allowing users to reset their selections and return to the full list effortlessly. The bottom section of the page displays the evaluation records. When no papers have been checked yet, the dashboard shows a helpful message indicating that no results are available and encourages users to begin checking papers.

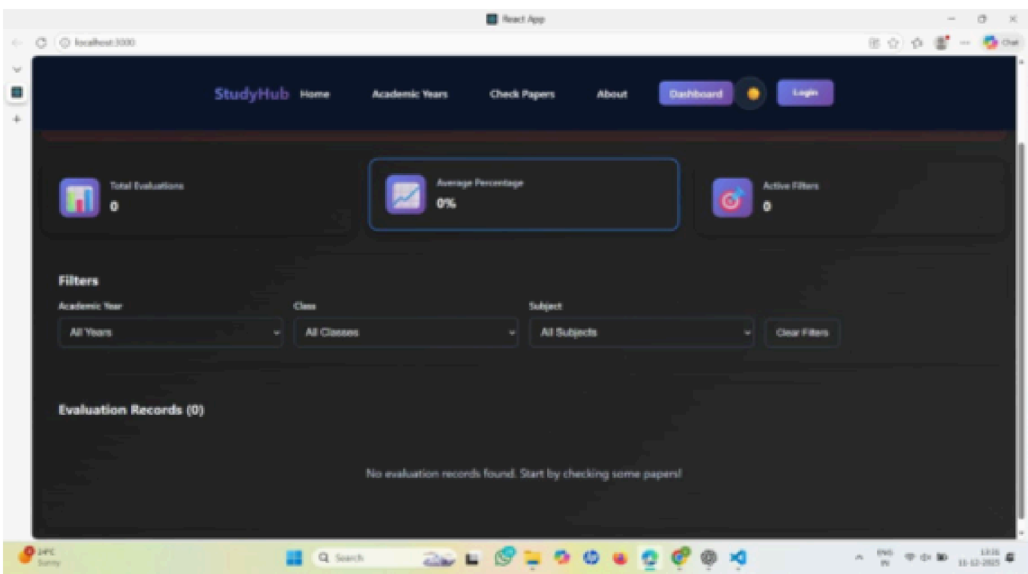



**Fig: 5.5 Dashboard Page**

## 5.6 Login Page:

The Login Page is the entry point for users to access the system. It provides a clean, modern, and user-friendly interface where users can sign in using their registered email and password. The design focuses on simplicity, making it easy for teachers, students, or administrators to log in without confusion.

At the center of the page, there is a small login box that contains two input fields one for the user's email and the other for their password. After entering the details, the user can click the Login button to access their dashboard and use features like checking papers, viewing academic years, or managing uploaded documents. If the login information is incorrect or not registered, the user will not be able to enter the system.

The page also includes a small link for new users who don't yet have an account. By clicking Register, they can move to the signup page and create a new account. The entire layout uses soft gradient colors and a clean card design, making the login process simple, secure, and visually appealing.

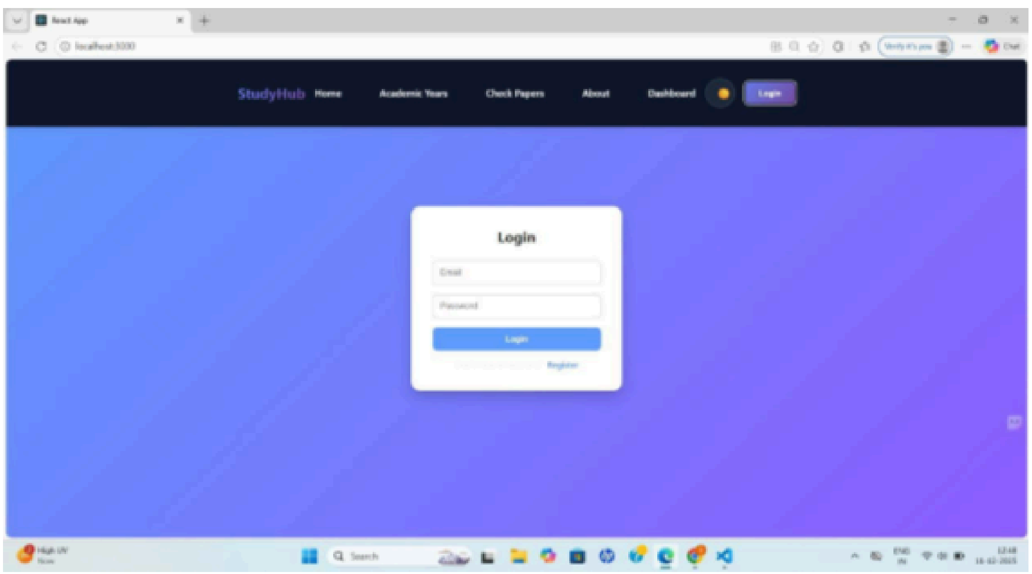



**Fig 5.6 Login Page**

# CONCLUSION AND FUTURE SCOPE

- **Conclusion:**

The Digital Marking System developed in this project successfully demonstrates how technology can modernize and improve the traditional evaluation process. By integrating OCR, automated validation, secure data handling, and a user-friendly dashboard, the system reduces manual workload, minimizes human error, and ensures faster and more consistent assessment of student submissions. The project shows that combining automation with supervised human review significantly enhances accuracy and reliability compared to conventional paper-based evaluation.

The system's ability to accept only valid file formats (such as PDFs), recognize roll numbers, extract necessary text, and display results efficiently ensures transparency and integrity in the marking process. Additionally, the digital interface contributes to better organization of answer scripts, easy retrieval of records, and improved examiner productivity.

Overall, this project demonstrates that digital marking is not just a technological upgrade but a practical solution to the limitations of manual evaluation. It lays the foundation for future improvements such as integrating machine learning for answer evaluation, expanding the types of questions that can be auto-graded, strengthening security features, and enabling full-scale deployment across academic institutions. The Digital Marking System thus represents a significant step toward smarter, faster, and more reliable educational assessment.

- ## Future Scope:

The future of digital marking systems holds significant promise as advancements in machine learning, OCR, and educational technology continue to evolve. Modern assessment environments demand faster, more accurate, and scalable solutions for evaluating large volumes of answer sheets. Digital marking systems can transform traditional evaluation by providing semi-automated or fully automated grading while maintaining consistency and reducing teacher workload. The integration of advanced data analytics, AI-based evaluation, and cloud-enabled infrastructures offers the potential for real-time assessment, personalized feedback, and predictive analytics for student performance.

The digital marking system is designed to process scanned answer sheets efficiently by extracting textual content using OCR technology and assisting teachers in grading. The system supports multiple input formats, including PDFs and images, and provides a web-based interface where educators can upload, view, and evaluate responses. The workflow begins with PDF extraction using PyMuPDF, followed by conversion into high-resolution images via pdf2image. These images undergo preprocessing using OpenCV to enhance clarity and correct skew, which improves OCR accuracy. Tesseract OCR then extracts the text from each page, which is structured according to questions and displayed on the frontend for teacher review.

The system significantly reduces manual effort while improving consistency and transparency in marking. Teachers can focus on evaluating content rather than transcribing or organizing answers, and the extracted data can be stored for further analysis, reporting, or longitudinal tracking of student performance. Although the current system is semi-automated, it lays a foundation for fully automated grading in the future, including handling descriptive answers, diagrams, and multi-language responses.

**1. Future Enhancements**

- **Integration with AI-Based Evaluation**

Future versions of the system could incorporate machine learning models capable of automated grading for descriptive answers. By training models on historical answer sheets and marking schemes, the system could predict scores and provide marking suggestions, significantly reducing the manual workload. These models could also recognize patterns in student responses, identify common errors, and generate personalized feedback, supporting targeted learning interventions.

- **Support for Multilingual and Handwritten Content**

Current OCR engines perform best with standard handwriting and specific languages. Future enhancements could integrate AI-powered handwriting recognition capable of accurately reading multiple languages and varying handwriting styles. This would allow the system to support diverse classrooms and institutions with regional language scripts.

- **Real-Time Evaluation and Analytics**

As technology advances, digital marking systems could provide near-real-time evaluation, allowing teachers to access partially processed answer sheets immediately after submission. Coupled with analytics, the system could generate insights such as student performance trends, question difficulty analysis, and overall assessment statistics, aiding educators in curriculum planning and intervention strategies.

- **Integration with Educational Platforms**

The system can be integrated with Learning Management Systems (LMS) or institutional databases. Such integration would allow seamless synchronization of student records, automatic update of grades, and centralized storage for future analysis. Additionally, integration with cloud services would enable collaborative marking, where multiple teachers can evaluate large batches of answer sheets simultaneously, ensuring efficiency in large-scale examinations.

- **Enhanced Security and Privacy**

Future systems can implement advanced encryption, access controls, and data anonymization to ensure the confidentiality of student information. Compliance with privacy regulations will build trust among institutions and students, especially in scenarios where sensitive data, such as examination scores, is being processed.

- **Visualization and Reporting Tools**

The system could include advanced dashboards showing student performance metrics, question-wise analysis, and comparative class performance. Interactive charts, graphs, and heatmaps can help teachers quickly identify trends, understand learning gaps, and make data-driven decisions for student development.

**2. Workflow and Architecture :**

- **Answer Sheet Collection:** Students submit answer sheets in PDF or image formats. Multiple pages, handwritten responses, and various answer styles are supported.

- **PDF Extraction and Conversion:** The system uses PyMuPDF to extract pages from PDFs and pdf2image to convert each page into high-resolution images suitable for OCR.

- **Image Preprocessing:** OpenCV performs noise removal, contrast enhancement, skew correction, and other preprocessing tasks to improve text clarity for OCR recognition.

- **Text Extraction (OCR):** Tesseract OCR processes the images to extract handwritten or printed text. The extracted text is structured by question and section for further analysis.

- **Evaluation Assistance:** Extracted answers are displayed in a web interface where teachers can review, assign marks, and add comments. In future implementations, AI-based grading models could provide score suggestions.

- **Result Generation and Reporting:** Marks, comments, and extracted text are stored in a database. Dashboards can generate summaries, performance analytics, and question-wise statistics.

- **Feedback and Analytics:** Teachers can access insights regarding student performance trends, common mistakes, and assessment analytics, which can inform curriculum adjustments and student interventions.

### 3. Advantages of the Digital Marking System

- **Accelerates the Marking Process**

  Reduces the time required for evaluating large batches of answer sheets. Automates text extraction and organization, allowing teachers to focus on evaluation rather than administrative tasks.

- **Reduces Human Error**

  Minimizes mistakes caused by fatigue, oversight, or subjective judgment. Ensures consistent and accurate evaluation across multiple teachers and large student groups.

- **Ensures Uniform Evaluation Standards**

  Organizes answers by question and section for structured assessment. Facilitates direct comparison with model answers, improving consistency in marking.

- **Centralized Data Management**

  Stores extracted answers, marks, and teacher comments in a secure database. Enables easy retrieval for audits, performance tracking, or future reference.

- **Improves Accessibility and Flexibility**

  Allows teachers to evaluate answer sheets from any location via a web interface. Supports multiple file formats (PDF, images), adapting to various institutional workflows.

- **Environmentally Friendly**

  Reduces reliance on physical paper storage, supporting eco-friendly practices. Minimizes risk of data loss associated with traditional paper-based evaluation.

- **Modular and Scalable Design**

  Components for uploading, extracting, and displaying answers can be updated or extended independently.Can be integrated with LMS, analytics dashboards, or other educational tools.

# REFERENCES


[1] Han, J., Kamber, M., & Pei, J., *Data Mining: Concepts and Techniques*, Morgan Kaufmann Publishers, 3rd Edition, 2012.

[2] Mitchell, T. M., *Machine Learning*, McGraw-Hill Education, 1st Edition, 1997.

[3] Alpaydin, E., *Introduction to Machine Learning*, MIT Press, 4th Edition, 2020.

[4] Russell, S. J., & Norvig, P., *Artificial Intelligence: A Modern Approach*, Pearson, 4th Edition, 2020.

[5] Bishop, C. M., *Pattern Recognition and Machine Learning*, Springer, 1st Edition, 2006.

[6] Aggarwal, C. C., *Data Mining: The Textbook*, Springer, 1st Edition, 2015.

[7] Géron, A., *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*, O'Reilly Media, 2nd Edition, 2019.

[8] McKinney, W., *Python for Data Analysis*, O'Reilly Media, 2nd Edition, 2017.

[9] Provost, F., & Fawcett, T., *Data Science for Business*, O'Reilly Media, 1st Edition, 2013.

[10] James, G., Witten, D., Hastie, T., & Tibshirani, R., *An Introduction to Statistical Learning*, Springer, 2nd Edition, 2021.