OMR System Empowered by Machine Learning and Image Processing

Rushikesh G. Dongare[1], Prof. Tanuja S. Mulla[2], Gauri Y. Kakulte[3], Manjiri M. Tamkar[4], and Shraddha S. Tambekar[5]

[1]Bachelor of Engineering Student, Smt. Kashibai Navale College of Engineering, Pune

[2]Assistant Professor, Smt. Kashibai Navale College of Engineering, Pune

[3,4,5] Bachelor of Engineering Student, Smt. Kashibai Navale College of Engineering, Pune

Abstract - This study affords the design and development of an automatic Optical Mark recognition (OMR) system aimed toward revolutionizing the grading procedure for examination answer sheets. The manual managing of OMR sheets has long been an undertaking, often resulting in mistakes and delays. through incorporating advanced image recognition and machine learning algorithms, our OMR system automates the complete grading technique, drastically enhancing velocity, precision, and scalability. The machine processes scanned OMR solution sheets and correctly evaluates them, ensuring real-time comments and scalability for each school room and huge-scale aggressive tests. Our experimental results exhibit the system's capacity to gain a 98% accuracy price whilst keeping rapid processing speeds, making it a treasured device in academic environments.

Index Terms - Automated evaluation, Image processing, Machine learning, Optical mark recognition.

INTRODUCTION

Because the demand for big-scale, more than one-desire question (MCQ) assessments continues to grow, conventional guide grading strategies have come to be increasingly inadequate. Evaluators face several demanding situations, together with time inefficiencies and the capability for human mistakes, while dealing with excessive volumes of OMR sheets. these problems can influence the integrity and reliability of examination outcomes, underscoring the need for an automatic answer. Optical Mark reputation (OMR) era offers a promising alternative by automating the evaluation of answer sheets. thru using photo processing and gadget learning strategies, OMR structures can quick and as it should be hit upon marked solutions, providing educators and examiners with well timed, blunders-free effects. This paper explores the improvement of an automated OMR system that goals to conquer the constraints of guide grading, handing over advanced accuracy, faster turnaround times, and more desirable scalability for each small and big exam settings.

LITERATURE SURVEY

OMR structures have developed from simple sample recognition to superior system gaining knowledge of primarily based answers, appreciably improving grading accuracy for more than one-choice tests. Early structures struggled with troubles like image pleasant and alignment, but recent advancements in picture processing and device learning, particularly the usage of help vector machines (SVM) and deep studying fashions, have improved mark detection even in suboptimal conditions. these structures now provide extra scalability and adaptability to numerous exam codecs. however, challenges remain in coping with poor-best scans, ensuring safety, and permitting actual-time processing, that are present day areas of ongoing studies.

MATERIALS

The improvement and implementation of the Optical Mark reputation (OMR) machine for automated examination assessment worried a structured technique to make sure accuracy, performance, and scalability. This segment details the substances and techniques used, together with the equipment, strategies, algorithms, and techniques worried in creating the gadget.

I. Software Tools

- Python**:** The backend of the OMR system was developed using Python, which is well-suited for data processing and machine learning applications. Python libraries such as OpenCV were extensively used for image processing, while machine learning algorithms were implemented using Scikit-learn.
- OpenCV: This open-source computer vision library was used for the image processing tasks such as binarization, noise removal, and edge detection, crucial for preparing the OMR sheets for accurate mark detection.

- Flask: Flask was used to develop the web interface for the OMR system, allowing users to upload answer sheets, process results, and download evaluation reports.
- Pandas & NumPy: These libraries were utilized for handling data structures and efficiently processing the CSV-based answer keys and result sheets.

II. Hardware Tools

- Standard Computer System: A standard desktop with adequate computational power was used for development, testing, and processing. Specific system requirements may vary depending on the number of sheets being processed, but the system was built to handle high workloads effectively.

III. Datasets

- OMR Sheets: A dataset consisting of scanned OMR sheets in common formats (JPG, JPEG, PNG) was used for training and testing the system. These sheets varied in quality, marking intensity, and layout to ensure the robustness of the detection algorithms.
- Answer Keys: The correct answers for each OMR sheet were stored in CSV files. This structured format allowed for easy matching between the detected marks on the OMR sheets and the correct answers during the evaluation phase.

METHODS

The development process of the OMR system follows a structured workflow that ensures the efficient and accurate processing of OMR sheets from scanning to result generation. The workflow is segmented into the following phases:

I. Image Preprocessing

The preprocessing stage is critical for preparing the scanned OMR sheets for analysis. This step involves several image processing techniques to enhance the quality and readability of the sheets:

- Convolutional Neural Network (CNN) Implementation: The system employs a Convolutional Neural Network (CNN) to enhance mark detection accuracy on the OMR sheets. The CNN is designed to process the input images through multiple layers, where it learns to identify patterns and features associated with marked and unmarked bubbles. This hierarchical feature extraction allows the network to effectively handle variations in marking styles and image quality, leading to improved classification performance. By training on a diverse dataset of labeled OMR images, the CNN adapts to different marking intensities and conditions, thereby minimizing misclassifications and increasing overall grading reliability.
- Grayscale Conversion: The input images are first converted to grayscale to simplify processing. This reduces the complexity of the image by eliminating color information and focusing on intensity variations.

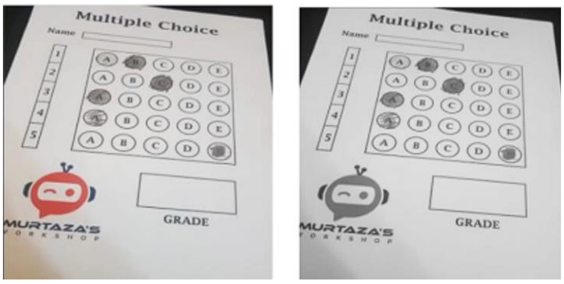

Figure I: Grey Scale Conversion of Image

- Noise Reduction: Gaussian filters are applied to the images to reduce noise. Noise can interfere with mark detection, particularly in poorly scanned or old OMR sheets. By smoothing the image, we reduce the likelihood of false positives in mark detection.
- Binarization: After noise reduction, binarization is performed to convert the image into a binary format (black and white). Thresholding techniques are used to separate the marked areas (black) from the background (white). Otsu's thresholding method is commonly employed to determine the optimal threshold automatically.
- Skew Correction: OMR sheets are often scanned at slight angles, which can distort the positioning of the marks. The Hough Transform method is applied to detect and correct any skew in the images, ensuring that the OMR sheet is properly aligned before proceeding to mark detection.

II. Mark Detection

In the mark detection phase, the system utilizes a Convolutional Neural Network (CNN) to enhance mark detection accuracy on the OMR sheets. The CNN is designed to process the input images through multiple layers, where it learns to identify patterns and features associated with marked and unmarked bubbles. This hierarchical feature extraction allows the network to effectively handle variations in marking styles and image quality, leading to improved classification performance. By training on a diverse dataset of labeled OMR images, the CNN adapts to different marking intensities and conditions, thereby minimizing misclassifications and increasing overall grading reliability. Additionally, the Support Vector Machine (SVM) is employed as a complementary method, potentially for final classification based on the features extracted by the CNN or to provide an additional layer of decision-making in the evaluation process. This combination leverages the strengths of both algorithms, enhancing the accuracy and reliability of the mark detection process.

III. Answer Matching and Scoring

In the answer matching and scoring phase, the system compares the detected marks on the OMR sheet against the answer key provided in CSV format. Each detected mark is matched with its corresponding question, and the system evaluates whether the response is correct, incorrect, or unattempted. The scoring algorithm calculates the total score by summing up the correct answers and accounting for any incorrect responses, with support for negative marking if required. This dynamic matching process ensures flexibility, allowing the system to adapt to different exam formats and grading rules, ultimately producing accurate scores and detailed evaluation reports.

III. Result Generation and Reporting

In the result generation and reporting phase, the system compiles the evaluation data into a structured CSV file, detailing the scores and performance metrics for each OMR sheet. This data includes the total score, correct, incorrect, and unattempted responses. The results are then presented through a user-friendly web interface, allowing users to view the evaluations in real-time and download the reports for further analysis. The system is designed to handle bulk processing, ensuring that even large volumes of OMR sheets can be processed efficiently, with results stored securely for future reference or analysis.

IV. Testing and Validation

The system was rigorously tested using a diverse dataset of 500 OMR sheets. These sheets were selected to represent a variety of real-world conditions, including variations in marking intensity, skew, and noise. The performance of the system was evaluated based on accuracy, speed, and scalability.

- Accuracy Testing: The accuracy of the system was measured by comparing its results to manually verified answers. The system consistently achieved a 98% accuracy rate across different conditions.
- Speed Testing: The average processing time per sheet was approximately 0.5 seconds. This makes the system suitable for large-scale examinations, where thousands of sheets need to be processed in a short time frame.
- Scalability Testing: The system's ability to handle high workloads was tested by processing batches of 10,000 OMR sheets. The system maintained high performance and responsiveness, demonstrating its capacity for scalability in large assessment environments.

V. Instant feedback based on analysis using AI-model

Once the candidate's OMR sheet is processed and evaluated, the AI model will analyze their performance across various topics and skill areas. Using these insights, the AI will design a personalized study roadmap that highlights strengths, pinpoints areas needing improvement, and recommends targeted resources or practice exercises. The roadmap will be structured to guide the candidate through a prioritized sequence of topics, offering suggestions for time allocation, specific study materials, and practice tests to solidify understanding. Additionally, it will include periodic milestones to help track progress, encouraging consistent improvement and ultimately enhancing the candidate's overall performance.

DISCUSSION

To provide a comprehensive understanding of the OMR scanner system's functionalities and interactions with its users, we employed a use case diagram. This visual representation offers a clear overview of the system's key features and the roles of different actors involved.
The diagram illustrates two primary actors: Students and Administration. Students are responsible for registering, scanning answer sheets, and viewing individual results. Administrators manage user accounts, scan answer keys, and analyze overall results.

The core use cases within the system include Registration, Login, Scan & Upload Answer Sheet, Scan & Upload Answer Key, Display Individual Result, and Display Result Analysis. The "Scan & Upload Answer Sheet" and "Scan & Upload Answer Key" use cases are closely related, as the answer key is necessary for grading the answer sheets. Similarly, the "Display Individual Result" and "Display Result Analysis" use cases are interconnected, as individual results contribute to the overall analysis.

The use case diagram effectively defines the boundaries of the OMR scanner system, outlining the interactions between the system and its users. By analyzing this diagram, we can gain valuable insights into the system's design, functionality, and potential for improvement.

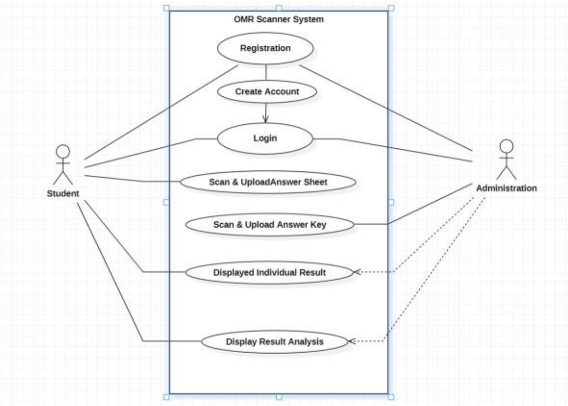

Figure II: Use Case Diagram

CONCLUSION AND FUTURE SCOPE

In conclusion, the proposed OMR scanning system aims to provide a user-friendly and efficient solution for processing and evaluating OMR sheets, thereby automating the grading process to minimize errors and enhance the reliability of assessments in educational environments. Looking ahead, several future enhancements are planned, including the development of real-time processing capabilities that will allow for instant feedback during live quizzes and exams, along with a mobile application that enables users to scan and process OMR sheets using their smartphones, making the system more accessible and convenient. To further protect sensitive data, advanced encryption and authentication methods will be implemented. Additionally, the system will introduce detailed analytics features that provide educators with valuable insights into student performance, helping to identify trends and target areas for improvement. An exciting prospect involves integrating artificial intelligence (AI) to facilitate personalized learning experiences, where AI algorithms analyze individual performance data to tailor feedback and suggest resources, thereby enhancing student engagement and support. Finally, integrating the OMR system with existing Learning Management Systems (LMS) will create a more comprehensive solution that seamlessly fits into modern educational workflows, ensuring the system remains at the forefront of educational technology and contributes meaningfully to the evolution of assessment practices.

REFERENCES

[1] Supriya N. Kulmethe, Hirendra Hajare "E-assessment Using Image Processing in Infinity Exam", International Journal of Advance Research, Ideas and Innovations in Technology (IJARIIT) - 2020.

[2] G. Himabindu, A. Reeta, A. Srinivas, Manikanta S. Manogna "Evaluation of Optical Mark Recognition (OMR) Sheet Using Computer Vision", International Research Journal of Modernization in Engineering Technology and Science (IRJMETS) - 2023.

[3] Nikita Kakade, Dr. R. C. Jaiswal "OMR Sheet Evaluation Using Image Processing", Journal of Emerging Technologies and Innovative Research (JETIR) - 2017.

[4] Supriya N. Kulmethe, Hirendra Hajare "E-assessment using image processing in infinity exam", International Journal of Advance Research, Ideas and Innovations in Technology (IJARIIT) - 2020.

[5] Dhananjay Kulkarni, Ankit Thakur, Jitendra Kshirsagar, Y. Ravi Raju "Automatic OMR Answer Sheet Evaluation using Efficient & Reliable OCR System", International Journal of Advanced Research in Computer and Communication Engineering (IJARCCE) - 2017.