



ASIAN BULLETIN OF BIG DATA MANAGEMENT

<http://abbdm.com/>

ISSN (Print): 2959-0795

ISSN (online): 2959-0809

Real-time Bubble Sheet Detection and Evaluation Using Object Recognition

Nimra Razzaq*, Infal Farooq, Khadija Noor, Sadia Abdullah, Rana Hassan Ajmal, Ali Murtaza

Chronicle

Article history

Received: Aug 24, 2025**Received in the revised format:** Sept 25, 2025**Accepted:** Oct 3 2025**Available online:** Oct 20 2025**Nimra Razzaq*, Infal Farooq, & Sadia Abdullah**, are currently affiliated with the Department of computer science University of agriculture Faisalabad Pakistan.**Email:** Csnimra@gmail.com**Email:** Infal.khalid@gmail.com**Email:** sadiaabduallah475@gmail.com**Khadija Noor** is currently affiliated with the Department of Software Engineering University of Punjab Pakistan.**Email:** khadijaaranoor2002@gmail.com**Rana Hassan Ajmal** is currently affiliated with the The Millennium Universal College Pakistan.**Email:** hassan.ajmal@tmuc.edu.pk**Ali Murtaza** is currently affiliated with the University of Education Township, Lahore Department of Information Sciences, Pakistan.**Email:**personal.alimurtaza@hotmail.com

Corresponding Author*

Keywords: Image Processing; Object Detection; Bubble Sheet; Object Recognition; Exam Evaluation; Automatic Grading.

Abstract

Optical Mark Recognition (OMR) is a technology widely used in education to digitize large amounts of paper data. It is often used for scoring multiple-choice tests. This can be time-consuming when manually grading. OMR technology facilitates this process by providing a controlled format for students to submit their answers. Without the need for manual scoring, OMR is a well-known data entry method and an important technique in human-computer interaction. It is widely used in assessment tests in universities, colleges, questionnaire forms, and competitive examinations. There are many applications for computer image processing and recognition in today's technology-driven world. Existing OMR techniques have limitations, such as the need for special paper, reliance on high quality scanning to achieve accurate results, and to overcome these challenges this research introduces a new image-based OMR method. This reduces the need for special paper and increasing the detection accuracy of the proposed system which is highly reliable by using advanced object recognition techniques that can detect partially filled bubbles with some Marks are slightly blurry or blurred. Additionally, the research includes an automated questionnaire creation system based on specific needs. The generated question paper comes with an answer key. The system will evaluate the completion of the bubble sheet and create the final output file in Excel format. This technique allows for cost-effective optimization of question papers and guarantees effective evaluation in standard reports. Tests have shown the system's accuracy to be more than 99%, effectively scanning and analyzing filled, partially filled bubbles.

© 2025 The Asian Academy of Business and social science research Ltd Pakistan.

INTRODUCTION

Optical mark recognition (OMR) is the process of capturing human marking information from document forms such as surveys and tests. Often used with multiple choice questions, OMR technology has advanced greatly over the years. The application of optical mark recognition (OMR) technology has been extensively studied to automate the scoring process of multiple-choice tests. A popular form of written evaluation that is thought to be straightforward and simple to grade, especially when working with a large number of students, is the multiple-choice question (MCQ) test. Usually, there are a lot of questions on the tests that cover different subjects. These tests can be evaluated using four-step automated multiple-choice question assessment systems. The answer sheet must first be converted into a digital format, which is essentially done with scanners. As a less expensive option, some systems have recently started using digital cameras. Finding the student ID and

response boxes on the answer sheet is the second step. By using a particular template called an optical answer sheet or scantron sheet, this problem can be fixed Spadaccini et al.[1]and Chai et al. [2]Using pattern or optical character recognition (OCR) techniques to identify the student ID is the third step Fisteus et al. [3] The identification of the pupils' responses is the fourth phase. Usually, this identification is carried out by using simple image processing techniques, presuming that there are no crossed-out or erased answers and that each question has a single dark-filled answer box Chai et al. [2] and Nguyen et al.[4].

These software-based OMR solutions have many advantages, such as reduced costs. Increased scalability and increased flexibility Different from other hardware The OMR software system can be easily updated and upgraded over time. This ensures that the system keeps up with the latest technological advances. Existing research on bubble sheet detection can't detect the bubbles correctly there remain lack of detection between partially filled, blurry, or overfilled bubbles that reduced the overall accuracy also they mainly focus on developing algorithms and software solutions for automatic detection and evaluation of bubbles added to existing bubble sheets.

But there is a gap in the literature describing a comprehensive system that combines the creation of multiple choice questionnaires (MCQs) and bubble sheets. Traditional methods often rely on manual creation of sheets or the use of specialized optical mark recognition (OMR) equipment, which leads to inefficiencies. Resource limitations and inaccuracies that may occur in the verification process. Processing times are significantly more efficient than previously available systems. It also creates questionnaires with predefined data bank. The Contributions of this study are summarized in the following

- Generation of question paper sheets based on question bank dataset.
- Development of algorithm to detect bubbles on the input bubble sheets.
- Formation of bubble sheet evaluation based on detected bubbles with comparison to the data set.

BACKGROUND STUDY

Optical Mark Recognition (OMR) is a process of capturing human-marked data from document forms like surveys and tests, widely used in multiple-choice questions. It has progressed significantly and is now common in educational and testing institutions for easier, faster, more powerful, and cheaper evaluation. OMR systems typically verify answer sheets, detect responses, and compare them to programmed correct answers. OMR-based evaluation is preferred over manual methods for voluminous data, large numbers of sources, multiple-choice questionnaires, and when high accuracy is needed. Various researchers have explored automated grading systems using different approaches. Abbas et al.[11]Compared a mobile OMR app (Zip Grade) with manual checking and found no significant difference in accuracy, suggesting economical OMR technologies can ease instructor burden. Atasoy et al.[13] created a mobile device-based image processing technique using a webcam for real-time automated grading of student response sheets with high success rates under various conditions. Karunanayake et al. [23] developed a cost-effective OMR sheet evaluation method using a cheap web camera and template matching, capable of evaluating any MCQ paper format. Bayar et al.[14] utilized Hough Transform to develop an intelligent grading system for multiple-choice exams, testing it on over 1000 exam papers. Catalan at al.[15] proposed a framework for automated MCQ scoring using readily available software like Octave and spreadsheets,

achieving great scoring accuracy. Khan et al.[24] proposed a user-friendly and cost-effective approach for MCQs marking using a camera and computer, accepting any mark on bubbles and saving significant time. Fahmi et al.[18] used thresholding and region merging with a Webcam for correcting multiple-choice answer sheets with good accuracy, able to detect various markings. Ascencio et al.[12] discussed the implementation of a fully automated exam test grading application using computer vision, Python, and OpenCV for contour and mark detection. Kommey et al.[25] proposed an android software application using image processing to scan patterns and score multiple-choice scripts automatically with a 90% accuracy rate, utilizing OpenCV and Flask in Python.

Himabindu et al.[19] provided a moderately cost image-based OMR approach capable of processing thin papers and low-printing accuracy sheets, achieving results within 60 seconds. Kanjalkar et al. [22] developed an algorithm evaluating student replies with high average accuracy (99.12%) using image processing techniques and providing performance analytics. To the best of our knowledge, Current approaches in the literature cannot capable of detecting blurry and partially filled bubbles also not able to validate results, generate exam question papers then evaluate final results in one go. Moreover, our method considers an additional contribution called "Automatic paper Generation" for the sake of bubble sheet. The proposed method exhibits enhanced accuracy and efficiency while providing unique functionality. The scanned sheets give more than 99% accurate results that making the system as a significant improvement in OMR technology

METHODOLOGY

The foundation of this research endeavor is outlined through the tools, technologies, and procedures employed in detection bubble sheet and evaluation using object recognition techniques. Below is framework of processing.

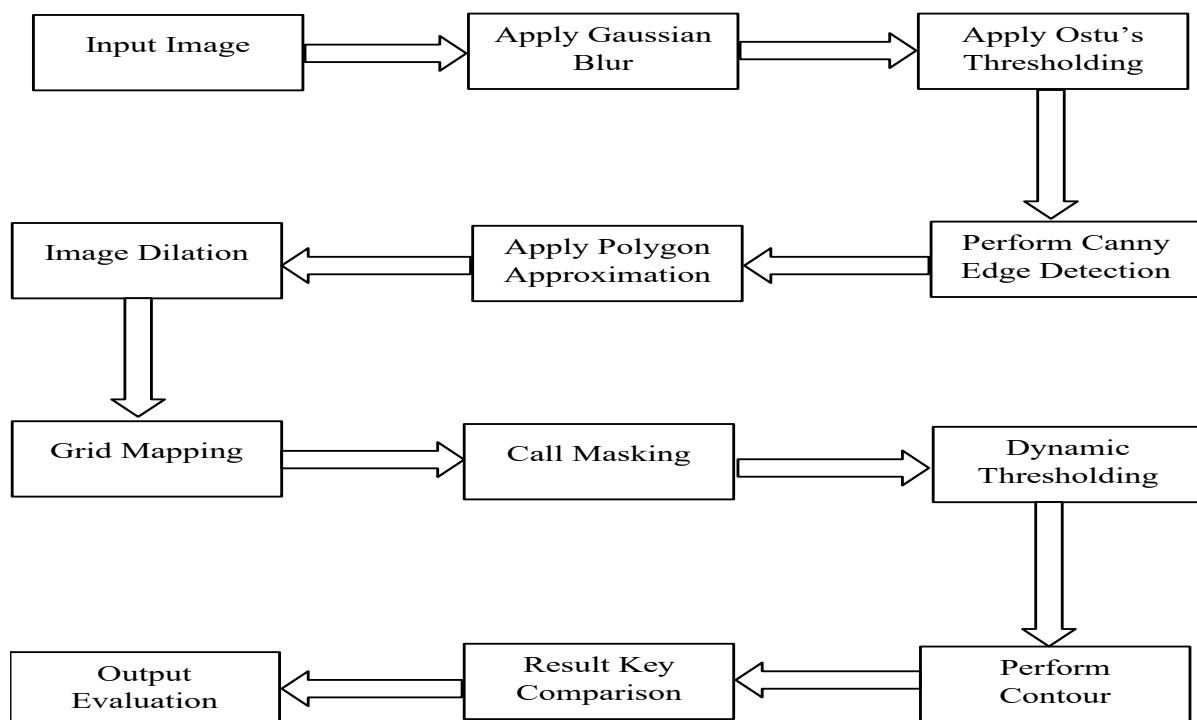


Figure 1.
Framework of bubble sheet processing Diagram

The dataset used in this study was specifically collected and tailored to evaluate and optimize the bubble sheet detection system. This dataset was designed to include various layouts, grid structures, and marker configurations, offering a unique challenge to test the robustness and adaptability of the system. Unlike existing datasets, the curated dataset for this study emphasizes compatibility with an open-source approach, ensuring flexibility and scalability for diverse educational contexts. Examples of dataset samples before pre-processing are shown in Figure 2.

Figure 2.
Dataset sample Sheet
Preprocessing

To standardize input images and improve detection accuracy, a series of preprocessing steps were applied:

Noise Removal: High-frequency noise was eliminated using a Gaussian filter, dynamically adjusting the kernel size and σ value based on image resolution. This ensured optimal performance across different image sizes, as described by Equation(1)

$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2 + y^2}{2\sigma^2}}$$

Eq. (1) 2D Gaussian distribution function, where x and y are 0 at the center of the kernel.

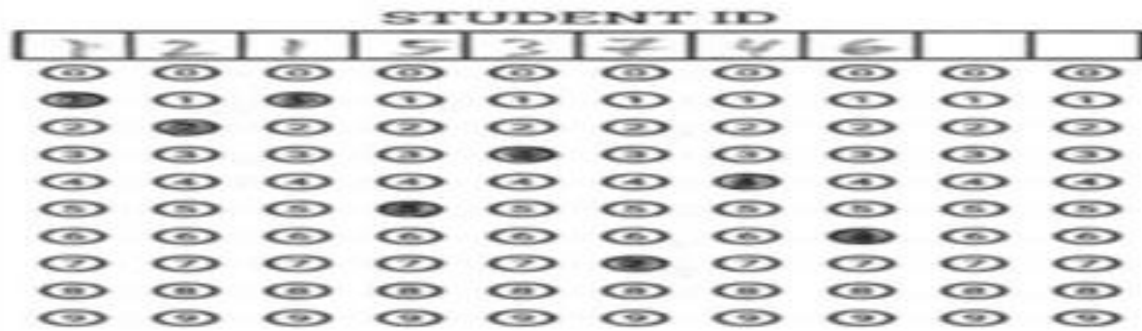


Figure 3.

illustrates the effect of Gaussian noise removal on a sample scanned sheet.

4.2.2 Thresholding: Images were converted to grayscale and binarized using Otsu's thresholding technique. This step enhanced critical features for subsequent processing by classifying pixels as either black or white. An example of this transformation is presented in Figure 4.

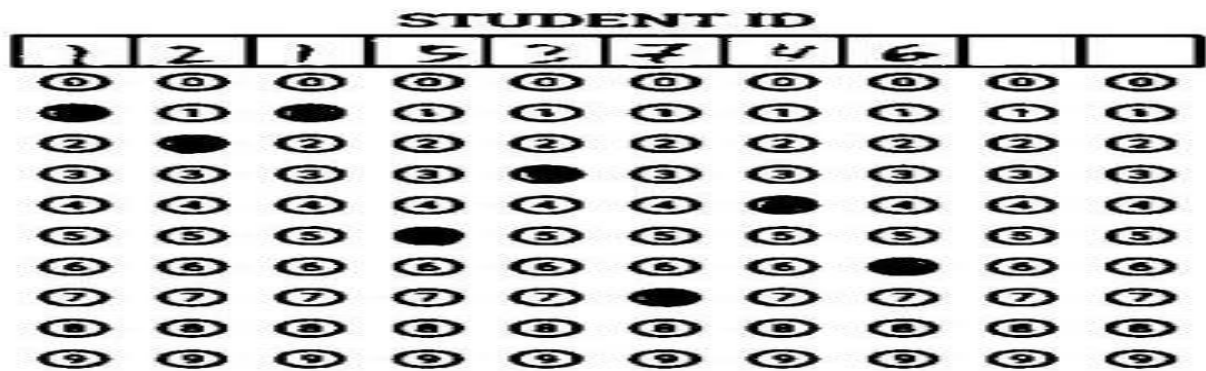


Figure 4.

Sample portion of scanned sheet after thresholding is applied.

Corner and Marker Detection

To accurately establish a coordinate system for evaluating responses, the system identified specific markers on the sheet. Key steps included:

Edge Detection: The Canny edge detection algorithm was used to identify prominent edges by analyzing gradients in the image. Non-maximum suppression and double thresholding ensured precise edge identification while minimizing noise. Figure 5 demonstrates the edge detection process.

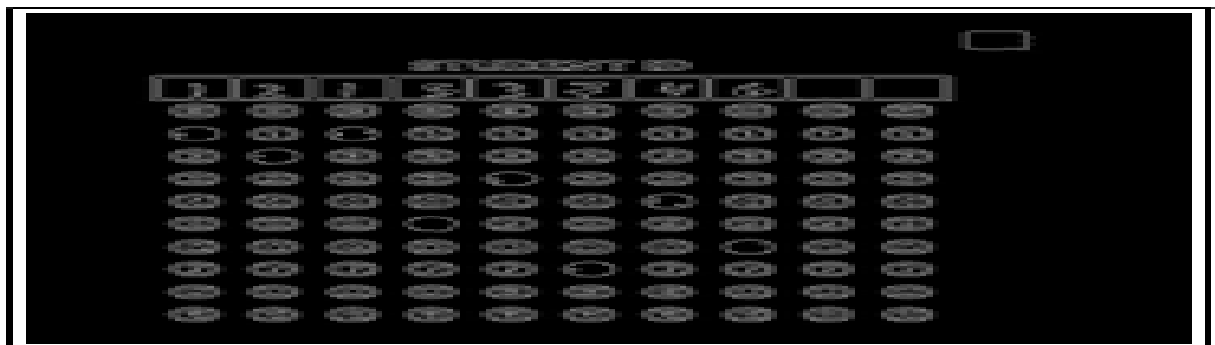


Figure 5.

Sample portion of scanned sheet after edge detection

L-Mark Identification: A unique marker in the top-left corner, referred to as the L-Mark, served as a reference point. The system evaluated polygons based on vertex count,

angles, and side lengths to reliably detect the L-Mark. This innovative approach improved detection accuracy under varying conditions.

Grid Establishment: Using the detected L-Mark and additional corner markers, the system constructed a uniform grid that divided the sheet into cells. Precise coordinate transformations

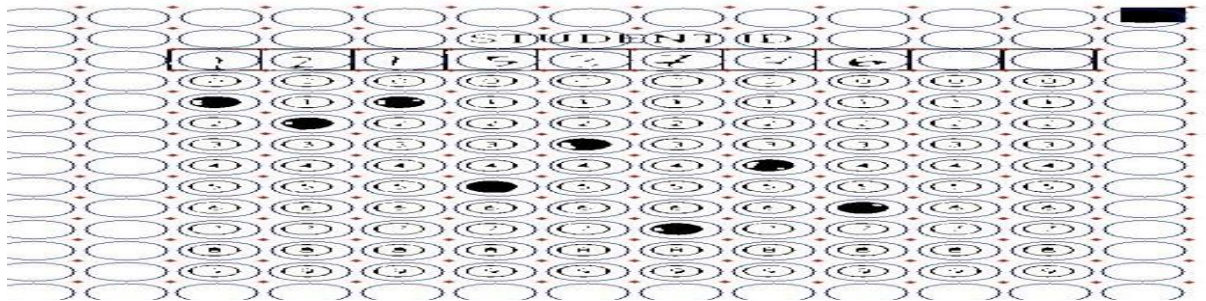


Figure 6.

Demonstration of the grid and cell masks on a sample portion of a scanned sheet.

4.2.6 Image Dilation: To address potential data loss caused by Gaussian blurring, a 3×3 kernel dilation process was applied. This enhanced the distinction between filled and unfilled bubbles, particularly in low-resolution images. Figure 7 shows an example of a dilated image.

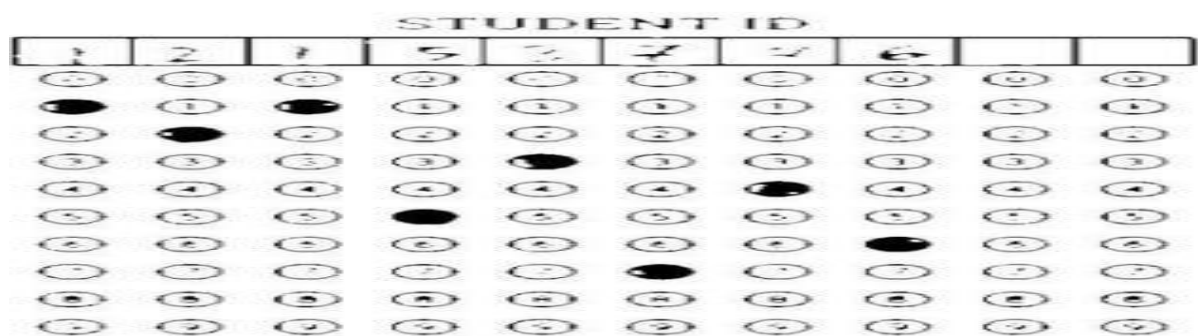


Figure 7.

Sample portion of scanned sheet after dilation is applied.

Cell Masking: Circular masks were applied to each grid cell, isolating relevant areas and removing extraneous information. The mask diameter was set to 75% of the cell width, as illustrated in Figure 8.

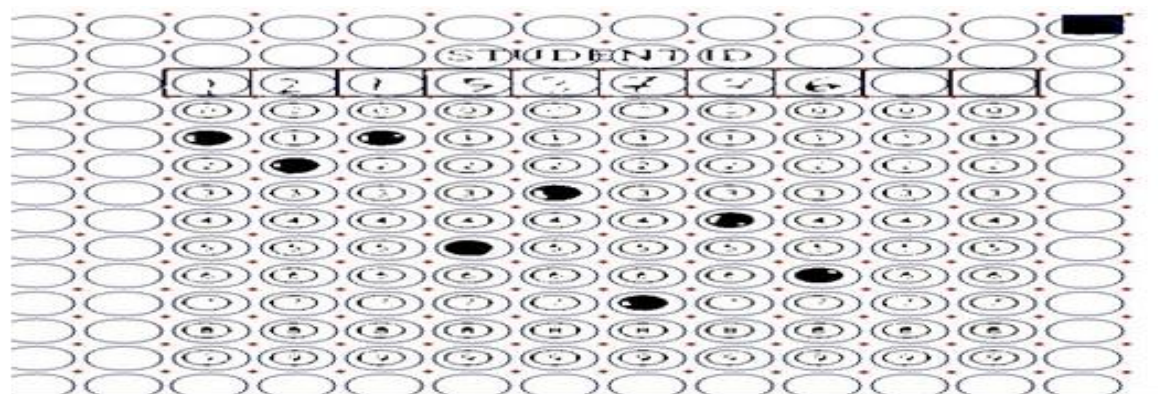


Figure 8.

Demonstration of the grid and cell masks on a sample portion of a scanned sheet

Dynamic Fill Thresholding: The fill fraction of each cell was calculated by dividing the number of black pixels by the total pixels within the mask. A dynamic threshold was computed to classify cells as filled or unfilled. This adaptive approach ensured high accuracy even in variable lighting and printing conditions.

Framework Implementation

The study introduced an integrated framework that combines multiple technologies to achieve efficient and accurate bubble sheet detection a PHP Application web-based tool for generating customization multiple-choice question papers. The tool's intuitive interface streamlined the process for educators, making it accessible to users with minimal technical expertise. Python-Based Algorithms do Core detection tasks, such as image preprocessing, edge detection, and bubble analysis, were implemented using Python libraries like OpenCV and NumPy. These algorithms optimized processing speed and accuracy using .NET Platform A supplementary application was developed for advanced image analysis and seamless integration with existing educational tools, enhancing the overall functionality of the system. Figure 9 displayed the paper generation interface.

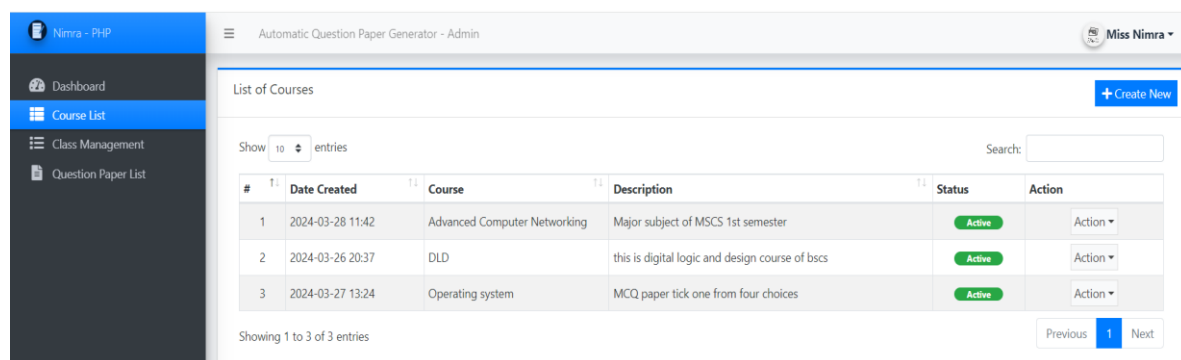


Figure 9.
Automatic Question paper Generation Application.
Evaluation Findings and Discussion

The system's performance was evaluated using accuracy metrics, ensuring reliability and robustness across diverse datasets. Experimental results demonstrated the efficacy of the proposed framework in automating bubble sheet detection and grading processes. Comparative analysis with existing solutions revealed the system's superior accuracy, adaptability, and open-source benefits.

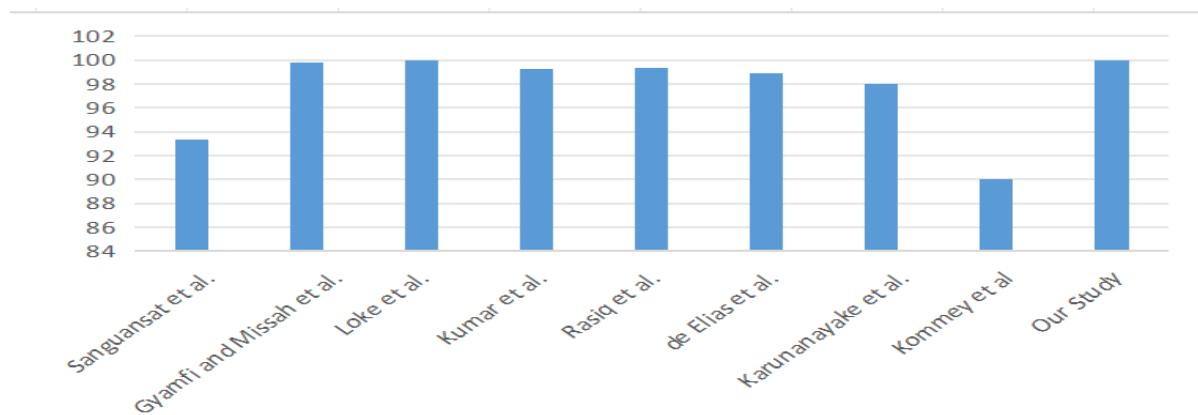


Figure 10.
Proposed system's Accuracy Comparison with Existing Systems

The developed system integrates advanced image processing algorithms for bubble sheet detection. This includes the Canny edge detection technique, which achieves an accuracy rate of 99.98%, significantly outperforming traditional systems by addressing common challenges such as partial marks and irregularities. The algorithm incorporates a robust error-handling mechanism to ensure that ambiguous documents are flagged for review rather than processed incorrectly. Batch processing capabilities allow for the evaluation of multiple documents simultaneously. It has a processing speed of up to 600 documents, faster than sequential systems.

Also automatic MCQ paper generation web application for efficient question management and paper generation demonstrated strong performance Leveraging the MySQL database, the user-friendly interface allows for customizable question selection and streamlined output in PDF format, ensuring seamless integration into the study workflow. Overall, this system contribute to greater accuracy Faster processing and efficiency present a practical, reliable and innovative solution to test management. System performance Especially in terms of accuracy and processing speed makes the system different from existing solutions. These characteristics are further illustrated in the table 1 that represent clear comparison of proposed study with existing literature.

Table 1.
Results Comparison Proposed Study with Existing

Study	year	Methodology	Accuracy (%)	Reading Speed(seconds)	Special Feature
Rasiq et al.	2020	Grid based smart phone image	99.40	Uncertain	No hardware required
Ascencio et al.	2021	Grading with OpenCV	97	Uncertain	Fully automatic grader
Kommay et al.	2022	Android image processing	~ 90.0	Uncertain	Mobile app
Largo et al.	2022	Mobile OMR	90.1	Uncertain	Fast mobile based
Faith et al.	2023	C# Web+ OpenCv	~99.0	Parallal fast	Intermediate library+web framework
Himabindu et al.	2023	Computer Vision	95	~60 sec(batch)	Thin paper and low quality scan handling
Kanjalkar et al.	2023	Python with Image processing	99.12	Uncertain	Perform analytics integration
Chen and zeng	2024	YOLOv8 Bubble detection	>99	Uncertain	Real time multi industry application
Our Study	2025	Object recognition image processing	99.98	0.10-25	Partial/blurry bubble handling fast result

The figure 11 shows speed comparison of the proposed system with existing solutions. It significantly outperforms traditional sequencing systems. This result highlights the system's efficiency and suitability for large-scale operations. These results highlight the practical advantages of the proposed system. This leads to a reliable and scalable solution for automated assessment processes.

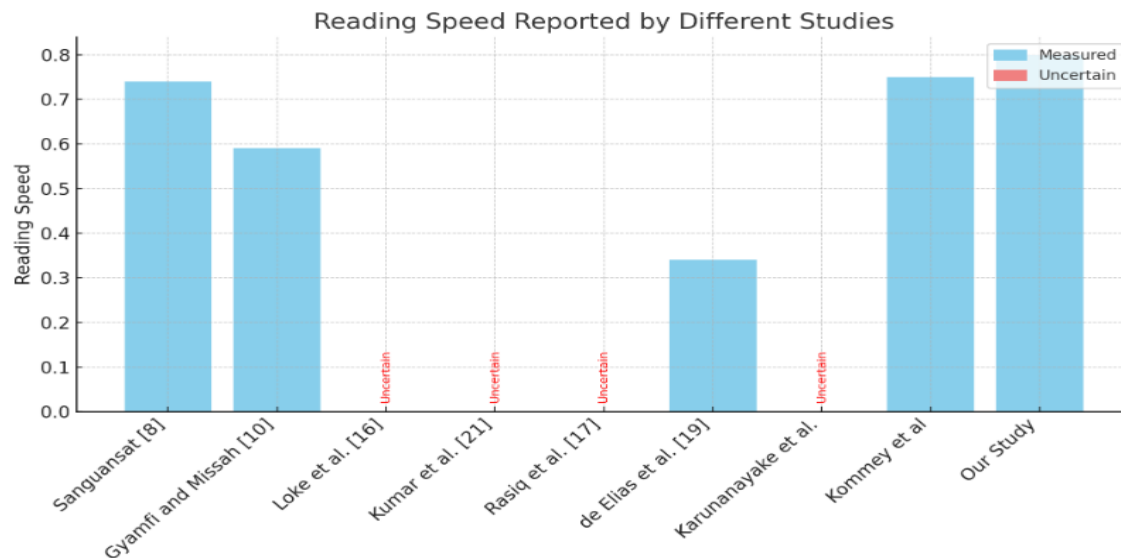


Figure 11.
Speed Comparison

CONCLUSION

This research presents a state-of-the-art framework for automatic MCQ paper generation and Bubble sheet evaluation, highlighting algorithmic advances in addressing limitations in previous work. Dynamic fill threshold algorithm which adaptively evaluates bubble markings evaluation according to different filling ratios which is inconsistent Even in the situation as well as setting a new standard for accurately distinguishing between marked and unmarked responses. A custom polygon-based corner detection algorithm ensures accurate localization of key marks. Previous methods struggled to deal with challenges such as rotation or partial Scanned characters. enhanced pre-processing techniques such as Dynamic Gaussian filtering and Otsu's threshold to further refine image quality Strong noise reduction and optimal sharpness are guaranteed for evaluation. Together, these innovations offer a state-of-the-art system. It integrates seamlessly with modern technologies like OpenCV, Python, and MySQL. By filling in the gaps in previous methods and introducing scalable and adaptable solutions. This research provides transformational tools for educational institutions. Improve the reliability and accuracy of automatic assessment systems.

DECLARATIONS

Acknowledgement: We appreciate the generous support from all the contributor to the research and their different affiliations.

Funding: No funding body in the public, private, or nonprofit sectors provided a particular grant for this research.

Availability of data and material: In the approach, the data sources for the variables are stated.

Authors' contributions: Each author participated equally in the creation of this work.

Conflicts of Interest: The authors declare no conflict of interest.

Consent to Participate: Yes

Consent for publication and Ethical approval: Because this study does not include human or animal data, ethical approval is not required for publication. All authors have given their consent.

REFERENCES

- A. Spadaccini and V. Rizzo, "A multiple-choice test recognition system based on the Gamera framework," arXiv preprint arXiv:1105.3834, 2011.
- D. Chai, "Automated marking of printed multiple choice answer sheets," in Proc. IEEE Int. Conf. Teaching, Assessment, and Learning for Engineering (TALE), 2016, pp. 145–149.
- J. A. Fisteus, A. Pardo, and N. F. García, "Grading multiple choice exams with low-cost and portable computer-vision techniques," J. Sci. Educ. Technol., vol. 22, pp. 560–571, 2013.
- T. D. Nguyen, Q. H. Manh, P. B. Minh, L. N. Thanh, and T. M. Hoang, "Efficient and reliable camera-based multiple-choice test grading system," in Proc. Int. Conf. Advanced Technologies for Communications (ATC), 2011, pp. 268–271.
- S. Hussmann and P. W. Deng, "A high-speed optical mark reader hardware implementation at low cost using programmable logic," Real-Time Imaging, vol. 11, no. 1, pp. 19–30, 2005.
- A. M. Smith, "Optical mark reading—making it easy for users," in Proc. 9th Annu. ACM SIGUCCS Conf. User Services, 1981, pp. 257–263.
- D. Lopresti, G. Nagy, and E. B. Smith, "A document analysis system for supporting electronic voting research," in Proc. 8th IAPR Int. Workshop on Document Analysis Systems (DAS), 2008, pp. 167–174.
- P. Sanguansat, "Robust and low-cost optical mark recognition for automated data entry," in Proc. 12th Int. Conf. Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), 2015, pp. 1–5.
- R. T. China, F. A. Zampiroli, R. P. O. Neves, and J. A. Quilici-Gonzalez, "An application for automatic multiple-choice test grading on Android," Rev. Bras. Inic. Cient., vol. 3, no. 2, 2016.
- E. O. Gyamfi and Y. M. Missah, "Pixel-based unsupervised classification approach for information detection on optical markup recognition sheet," Adv. Sci. Technol. Eng. Syst. J., vol. 2, no. 4, pp. 121–132, 2017.
- A. Abbas, "An automatic system to grade multiple choice questions paper based exams," J. Univ. Anbar Pure Sci., vol. 3, pp. 174–181, 2009.
- H. E. Ascencio, C. F. Peña, K. R. Vásquez, M. Cardona, and S. Gutiérrez, "Automatic multiple choice test grader using computer vision," in Proc. 3rd IEEE Mexican Humanitarian Technology Conf. (MHTC), 2021, pp. 65–72.
- H. Atasoy, E. Yildirim, Y. Kutlu, and K. Tohma, "Webcam based real-time robust optical mark recognition," in Lecture Notes in Computer Science, vol. 9490, Springer, 2015, pp. 449–456.
- G. Bayar, "The use of Hough Transform to develop an intelligent grading system for the multiple choice exam papers," Karaelmas Fen Müh. Derg., vol. 6, pp. 100–104, 2016.
- J. A. Catalan, "A framework for automated multiple-choice exam scoring with digital image and assorted processing using readily available software," in Proc. DLSU Research Congress, 2017, pp. 1–5.
- V. Chidrewar, J. Yang, and D. Moon, "Mobile based auto grading of answer sheets," unpublished manuscript, 2013.
- R. C. Dharmik, S. Rangari, S. Jain, A. Nilawar, G. Deshmukh, and A. Yeole, "Optical mark recognition evaluation system using dual-component approach," Int. J. Intell. Syst. Appl. Eng., vol. 12, pp. 349–353, 2024.
- H. Fahmi, M. Zarlis, H. Mawengkang, N. Zandrato, and Sulindawaty, "The using of thresholding and region merging algorithm for correcting the multiple choice answer sheets," J. Phys. Conf. Ser., vol. 1255, 2019.
- G. Himabindu, A. Reeta, A. S. Manikanta, and S. Manogna, "Evaluation of optical mark recognition (OMR) sheet using computer vision," Int. Res. J. Mod. Eng. Technol. Sci., pp. 5–9, 2023.
- A. Info, "Image processing-based automated multiple-choice grading system," Int. J. Sci. Res. Eng. Technol., vol. 11, pp. 64–73, 2023.

- N. Kakade and D. R. C. Jaiswal, "OMR sheet evaluation using image processing," J. Emerg. Technol. Innov. Res., vol. 4, pp. 640–643, 2017.
- P. Kanjalkar, P. Chinchole, A. Chitre, J. Kanjalkar, and P. Sharma, "Economical solution to automatic evaluation of an OMR sheet using image processing," in Proc. Int. Conf. Data Management, Analytics & Innovation, 2023, pp. 665–683. Springer.
- N. Karunanayake, "OMR sheet evaluation by web camera using template matching approach," Int. J. Res. Emerg. Sci. Technol., vol. 8, pp. 15–18, 2015.
- I. Khan, S. U. Rahman, and F. Alam, "An efficient, cost effective and user friendly approach for MCQs treatment," Proc. Pakistan Acad. Sci. Part A, vol. 55, pp. 39–44, 2018.
- B. Kommey, E. Keelson, F. Samuel, S. Twum-Asare, and K. K. Akuffo, "Automatic multiple choice examination questions marking and grade generator software," IPTEK J. Technol. Sci., vol. 33, no. 3, pp. 175–180, 2022.



2025 by the authors; The Asian Academy of Business and social science research Ltd Pakistan. This is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).