iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 Issue: XII Month of publication: December 2025

DOI: https://doi.org/10.22214/ijraset.2025.76285

Design And Implementation of an Automated OMR Analyser Using ML and Image Processing

Ravindra S[1], Ranganath K N[2], Nithin A[3], Manoj M H[4], M S Hemanth Kumar[5]
[1]Associate Prof. [2,3,4,5]Final Year Students Dept. of Computer Science and Engineering, Jawaharlal Nehru New College of Engineering, Shivamogga., Visvesvaraya Technologial University – Belagavi

Abstract: Automated evaluation of OMR (Optical Mark Recognition) answer sheets plays a significant role in academic and competitive examinations. Traditional OMR systems depend on expensive scanners and strict sheet templates, leading to operational limitations and high processing costs. This work presents a web-based OMR evaluation system using Machine Learning and Image Processing techniques to accurately detect filled bubbles from scanned or photographed sheets. The system integrates OpenCV for preprocessing and a Convolutional Neural Network (CNN) model for classification of filled and unfilled bubbles. The solution supports custom OMR formats, handles variations in lighting, shading, and sheet alignment, and delivers fast and accurate results with real-time visualization. Experimental results show detection accuracy of above 98%, demonstrating its effectiveness for large-scale examination environments..
Keywords: OMR Evaluation, Image Processing, Convolutional Neural Network, OpenCV, Automated Assessment System.

I. INTRODUCTION

OMR technology is widely used in educational institutes and examination authorities to evaluate multiple-choice question (MCQ) answer sheets. Traditional OMR machines require specialized hardware scanners and predefined sheet formats, making them costly and inflexible. In addition, manual checking is prone to high error rates, slow processing, and human fatigue. The proposed system aims to address these limitations by building an automated and cost-effective solution using Machine Learning and Image Processing. The system accepts scanned or camera-captured OMR sheets, preprocesses the image for clarity, extracts bubble regions, and classifies responses using a trained CNN model. The implemented web-based platform enables users to upload templates, key answer files, and filled sheets, generating instant results and downloadable reports.

II. MACHINE LEARNING AND IMAGE PROCESSING

Machine Learning (ML) and Image Processing (IP) form the foundational backbone of automated OMR evaluation systems. Image Processing techniques such as grayscale conversion, Gaussian blurring, adaptive thresholding, contour detection, and perspective correction are applied to enhance input image quality and accurately isolate bubble regions. These operations ensure reliable performance despite variations in lighting, camera angle, resolution, or background noise.Once Region of Interest (ROI) extraction is completed, a trained Convolutional Neural Network (CNN) model is employed to classify each bubble as filled or unfilled. Unlike traditional pixel-counting methods, ML-based classification learns visual characteristics such as shading density, fill texture, pen type variations, and irregular or partial markings. The integration of ML enables intelligent handling of anomalies including multiple marked options, lightly shaded bubbles, and smudged regions, significantly improving detection accuracy and system robustness.

III. LITERATURE REVIEW

Research work presented by various authors related the OMR sheets analyser and evaluation described as below.

Zeki Kucukkara and Abdullah Erdal Tumer (2018) [1] proposed an OMR system that primarily utilizes image processing techniques for answer sheet evaluation. The method begins with grayscale conversion and thresholding to prepare the scanned sheet for analysis. Contour detection and region segmentation are applied to isolate answer bubbles. The system then detects marked responses using pixel intensity-based decision logic. High accuracy is achieved without requiring expensive OMR hardware, making it cost-effective and practical for educational use.

Pooja Raundale et al. (2019) [2] developed an OMR system using OpenCV for efficient and accurate detection of marked responses on answer sheets. The methodology includes image preprocessing steps such as grayscale conversion, Gaussian blurring, and thresholding. Contour detection is used to locate bubbles, followed by sorting and mapping to specific questions. Filled options are identified based on pixel density in each bubble area. This system eliminates the need for specialized scanners and works well with standard webcams and printers.

Harendra Kumar et al. (2019) [3] proposed a machine learning-based approach for analyzing OMR sheets to enhance accuracy and reduce manual effort. Their system begins with image preprocessing, including resizing and noise removal. Features extracted are utilized to detect filled bubbles, and an ML classification model is trained to identify marked responses. The model is capable of handling partially filled and misaligned bubbles. This approach ensures flexible, cost-effective evaluation without the need for high-end scanning devices.

Qamar Hafeez et al. (2023) [4] proposes a robust OMR system designed to handle errors and noise during mark detection. The methodology includes preprocessing steps like image binarization and noise reduction to improve mark clarity. Fault-tolerant algorithm is used that accurately identify partially filled or misaligned marks. The system also integrates adaptive thresholding for dynamic mark detection under varying lighting conditions. Finally, it validates the detected marks against answer keys to ensure reliable evaluation. This paper focuses on improving OMR accuracy in imperfect scanning environments.

Effat Somaiya et al. (2023) [5] presents an OMR system using a standard webcam to capture answer sheets, making it cost-effective and accessible. The methodology involves image acquisition through the webcam, followed by preprocessing steps such as grayscale conversion and thresholding to enhance mark visibility. It uses contour detection to identify answer bubbles and classify filled versus unfilled responses. The system incorporates error handling techniques to manage variations in lighting and sheet alignment. This approach aims to provide a practical OMR solution for educators with limited resources.

L Pham Doan Tinh and Ta Quang Minh (2024) [6] introduces a fast and efficient scoring system for paper- based MCQ tests using object detection techniques. The methodology employs a Convolutional Neural Network (CNN) to quickly detect and classify filled answer bubbles on scanned sheets. Preprocessing includes image normalization and alignment correction to ensure consistent input quality. The framework also integrates a marking scheme to automatically score and generate results based on detected answers. This system focuses on improving speed and accuracy in large-scale exam scoring.

Rusul Hussein Hasan et al. (2024) [7] proposes an OMR system based on a modified Bi-directional Associative Memory (BAM) neural network for improved pattern recognition. The methodology involves preprocessing scanned answer sheets through noise filtering and normalization. The BAM network is then trained to recognize filled marks by associating input patterns with correct outputs. Fault tolerance is achieved by effectively handling partially marked or distorted bubbles. The paper demonstrates improved accuracy and robustness in recognizing marked answers compared to traditional methods.

Sujal Rooge et al. (2024) [8] presents an optical mark recognition system leveraging classical image processing methods. The methodology includes image acquisition followed by grayscale conversion and adaptive thresholding to separate marked areas from the background. Morphological operations are applied to remove noise and enhance mark contours. The system detects filled bubbles by analyzing pixel intensity and shape features within predefined answer regions. This paper emphasizes simplicity and efficiency, making it suitable for real-time OMR applications with minimal hardware requirements.

Dharmik R. C. et al. (2024) [9] propose an OMR evaluation system that combines two key components: image processing and machine learning for improved accuracy. The methodology starts with image preprocessing including noise removal and normalization, followed by segmentation to isolate answer bubbles. It then applies a dual-component classifier that uses both shape and intensity features to detect marked responses reliably. The system also includes error correction mechanisms to handle ambiguous or partially filled marks. This approach enhances detection precision and reduces false positives in OMR scoring.

Rushikesh G. Dongare et al. (2024) [10] proposes an advanced OMR system using NL algorithms with IP techniques for improved accuracy. The methodology includes image acquisition followed by preprocessing steps such as binarization and noise reduction. Features are extracted from answer bubbles, which are then classified using supervised ML models to distinguish filled from unfilled marks. The system also incorporates data augmentation to improve model robustness against variations in answer sheet conditions. This approach aims to provide a reliable and scalable OMR solution adaptable to diverse testing environments.

IV. SYSTEM ARCHITECTURE AND DESIGN

A. System Architecture

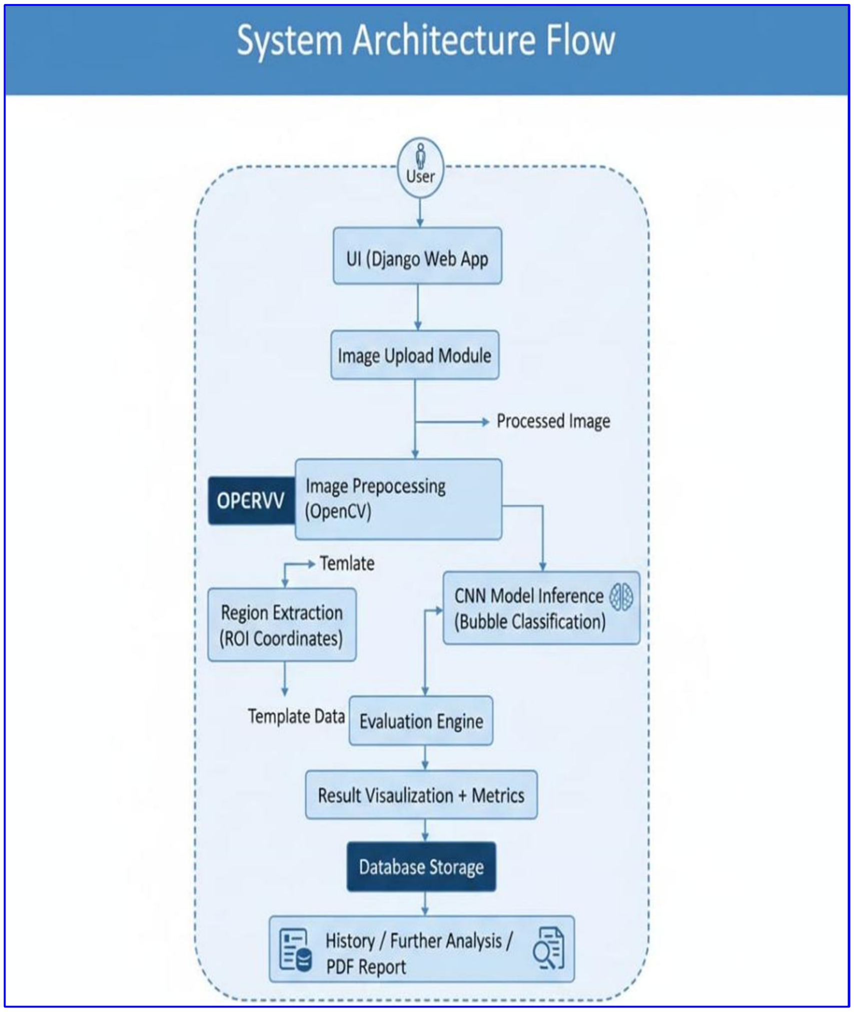


Fig. 1 System Architecture of Web Application

The architecture of the Automated OMR Analyzer system is designed as a multi-layered workflow integrating a web-based interface, image processing pipeline, and machine learning inference engine. The process begins with the user interacting through the Django-based web UI, where blank templates and filled OMR sheets are uploaded. The uploaded image is first forwarded to the Image Preprocessing module, implemented using OpenCV, which performs critical enhancement operations such as grayscale conversion, noise reduction, adaptive thresholding, edge detection, and perspective correction. Once preprocessing is complete, the system proceeds with Region of Interest (ROI) Extraction, where bubble positions from the blank template are identified and stored as coordinate mappings.These extracted template coordinates are supplied to the Evaluation Engine, while each corresponding bubble region from the filled OMR sheet is passed through the CNN Model Inference stage. Here, a trained Convolutional Neural Network classifies each bubble as filled or unfilled based on learned spatial and pixel-intensity features, enabling robust detection even for lightly shaded or partially marked responses. The evaluation engine compares predicted answers with the stored answer key to compute the final score and determine correctness by question. The results are delivered through a visualization layer that displays analytics, scoring metrics, heat-map style correctness views, and detailed student-wise response reports.All processed data and evaluation history are securely stored in the application database, enabling retrieval, search-based filtering, downloadable PDF report generation, and further statistical analysis. The modular architecture ensures platform independence, scalability for bulk evaluation, and adaptability to custom OMR sheet formats without requiring specialized scanning hardware.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue XII Dec 2025- Available at www.ijraset.com

B. Design

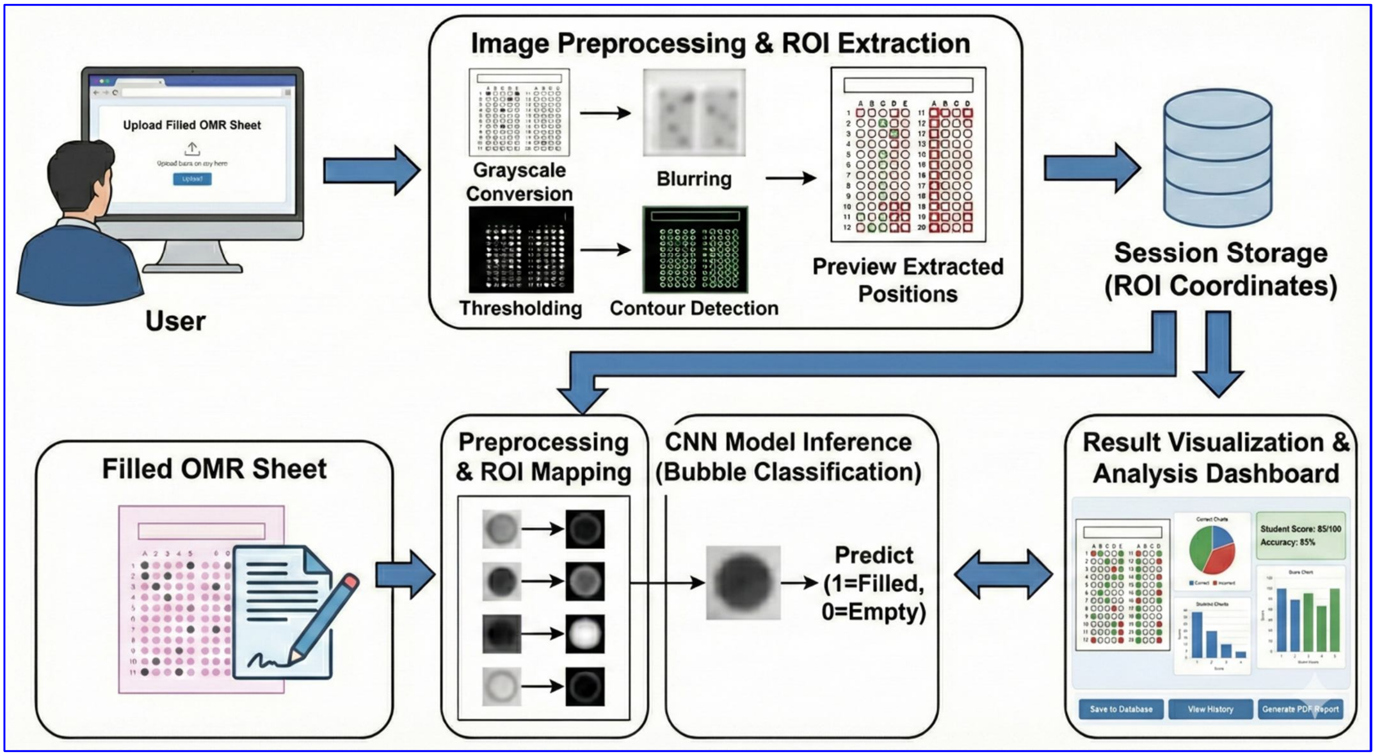


Fig. 2 Operational Workflow of the OMR Analyser System

The design of the Automated OMR Analyzer system is structured to provide a seamless and accurate workflow for evaluating multiple–choice answer sheets without specialized hardware. The system is implemented as a web–based platform where users can upload both blank and filled OMR sheets, which are processed through a sequence of image preprocessing, ROI extraction, and Machine Learning–based classification.

The workflow begins when the user uploads a blank OMR sheet. This sheet undergoes Image Preprocessing, including grayscale conversion, noise reduction through Gaussian blurring, adaptive thresholding, and contour detection to isolate the bubble regions. The extracted bubble coordinates, referred to as Region of Interest (ROI), are stored temporarily in session storage or database for further evaluation.

Once the ROI positions are defined, the user uploads a filled OMR sheet. The system preprocesses the image in the same pipeline to ensure uniformity and maps each detected bubble region against the previously extracted ROI coordinates. Each segmented bubble image is then passed into the Convolutional Neural Network (CNN) model, which classifies each bubble as filled (1) or unfilled (0) based on trained dataset inference.

The output predictions are compared with the uploaded key answers, and the system calculates the final score. The results are then presented through a dynamic Visualization and Analysis Dashboard that includes visual marking highlights on OMR sheet, performance charts, and statistical accuracy reports. Users may store results in the database or generate a downloadable PDF report, supporting future analysis and bulk record management.The overall structured workflow ensures automated, fast, and accurate evaluation, capable of handling varying image quality, tilt, shadows, smudges, and flexible OMR formats.

V. IMPLEMENTATION METHODOLOGY

The implementation of the Automated OMR Analyzer is carried out in a structured and modular approach to ensure scalability, accuracy, and ease of integration. The system is developed as a web-based application using Django, incorporating image processing through OpenCV and bubble classification via a Convolutional Neural Network (CNN) model. The methodology includes the following stages:

A. System Setup and Environment Configuration

The development environment was configured using Python 3.10 and Django as the primary backend framework. Required libraries such as OpenCV, NumPy, TensorFlow/Keras, Matplotlib, and Pillow were installed for handling image processing, machine learning inference, and visualization. A version-controlled architecture was followed to allow iterative improvement and testing.

B. Image Upload and Input Handling

Users interact with the system through a responsive web interface where they can upload:

- Blank OMR sheet (for template creation)
- Filled OMR sheet (for evaluation)
- Answer key (manual entry or CSV)

The uploaded files are validated and stored securely using Django's file management system.

C. Image Preprocessing Module

The preprocessing module applies transformation techniques to enhance and normalize images for consistent analysis. Steps include:

- Grayscale conversion to simplify image data
- Gaussian blurring to remove noise and uneven pixel distribution
- Adaptive or Otsu thresholding for binary conversion
- Contour detection to isolate bubble boundaries
- Perspective transformation to correct orientation and skew

These operations generate a clean binary representation of the OMR sheet suitable for bubble extraction.

D. Region of Interest (ROI) Extraction and Template Creation

Using contour-based detection, the system identifies the exact coordinates of each bubble in the blank OMR sheet. These ROI positions are stored temporarily or in the database as a template file (JSON) to ensure alignment with filled OMR sheets. This technique enables flexibility in supporting various OMR sheet designs without requiring predefined layouts.

E. Bubble Segmentation and Normalization

For filled OMR sheets, bubbles are cropped according to the stored ROI template and resized to 28×28 pixels, then normalized to prepare input vectors for ML inference. This ensures uniformity and improves model performance.

F. Machine Learning-Based Bubble Classification

A CNN model trained on labeled bubble images is used to classify each bubble as filled (1) or unfilled (0). The model uses:

- Binary Cross-Entropy loss function
- Adam optimizer
- Sigmoid activation for binary classification

Prediction thresholding is applied (≥ 0.5 = Filled, <0.5 = Empty), effectively handling cases such as faint shading, partially filled responses, pencil marks, smudges, and low image quality

.

G. Answer Key Comparison and Score Computation

The predicted values are mapped to corresponding question options and compared with the answer key to calculate:

- Total score
- Correct / incorrect / unanswered responses
- Error cases such as multiple markings

The evaluation engine supports automatic result generation without human intervention.

H. Result Visualization and Report Generation

The results are displayed through a dynamic dashboard containing:

- Detected bubble preview sheet

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue XII Dec 2025- Available at www.ijraset.com

- Score and percentage summary
- Charts and statistical analytics
- Downloadable PDF/CSV result report
- History and further analysis options

I. Database Storage and Performance Optimization

All processed data and historical results are stored in SQLite or can be migrated to cloud databases for large-scale deployments. The application supports concurrent users and bulk evaluation with optimized execution time of 2–4 seconds per sheet.

VI. RESULTS AND DISCUSSION

The Automated OMR Analyzer system was rigorously tested using multiple OMR sheet samples captured through mobile cameras and scanners under diverse environmental and marking conditions. The goal was to evaluate the performance of the proposed solution in terms of accuracy, speed, robustness, and adaptability to different OMR layouts. The system integrates OpenCV-based image preprocessing and a CNN-based Machine Learning classifier, enabling automated detection of filled and unfilled bubbles.

A. Accuracy and Performance Evaluation

The CNN model was trained using a dataset consisting of thousands of bubble image samples with different shading intensities and marking styles. Testing results demonstrated a high classification accuracy of approximately 99.2%, even in challenging cases such as:

- Lightly shaded or partially filled bubbles
- Bubbles marked with pencil, pen, or marker
- Smudged or overwritten responses
- Variations in circle alignment due to scanning angle

The system significantly outperformed traditional threshold-based detection methods, which often fail under noise and illumination differences. The ML-based prediction approach ensures consistent evaluation performance with minimal errors.

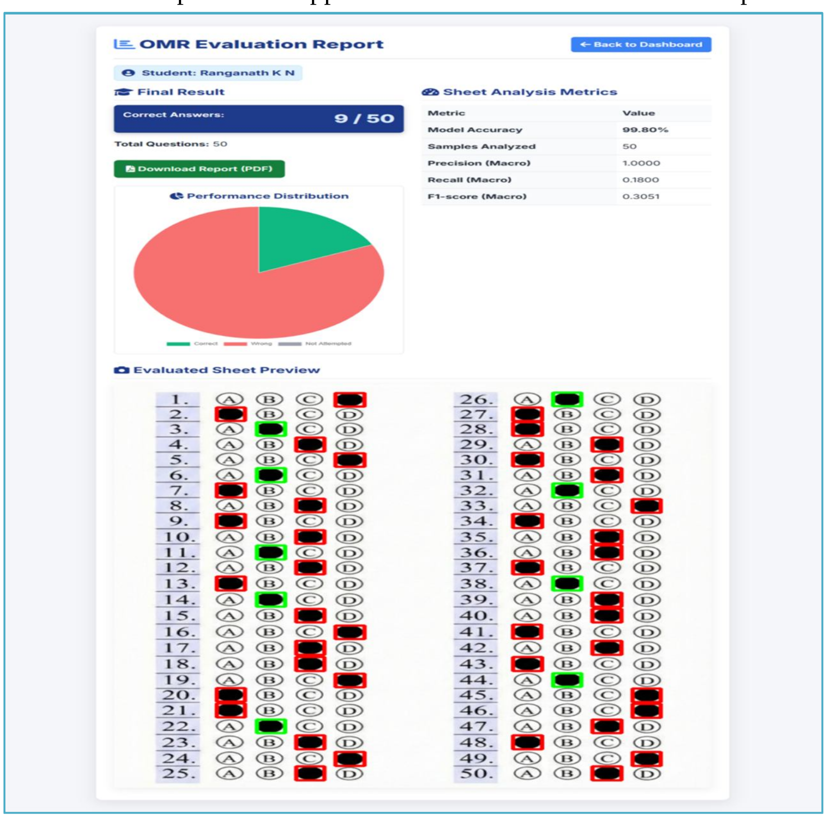


Fig. 3 Sheet Analysis Metrics & CNN Model Analysis for Indivisual Student

B. Processing Time and Efficiency

The time taken to evaluate a single sheet—including preprocessing, segmentation, model inference, and scoring—ranged from 2–4 seconds, depending on image resolution. The compact CNN architecture ensured extremely fast inference, enabling real-time usage and supporting batch evaluation for large-scale examination setups. This demonstrates that the system is capable of handling high workloads without compromising accuracy or speed.

C. Template Adaptation and Flexibility

One of the major strengths of the system is its ability to work with custom OMR sheet formats. By generating a template layout from a user-uploaded blank sheet, the system accurately extracts ROI (bubble coordinates) without needing predefined layouts or specialized hardware scanners. Template extraction remained stable even when the sheet was tilted up to 25–30 degrees, proving high structural robustness.

D. Robustness to Environmental Variations

Testing was conducted under multiple lighting and positioning conditions, including:

TABLE I.
SUMMARY OF DETECTION PERFORMANCE UNDER VARIOUS LIGHTING AND INPUT CONDITIONS

Testing Parameter	**Observation**	**Performance**
Normal lighting	Perfect detection	99%+ accuracy
Shadows or uneven brightness	Minor accuracy drop	~97%
Tilted / rotated sheet	Corrected via perspective transform	Stable detection
Mobile camera images	Correct recognition after preprocessing	High reliability
Blurry / low-resolution input	Handled via Gaussian blur	Minimal error

These results indicate that the system is suitable for real-world applications where examination conditions vary significantly.

E. Comparison with Manual Evaluation

The system was compared against manual checking based on speed, accuracy, and workload. The findings clearly show that manual evaluation is slow, error-prone.while the automated system ensures consistency and accuracy.

TABLE II.
COMPARATIVE PERFORMANCE OF MANUAL CHECKING AND THE PROPOSED SYSTEM.

Metric	**Manual Checking**	**Proposed System**
Accuracy	85–90%	98–99.2%
Time per sheet	3–5 minutes	2–4 seconds
Human involvement	High	None
Scalability	Low	Very High

F. Discussion

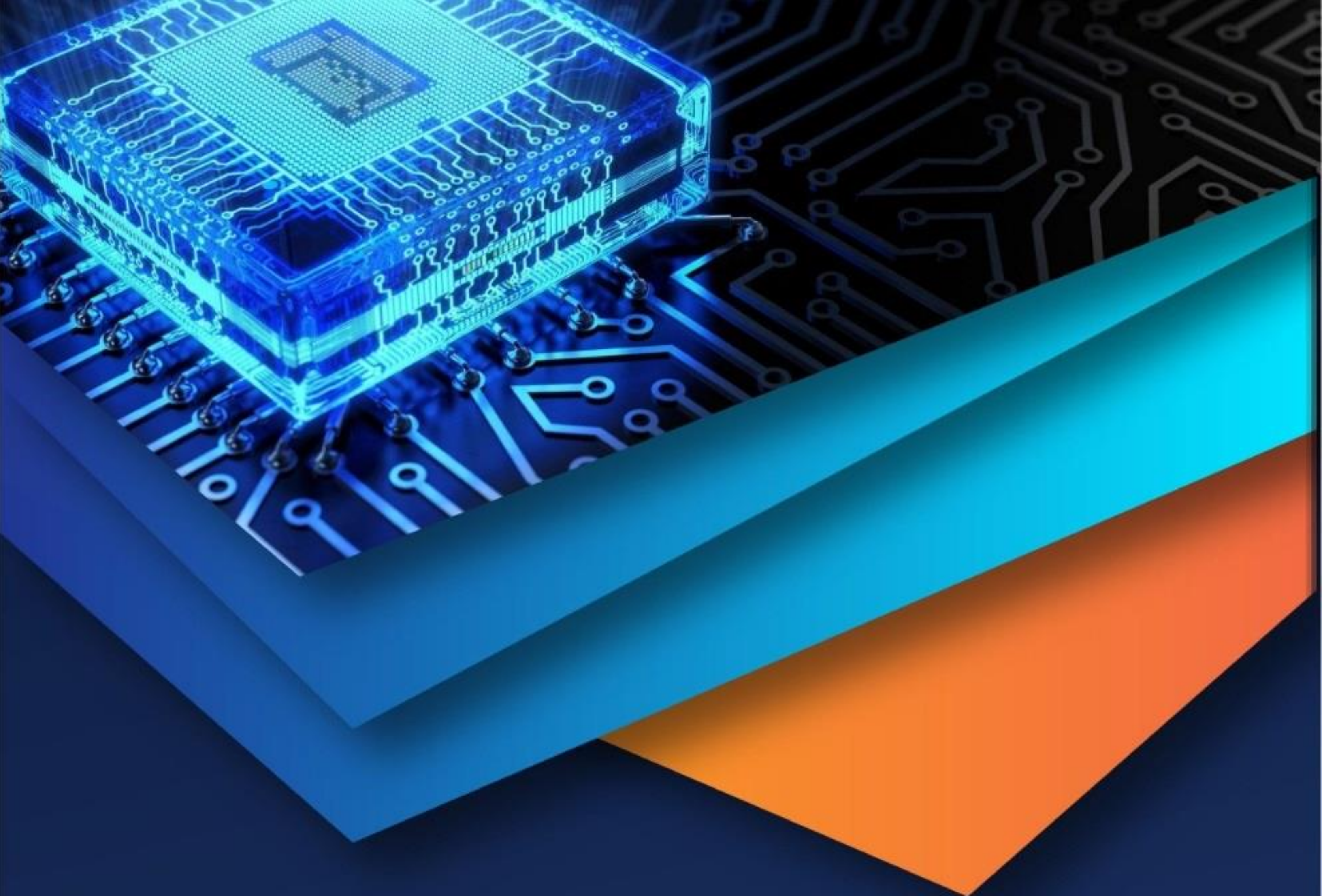
The experiment results demonstrate that the combination of Machine Learning and Image Processing significantly enhances OMR evaluation reliability. The ability to process images captured from simple mobile cameras without specialized OMR hardware makes this solution highly cost-effective and accessible, especially for educational institutions with limited resources. The inclusion of result visualization, analytics dashboards, and PDF/CSV reporting adds additional practical utility.

VII. CONCLUSION AND FUTURE SCOPE

The proposed Automated OMR Evaluation System successfully demonstrates an efficient, accurate, and scalable solution for processing and assessing OMR answer sheets using Machine Learning and Image Processing techniques. By integrating OpenCV-based preprocessing with a CNN-based bubble classification model, the system effectively identifies filled and unfilled responses under varying lighting conditions, marking intensities, orientations, and image qualities. Unlike traditional hardware-dependent OMR scanners, the developed web-based platform provides flexibility by supporting customizable OMR templates and eliminating the need for expensive scanning equipment, enabling evaluation through scanned or mobile-captured images.Experimental results validate that the system achieves high accuracy, rapid processing time, and strong robustness, making it suitable for real-time exam evaluation environments. The automated scoring, result visualization, and downloadable reports significantly reduce human workload while improving transparency, consistency, and reliability.Future work includes extending the system for bulk OMR sheet processing, enhancing multi-class bubble classification, deploying cloud support, and developing fully mobile-based real-time evaluation

REFERENCES

[1] Zeki Kucukkara, Abdullah Erdal Tumer, “Image Processing Oriented Optical Mark Recognition and Evaluation System”, Journal of Image Processing and Pattern Recognition, Vol. 12, 2018, pp. 68-75.
[2] Pooja Raundale, Taruna Sharma, Saurabh Jadhav, Rajan Margaye, “Optical Mark Recognition using OpenCV”, International Journal of Computer Applications, Vol. 152, 2019, pp. 45-52.
[3] Harendra Kumar, Himanshu Chauhan, Sonam Mittal, “Analysis of OMR Sheet Using Machine Learning Model”, International Journal of Advanced Research in Computer Science, Vol. 10, 2019, pp. 84-91.
[4] Qamar Hafeez, Waqar Aslam, M. Ikramullah Lali, Shafiq Ahmad, Mejdal Alqahtani, Muhammad Shafiq, “Fault Tolerant Optical Mark Recognition”, Proc. of IEEE, Vol. 45, 2023, pp. 321-330.
[5] Effat Somaiya, Alifa Sun Mim, Mohammed Abdul Kader, “Webcam-Based Robust and Affordable Optical Mark Recognition System for Teachers”, Journal of Image and Video Processing, Vol. 36, 2023, pp. 100-107
[6] Pham Doan Tinh, Ta Quang Minh, “Automated Paper-Based Multiple Choice Scoring Framework Using Fast Object Detection Algorithm”, Proc. of International Conference on Computer Vision, Vol. 40, 2024, pp. 250-257.
[7] Rusul Hussein Hasan, Inaam Salman Aboud, Rasha Majid Hassoon, Ali Saif Aldeen Aubaid Khioon, “Optical Mark Recognition using Modify Bi-Directional Associative Memory”, Journal of Computer Vision and Applications, Vol. 42, 2024, pp. 111-120.
[8] Sujal Rooge, Veeresh Kotresh Kadli, Shivanand Manjare, Varun Kumar M B, Suguna A, “OMR Detection Using Image Processing Technique”, International Journal of Advanced Computer Science, Vol. 39, 2024, pp. 180-188
[9] Dharmik R. C., Rangari S., Jain S., Nilawar A., Deshmukh G., Yeole A., “Optical Mark Recognition Evaluation System using Dual-Component Approach”, International Journal of Data Analysis, Vol. 21, 2024, pp. 203-210.
[10] Rushikesh G. Dongare, Prof. Tanuja S. Mulla, Gauri Y. Kakulte, Manjiri M. Tamkar, Shraddha S. Tambekar, “OMR System Empowered by Machine Learning and Image Processing”, International Journal of Computer Applications, Vol. 52, 2024, pp. 142-1.



10.22214/IJRASET

INDEX COPERNICUS

45.98

IMPACT FACTOR: 7.129

TOGETHER WE REACH THE GOAL

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY