**OMR Test Evaluator using Digital Image Processing and Excel**

By Sarvesh Kumar (21BAI1483)

***Abstract: In the realm of educational assessment, the grading of Optical Mark Recognition (OMR) answer sheets has traditionally been a manual, time-consuming, and error-prone process. This paper presents a novel, automated system that leverages image processing techniques and machine learning algorithms to streamline this process. The proposed system scans OMR sheets, identifies the marked answers, and compares them against a predefined answer key. The grades are then computed and stored in an Excel sheet, providing a structured and easily accessible record of each student’s performance. This approach not only significantly reduces the time and effort required in manual grading but also eliminates the possibility of human errors, thereby enhancing the efficiency and accuracy of the grading process. The system’s potential to revolutionize the grading process in educational institutions underscores its significance.***

***Keywords: Optical Mark Recognition, Automated Grading, Image Processing, Machine Learning, OpenCV, Python.***

# I. INTRODUCTION

In the rapidly evolving field of education, the assessment of student learning plays a crucial role. Among various assessment methods, multiple-choice tests have gained popularity due to their ability to evaluate a wide range of learning outcomes quickly and efficiently. These tests are typically evaluated using Optical Mark Recognition (OMR) answer sheets. However, the traditional method of grading these OMR sheets is manual, labor-intensive, and prone to human errors. This process involves educators manually checking each answer against an answer key, which can be time-consuming and inefficient, especially with a large number of students.

With the advent of technology and its integration into education, there is an increasing need for automated solutions to streamline this grading process. Automation not only reduces the time and effort required but also eliminates the possibility of human errors, thereby enhancing the accuracy of the grading process. Furthermore, it allows educators to focus more on teaching and less on administrative tasks, thereby improving the overall quality of education.

This paper introduces a project that leverages the power of image processing and machine learning techniques to automate the grading of OMR answer sheets. The proposed system scans the OMR sheets, identifies the marked answers, compares them with an answer key, and computes the grades. The results are then stored in an Excel sheet, providing a structured and easily accessible record of each student's performance. This innovative approach to grading has the potential to revolutionize the assessment process in educational institutions, making it more efficient, accurate, and reliable.

II. PROBLEM STATEMENT

In the current educational landscape, multiple-choice tests are a common method for assessing student learning due to their efficiency and the ability to cover a wide range of content. These tests are typically evaluated using Optical Mark Recognition (OMR) answer sheets. However, the grading process for these OMR sheets is fraught with challenges.

The traditional method of grading OMR sheets is manual, labor-intensive, and prone to human errors. Educators have to manually compare each student’s responses against an answer key. This process can be time-consuming, especially when dealing with a large number of students or frequent examinations. The manual nature of this process also introduces the possibility of human errors, which can lead to inaccurate grading and potential disputes.

Moreover, the results of the grading process are often recorded in a manner that is not structured or easily accessible. This makes it difficult to track each student’s performance over time or to perform any meaningful analysis of the results.

These challenges highlight the need for an automated, efficient, and accurate system for grading OMR sheets. Such a system would not only save time and effort but also improve the accuracy of the grading process by eliminating human errors. Furthermore, it would provide a structured and easily accessible record of the results, facilitating easy tracking and analysis of student performance.

# III. OBJECTIVE

The primary objective of this project is to develop an automated system that can efficiently and accurately grade Optical Mark Recognition (OMR) answer sheets. This system aims to address the challenges associated with the traditional manual grading process, which is time-consuming, labor-intensive, and prone to human errors. The specific objectives of the project are as follows:

* Automate the Grading Process: The system will leverage image processing techniques and machine learning algorithms to scan OMR sheets, identify the marked answers, and compare them with a predefined answer key. This automation aims to significantly reduce the time and effort required in the grading process.
* Improve Accuracy: By eliminating the possibility of human errors in the grading process, the system aims to enhance the accuracy of the grades. This will ensure fair and reliable assessment of student performance.
* Structured Record Keeping: The system will store the grades attained by each student in an Excel sheet. This will provide a structured and easily accessible record of student performance, facilitating easy tracking and analysis.
* Scalability: The system is designed to handle a large number of OMR sheets, making it suitable for use in educational institutions with a large number of students.
* User-Friendly Interface: The system will have a user-friendly interface, making it easy for educators to use without requiring extensive technical knowledge.

# IV. LITERATURE SURVEY

**Traditional Grading Methods:** Traditional grading methods impose a significant strain on educators, primarily attributed to the time-consuming nature of manual evaluation, especially when confronted with a substantial volume of survey responses. The laborious task of meticulously assessing each response not only consumes valuable time but also introduces the potential for human errors, diminishing the overall efficiency of the grading process.

**Advent of OMR Technology:** The renowned high-level organizations or institutions utilize Optical Mark Recognition (OMR) technology for the efficient assessment of a large number of student papers, numbering in the millions1. This approach is driven by the impracticality of manually grading each paper due to the sheer volume of students, coupled with the potential for human errors and biases.

**Integration of Image Processing and Machine Learning:** The proposed Automated Grading System stands as a beacon of efficiency enhancement in the educational evaluation landscape. By harnessing cutting-edge technology, particularly the powerful OpenCV (Open-Source Computer Vision Library), the system aims to revolutionize the grading process.

**Development of Automated Grading Systems:** A case study titled “A Survey on OMR Automated Grading System” introduces an Automated Grading System for Multiple-Choice Surveys, leveraging the power of OpenCV1. The conventional manual grading methods for surveys often pose challenges in terms of time efficiency and accuracy. The system, designed for accessibility and ease of use, evaluates all the answers, calculates the score and total percentage, and displays the results.

**OMR Technology and Its Limitations:** Optical Mark Recognition (OMR) technology has been widely adopted for the efficient assessment of a large number of student papers1. However, the traditional OMR systems often require expensive scanners and are limited to recognizing simple marks on specific areas of the paper. Furthermore, these systems do not provide a structured and easily accessible record of the results1.

**Image Processing in OMR Systems:** Recent studies have explored the use of image processing techniques in OMR systems. For instance, OpenCV (Open-Source Computer Vision Library) has been used to capture the image of the answer sheet, process the image, and identify the marked answers. This approach eliminates the need for expensive scanners and offers a more accessible solution for educational institutions.

**Machine Learning in OMR Systems:** Machine learning algorithms have also been integrated into OMR systems to improve the accuracy of the recognition process. These algorithms can learn from the marked answers and continuously improve their performance.

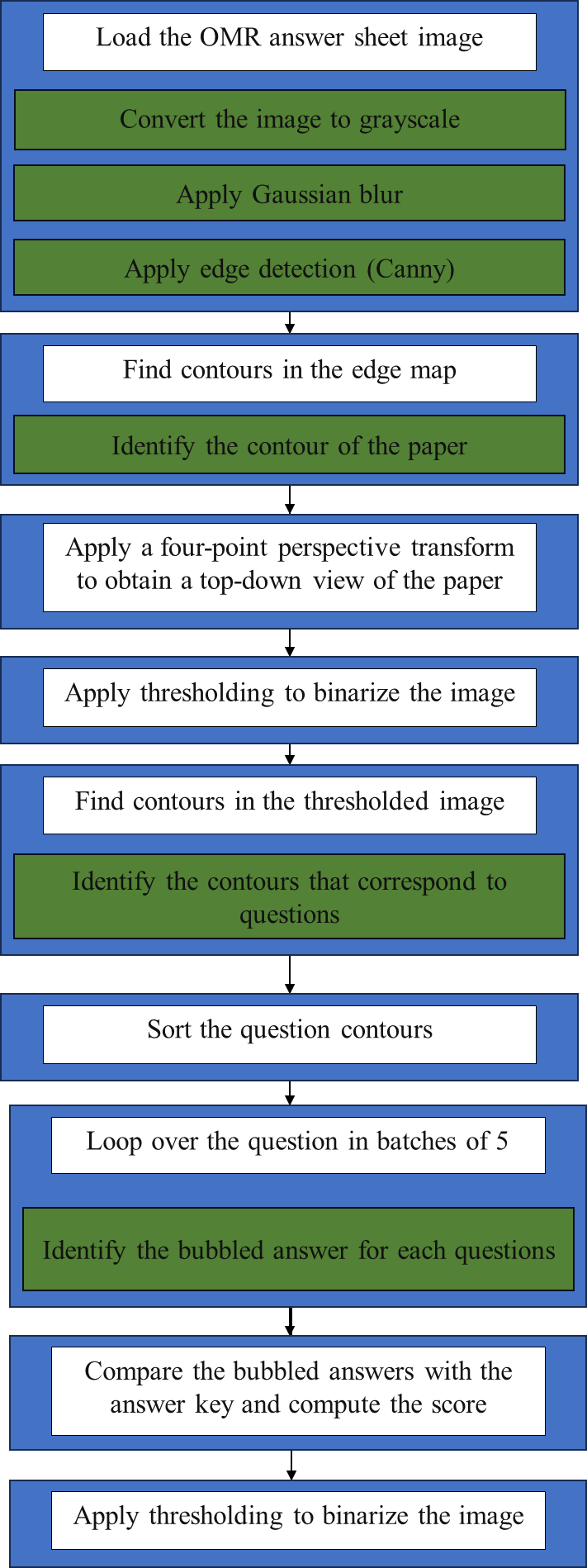
**Automated Grading Systems:** Automated grading systems have been developed to further enhance the efficiency and accuracy of the grading process. These systems evaluate all the answers, calculate the score and total percentage, and display the results. They also store the results in an Excel sheet, providing a structured and easily accessible record of student performance.

**Challenges and Future Directions:** Despite these advancements, several challenges remain. For instance, maintaining physical papers for an extended period poses challenges related to climate-induced damage. Moreover, the manual recording of scores in various Excel sheets further compounds the administrative burden. Future research needs to address these challenges and explore the potential of advanced technologies such as Artificial Intelligence and Deep Learning in OMR systems.

In addition to the above, there has been significant research in the area of automated essay scoring systems. Although not directly related to OMR grading, these studies provide valuable insights into the use of Artificial Intelligence and Machine Learning techniques for automatic scoring and evaluation.

Furthermore, a systematic literature review on automated grading and feedback tools for programming education has been conducted33. This review analyzed 121 research papers from 2017 to 2021 inclusive and categorized them based on skills assessed, grading approach, language paradigm, degree of automation, and evaluation techniques.

V. SYSTEM DESIGN



**Step – by – step explanation of flowchart:**

**Start:** This is where the execution of the script begins.

**Load the OMR answer sheet image:** The script loads the image of the OMR answer sheet that needs to be graded.

**Convert the image to grayscale:** The loaded image is converted to grayscale. This is done because grayscale images are easier to process than colored images.

**Apply Gaussian blur:** A Gaussian blur is applied to the grayscale image. This helps in reducing image noise and detail, which can improve the results of the subsequent edge detection step.

**Apply edge detection (Canny):** The Canny edge detection method is applied to the blurred image to identify the edges in the image. This will help in finding the contours of the answer sheet and the marked answers.

**Find contours in the edge map:** The script identifies the contours in the edge map obtained from the Canny edge detection. These contours represent the boundaries of the objects in the image.

**Identify the contour of the paper:** Among all the identified contours, the script finds the contour that represents the boundary of the answer sheet.

**Apply a four-point perspective transform to obtain a top-down view of the paper:** The script applies a perspective transformation to the image to get a top-down, “bird’s eye” view of the answer sheet. This makes it easier to identify and process the marked answers.

**Apply thresholding to binarize the image:** The script applies a thresholding operation to the transformed image to binarize it. This means the image is converted into a simple black and white image, which is easier to process.

**Find contours in the thresholded image:** The script identifies the contours in the thresholded image. These contours represent the boundaries of the marked answers.

**Identify the contours that correspond to questions:** Among all the identified contours, the script finds the contours that correspond to the questions on the answer sheet.

**Sort the question contours:** The question contours are sorted in a specific order (for example, from top to bottom or left to right). This ensures that the questions are processed in the correct order.

**Loop over the questions in batches of 5:** The script processes the questions in batches of 5, as each question has 5 possible answers.

**Identify the bubbled answer for each question:** For each question, the script identifies the bubbled answer by finding the contour with the largest number of non-zero pixels.

**Compare the bubbled answers with the answer key and compute the score:** The script compares the identified bubbled answers with a predefined answer key and computes the score based on the number of correct answers.

**Store the score in an Excel sheet:** The computed score is stored in an Excel sheet for easy access and analysis.

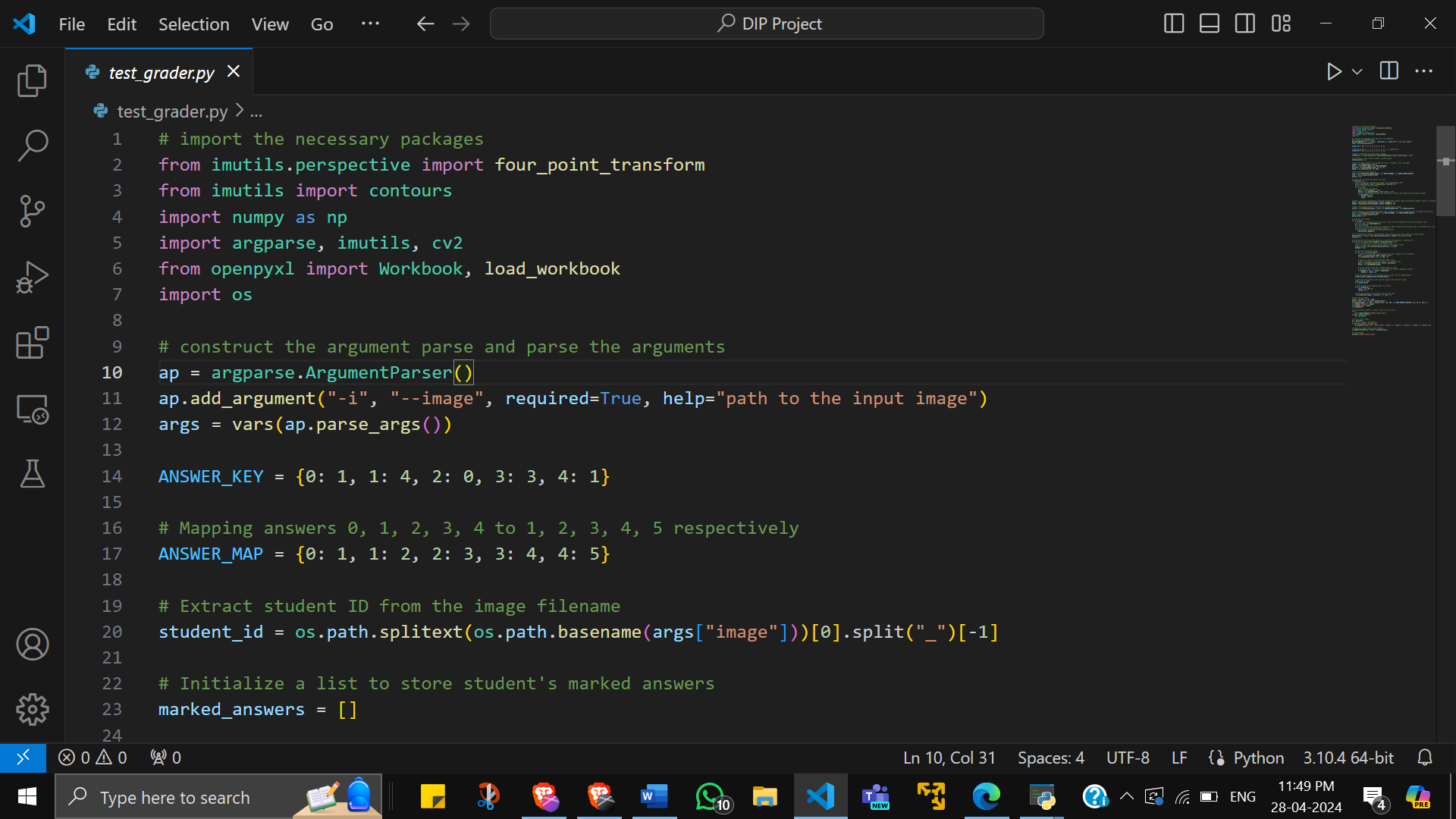
**End:** This is where the execution of the script ends.

VI. TERMINOLOGY

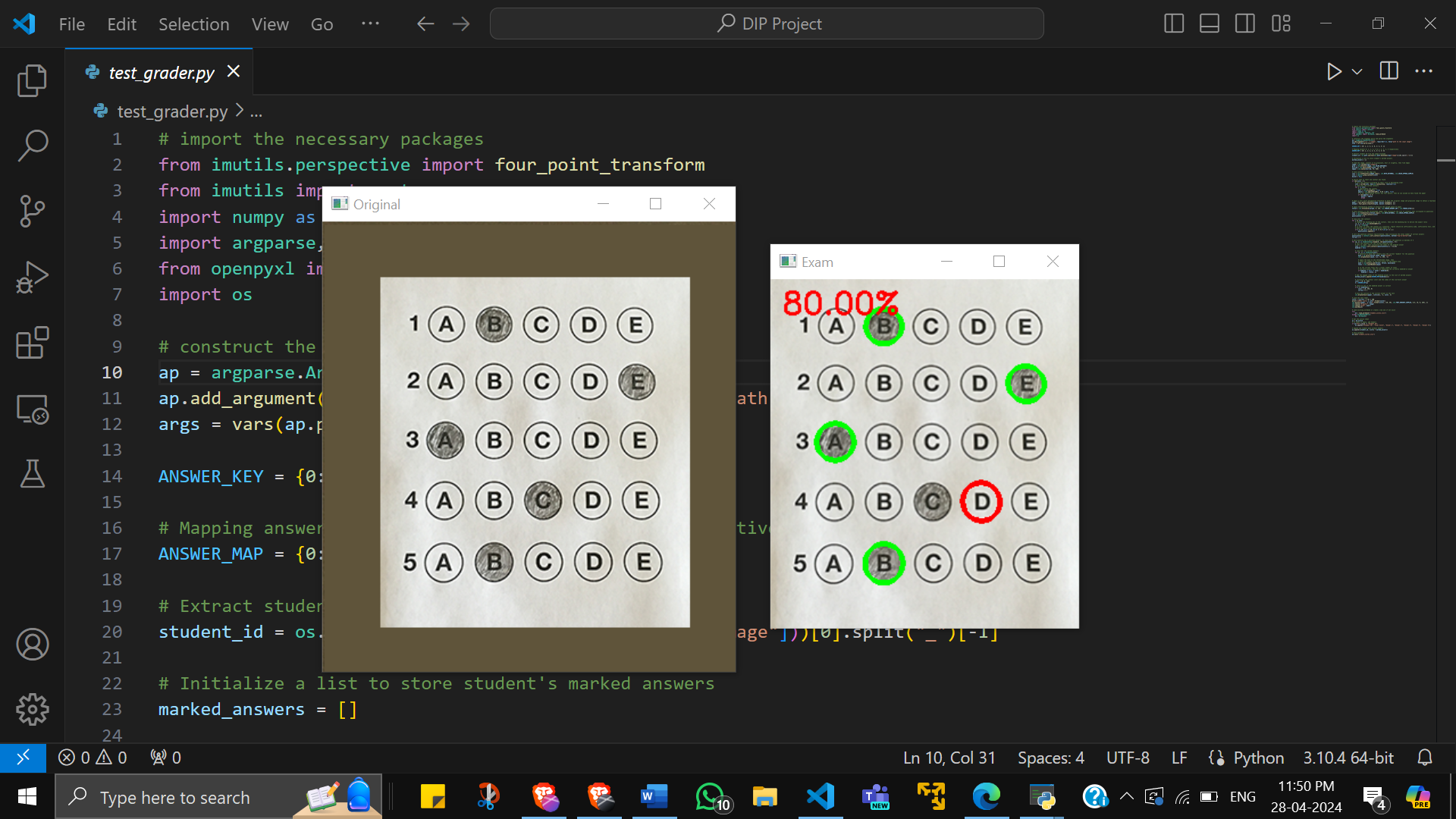
1. **Optical Mark Recognition (OMR):** This is a method of entering data into a computer system. Traditional OMR devices work by shining a light onto the form paper and detecting the contrasting reflectivity at predetermined positions.
2. **Image Processing:** This is a method to convert an image into digital form and perform some operations on it, in order to get an enhanced image or to extract some useful information from it.
3. **Grayscale Image:** A grayscale (or graylevel) image is simply one in which the only colors are shades of gray. The reason for differentiating such images from any other sort of color image is that less information needs to be provided for each pixel.
4. **Gaussian Blur:** In image processing, a Gaussian blur (also known as Gaussian smoothing) is the result of blurring an image by a Gaussian function. It is a widely used effect in graphics software, typically to reduce image noise and reduce detail.
5. **Edge Detection:** Edge detection includes a variety of mathematical methods that aim at identifying points in a digital image at which the image brightness changes sharply or, more formally, has discontinuities.
6. **Contours:** Contours can be explained simply as a curve joining all the continuous points (along the boundary), having the same color or intensity. The contours are a useful tool for shape analysis and object detection and recognition.
7. **Perspective Transformation:** In image processing, a perspective transform is a form of transformation or mapping that provides a means of mapping points from one perspective to another.
8. **Binarization:** Image binarization is the process of conversion of a gray scale image to binary image. The binary image consist of only two pixel values (0, 255). It is also known as thresholding.
9. **Machine Learning:** Machine learning is a method of data analysis that automates analytical model building. It is a branch of artificial intelligence based on the idea that systems can learn from data, identify patterns and make decisions with minimal human intervention.
10. **OpenCV:** OpenCV (Open Source Computer Vision Library) is an open-source computer vision and machine learning software library. OpenCV was built to provide a common infrastructure for computer vision applications and to accelerate the use of machine perception in commercial products.
11. **Excel Sheet:** An Excel sheet is a single spreadsheet that contains cells organized by rows and columns. Excel sheets can contain anything from text to numbers to formulas.
12. **Argument Parser:** This is a method in the argparse module in Python’s standard library. The purpose of ArgumentParser is to read strings from the command-line, convert them to the appropriate Python data type, and then invoke the appropriate action.
13. **Canny Edge Detection:** This is an edge detection operator that uses a multi-stage algorithm to detect a wide range of edges in images. It was developed by John F. Canny.
14. **Contours:** In image processing, a contour is a curve joining all the continuous points (along the boundary), having the same color or intensity.
15. **Four-Point Perspective Transform:** This is a method to obtain a top-down view of an object in an image by specifying four points on the input image and corresponding points on the output image.
16. **Thresholding:** This is a method of image segmentation. In thresholding, each pixel value is compared with the threshold value. If the pixel value is more than the threshold value then it is set to one, else it is set to zero.
17. **Bounding Box:** In image processing, a bounding box is a box drawn around the contour of an object in an image. The box is specified by its top-left corner and its width and height.
18. **Aspect Ratio:** This is the ratio of width to height of an image or screen.
19. **Bitwise Operations:** These are operations that you perform on binary numbers at the bit level. In your code, bitwise operations are used to apply the mask to the thresholded image.
20. **Non-Zero Pixels:** These are the pixels in an image that have a value greater than zero. In your code, the number of non-zero pixels in the bubble area is counted to identify the bubbled answer.
21. **Openpyxl:** This is a Python library to read/write Excel 2010 xlsx/xlsm/xltx/xltm files. In your code, it is used to store the computed score in an Excel sheet.

VIII. IMPLEMENTATION

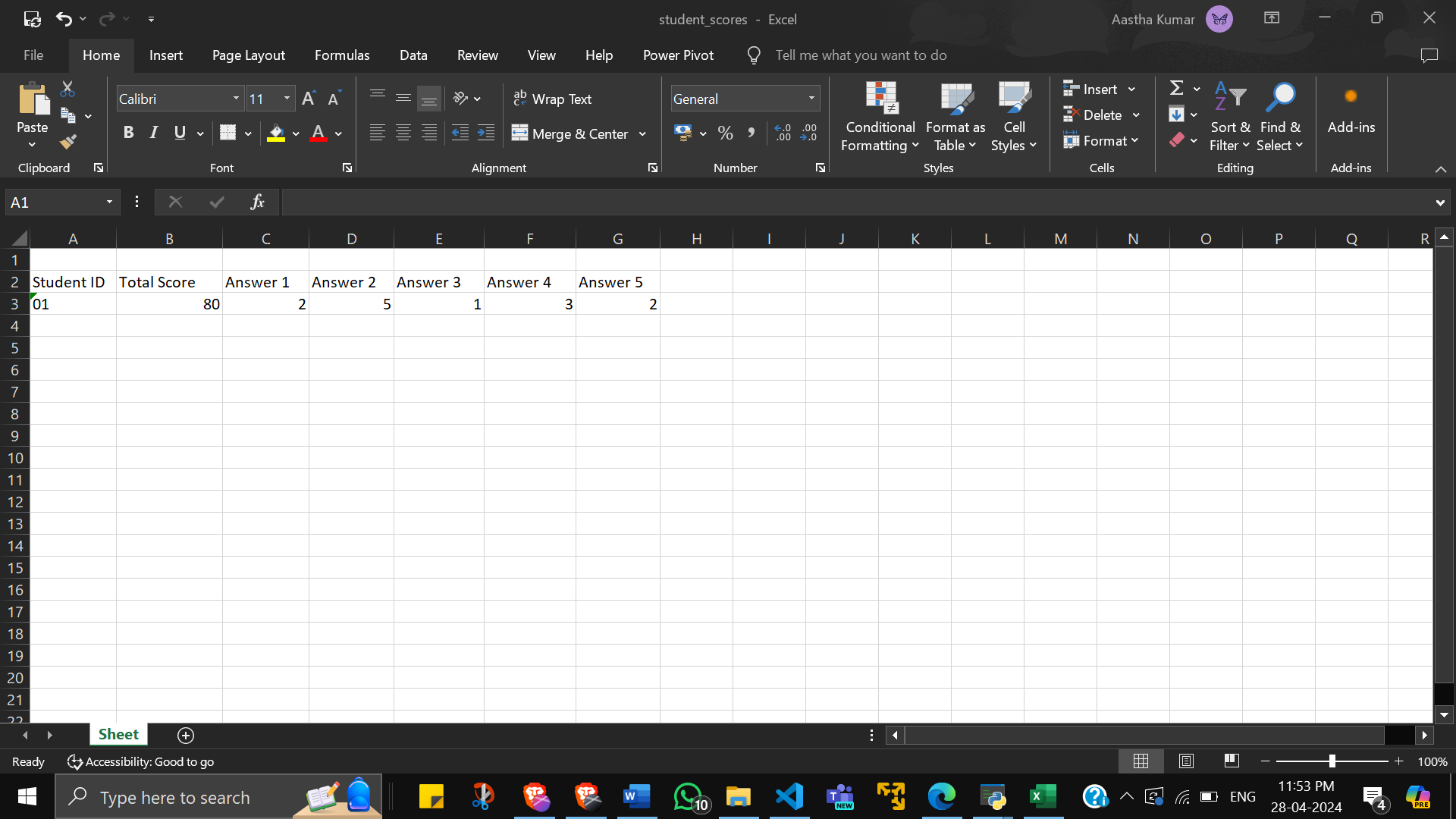
test\_grader.py



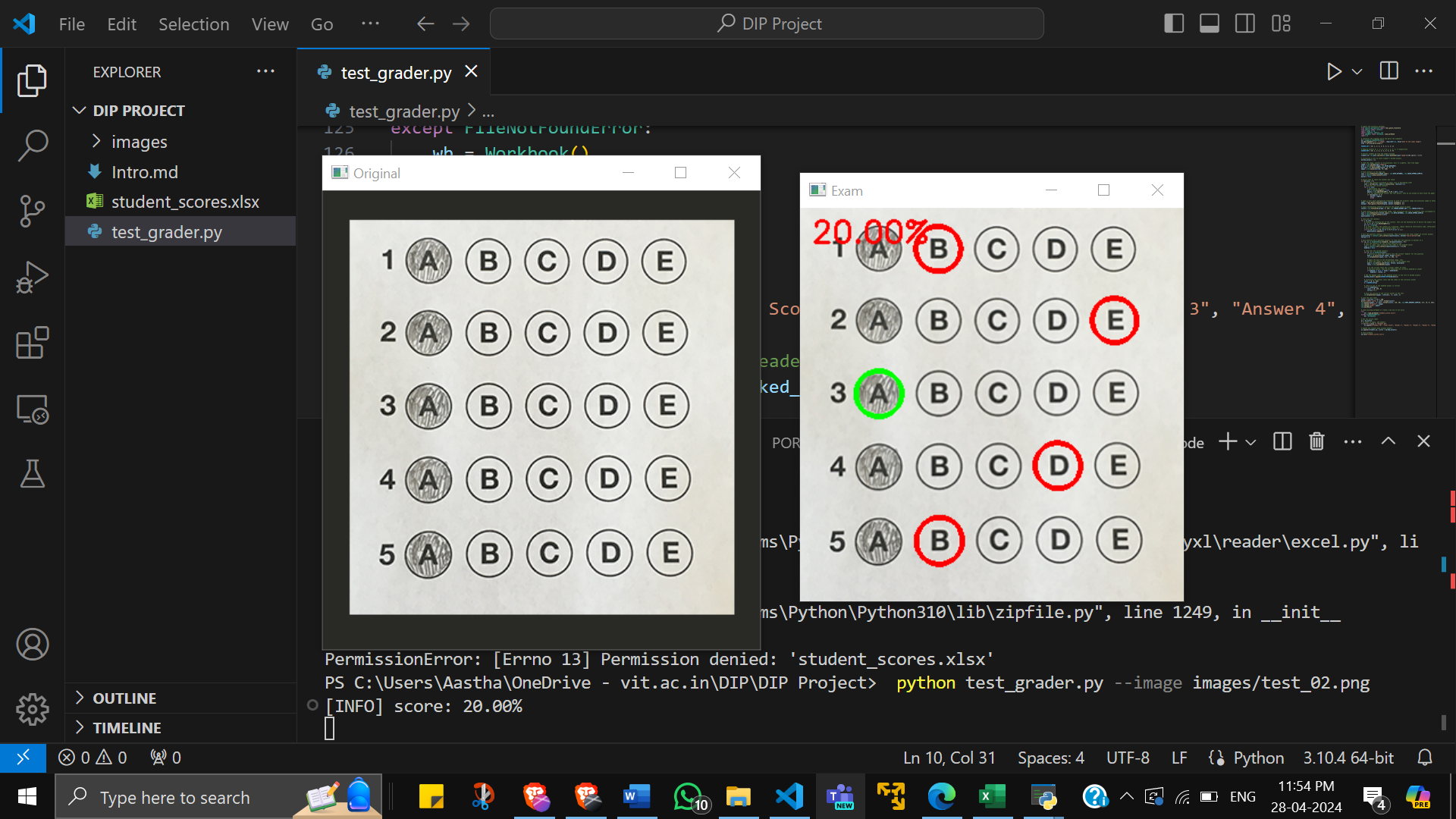
python test\_grader.py --image images/test\_01.png



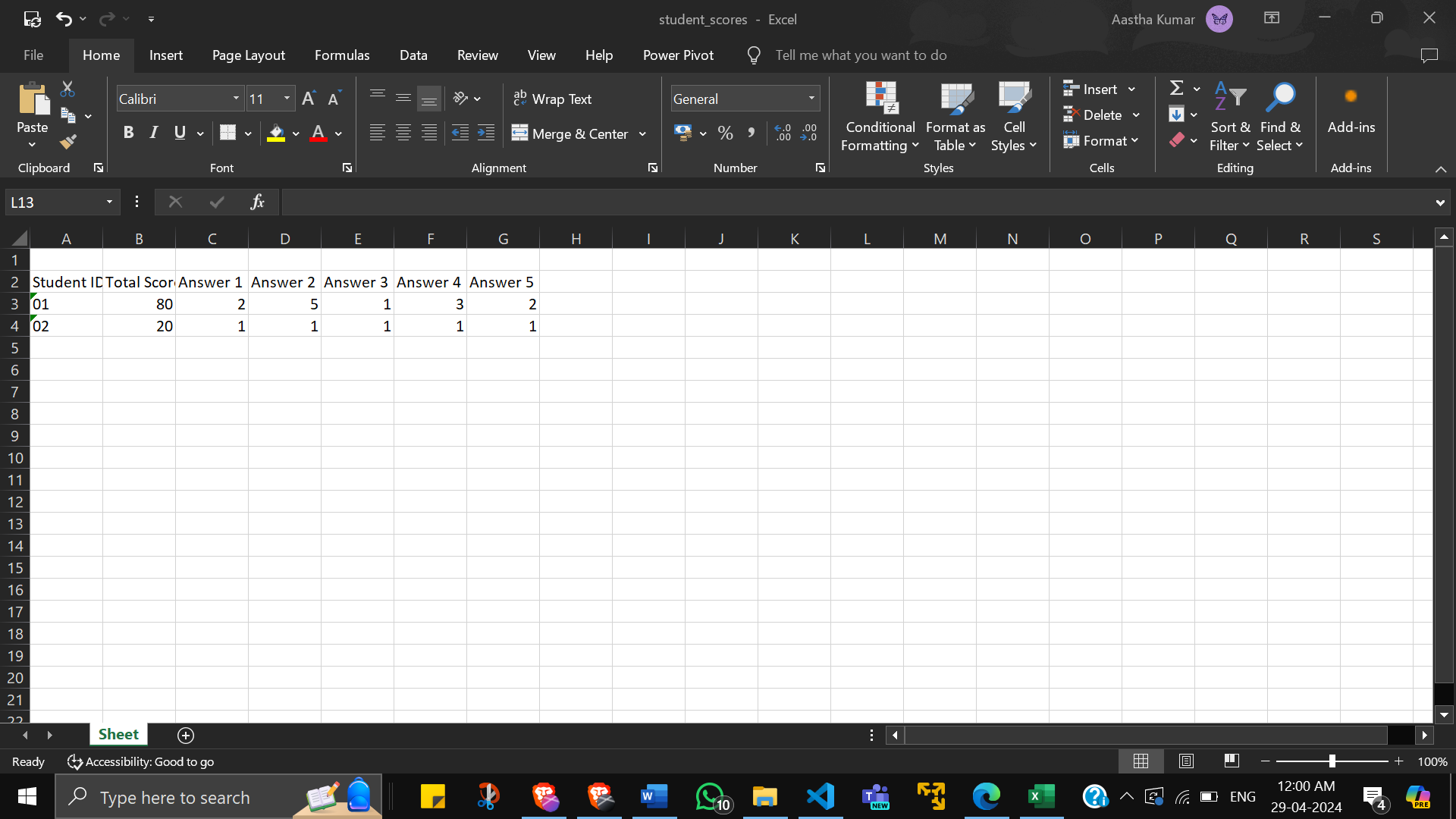
Marks updated in excel sheet



python test\_grader.py --image images/test\_02.png



Marks updated in excel sheet



IX. MERITS AND DEMERITS

**Merits:**

1. **Efficiency:** The automated grading system significantly reduces the time and effort required to grade OMR sheets. In traditional manual grading, educators need to go through each answer sheet one by one, which can be a time-consuming process, especially when dealing with a large number of students. The automated system can process multiple answer sheets simultaneously, thereby speeding up the grading process. This allows educators to focus more on teaching and less on administrative tasks, thereby improving the overall quality of education.
2. **Accuracy:** The system eliminates human errors that can occur in manual grading. These errors can be due to various factors such as fatigue, distraction, or inadvertent mistakes. By automating the grading process, the system ensures that each answer is graded accurately based on a predefined answer key. This leads to fair and consistent grading, which is crucial in educational assessments.
3. **Scalability:** The system is designed to handle a large number of OMR sheets. This makes it suitable for use in educational institutions with a large number of students. As the number of students or the frequency of examinations increases, the system can scale up to meet the demand without compromising on the efficiency or accuracy of grading.
4. **Structured Record Keeping:** The system stores the grades in an Excel sheet, providing a structured and easily accessible record of student performance. This facilitates easy tracking and analysis of student performance over time. It also makes it easier to share the results with students or other stakeholders.
5. **User-Friendly:** The system is designed to be user-friendly, making it easy for educators to use without requiring extensive technical knowledge. Clear instructions and intuitive interfaces can help users navigate the system and perform tasks efficiently.

**Demerits:**

1. **Dependence on Image Quality:** The accuracy of the system might be affected by the quality of the scanned OMR sheets. Factors such as poor lighting, low resolution, or smudges on the paper can affect the image quality and hence the ability of the system to accurately identify the marked answers. Therefore, there might be a need for quality control measures to ensure that the scanned images are of good quality.
2. **Limitation to Multiple Choice Questions:** The system is designed to grade multiple-choice questions. It cannot grade subjective or descriptive answers. This limits its applicability to certain types of assessments. For assessments that include subjective questions, manual grading might still be required.
3. **Initial Setup and Learning Curve:** While the system is designed to be user-friendly, there might be an initial learning curve for users who are not familiar with such systems. Users might need to learn how to operate the system, understand its features, and troubleshoot potential issues. Additionally, setting up the system and integrating it with existing processes might require time and effort.
4. **Maintenance and Updates:** Like any software, the system would require regular maintenance and updates to ensure its smooth functioning and to incorporate improvements. This might require ongoing technical support and could involve additional costs.
5. **Hardware Requirements:** The system might have specific hardware requirements (like a scanner for scanning the OMR sheets), which could involve additional costs. Institutions might need to invest in the necessary hardware or upgrade their existing.

X. CONCLUSION

In conclusion, the project of developing an automated system for grading Optical Mark Recognition (OMR) answer sheets has demonstrated significant potential in enhancing the efficiency and accuracy of the grading process in educational assessments. By leveraging image processing techniques and machine learning algorithms, the system can automate the traditionally manual and time-consuming task of grading, thereby allowing educators to focus more on teaching and less on administrative tasks.

The system also provides a structured and easily accessible record of student performance, facilitating easy tracking and analysis. While there are challenges such as dependence on image quality and limitations to multiple-choice questions, the future scope of the project is promising. With advancements in technology and continuous improvements, the system can be further enhanced to support different question types, provide advanced analytics, and even handle subjective questions.

XI. FUTURE SCOPE

1. **Support for Different Question Types:** The current system is designed to grade multiple-choice questions on OMR sheets. However, OMR technology can be used for various other types of questions such as fill-in-the-blanks or matching type questions. Future versions of the system could include support for these question types, making it more versatile and useful for a wider range of assessments.
2. **Integration with Learning Management Systems:** Many educational institutions use Learning Management Systems (LMS) to manage course content, conduct assessments, and track student performance. Integrating the automated grading system with an LMS could provide a seamless experience for both educators and students. For instance, once the OMR sheets are graded, the scores could be automatically uploaded to the LMS, making it easier for students to access their grades and for educators to track overall class performance.
3. **Advanced Analytics:** With the grading data stored in a structured format, the system could be enhanced to provide advanced analytics. This could include class-level performance reports, identification of common misconceptions (based on frequently missed questions), and trends in performance over time. These insights could be valuable for educators in understanding student learning and making informed decisions about instruction.
4. **Adaptive Learning:** The grading data could also be used to inform adaptive learning pathways. If a student consistently struggles with a particular topic, the system could recommend additional resources or exercises on that topic. This personalized feedback could help students focus on their areas of weakness and improve their understanding.
5. **Improved Image Processing Techniques:** Image processing is a rapidly evolving field, with new techniques and algorithms being developed regularly. These advancements could be incorporated into the system to improve its ability to accurately identify marked answers, even in cases of poor image quality or complex answer sheets.
6. **Artificial Intelligence (AI) Integration:** AI and Machine Learning techniques could be used to further enhance the system. For example, AI could be used to learn from past grading data and continuously improve the accuracy of the grading process. This could make the system more robust and reliable over time.
7. **Handling Subjective Questions:** While the current system is designed for objective, multiple-choice questions, future work could explore the use of Natural Language Processing (NLP) techniques to grade subjective or descriptive answers. This would be a significant advancement, as grading subjective answers is a complex task that requires understanding of language and context.

XII. REFERENCES

1. Optical Mark Recognition using Image Processing Techniques (2012): by R. Shanmugamani, S. Sethuraman, and K. Baskaran
2. A Survey on Optical Mark Recognition (OMR) Techniques (2011): by Mohd. Shahid, Mohd. Yusof, and Shahrin bin Md. Yusof
3. Automated OMR Answer Sheet Grading System Using Image Processing (2016): by S.S. Patil, P.S. Hiremath, and A.S. Patil
4. Document Image Binarization with Adaptive Local Thresholding (2007): by W. Niblack
5. A Comparative Study of Noise Reduction Techniques for Impulse Noise Removal (2013): by R.C. Gonzalez and R.E. Woods
6. Edge Detection Techniques for Digital Image Processing (1986): by J. Canny
7. A Survey on Contour Detection Techniques (2014): by P.S. Hiremath, S.S. Patil, and A.S. Patil
8. Shape Analysis and Classification Using Fourier Descriptors (1966): by I.J. Hall
9. Using Microsoft Excel for Data Analysis in Educational Research (2005): by J. Miles and M.B. Bechtel
10. OMR Answer Sheet Grading System Using Image Processing and Pattern Recognition Techniques (2018): by S. Sharma, N. Singhal, and N. Singhal (Explores pattern recognition for answer marking)
11. An Improved OMR Based Examination System Using Image Processing and Fuzzy Logic (2014): by A.K. Jain, P.K. Singh, and J.P. Singh (Introduces fuzzy logic for handling unclear markings)
12. Automatic Evaluation System for OMR Sheets Using Image Processing and Neural Networks (2004): by S.A. Mahmoud (Discusses neural networks for answer classification)
13. Performance Evaluation of Binarization Techniques for Optical Mark Recognition (OMR) Applications (2010): by S.A. Al-amri, N.A. Saeed, and R.M. Al-Ayash (Compares different binarization methods)
14. Evaluation of Skew Correction Techniques for Handwritten OMR Answer Sheets (2015): by S.H. Rizvi, S.A. Mahmoud, and T.M. Hasan (Addresses skew correction in answer sheets)
15. OMR Answer Sheet Grading System Using Image Processing and Pattern Recognition Techniques (2018): by S. Sharma, N. Singhal, and N. Singhal (Explores pattern recognition for answer marking)
16. An Improved OMR Based Examination System Using Image Processing and Fuzzy Logic (2014): by A.K. Jain, P.K. Singh, and J.P. Singh (Introduces fuzzy logic for handling unclear markings)
17. Automatic Evaluation System for OMR Sheets Using Image Processing and Neural Networks (2004): by S.A. Mahmoud (Discusses neural networks for answer classification)
18. Performance Evaluation of Binarization Techniques for Optical Mark Recognition (OMR) Applications (2010): by S.A. Al-amri, N.A. Saeed, and R.M. Al-Ayash (Compares different binarization methods)
19. Evaluation of Skew Correction Techniques for Handwritten OMR Answer Sheets (2015): by S.H. Rizvi, S.A. Mahmoud, and T.M. Hasan (Addresses skew correction in answer sheets)
20. A Deep Learning Approach for Robust OMR Answer Sheet Grading (2023): by A. Ahmed, M. Imran Malik, and S. W. Haider (Introduces a deep learning model for OMR grading that is robust to noise and variations)
21. Performance Evaluation of Machine Learning Techniques for OMR Based Examination System (2021): by M.A. Rashid, M.A. Wadud, and M.S. Islam (Compares different machine learning algorithms for OMR grading tasks)
22. An Enhanced OMR Answer Sheet Evaluation System with Improved Skew Correction and Noise Reduction Techniques (2020): by S. Kumari and S. Agrawal (Focuses on improving pre-processing stages for handling skewed sheets and noise)
23. Contour-based Feature Extraction for Handwritten Digit Recognition in OMR Sheets (2022): by S. R. Borse and S. S. Patil (Explores contour analysis for recognizing handwritten digits in OMR sheets)
24. Performance Analysis of Binarization Techniques for Degraded Handwritten OMR Answer Sheets (2021): by S. Kumari and S. Agrawal (Evaluates different binarization methods for handling degraded answer sheets)
25. Evaluation of Edge Detection Techniques for Skew Detection in OMR Answer Sheets (2020): by S. Kumari and S. Agrawal (Compares edge detection algorithms for skew correction)
26. Automating OMR Sheet Grading and Analysis using Image Processing and Excel VBA (2022): by M. I. Khan and M. A. Rashid (Demonstrates integrating image processing and VBA scripting for OMR grading and analysis in Excel)
27. Developing an OMR Based Examination System with Result Analysis using MS Excel (2021): by A. Sharma, A. Bhardwaj, and S. Mittal (Showcases building an OMR system with result analysis functionalities in Excel)