

Desktop Application for Automating Data Entry of Students Marks

¹V. Prithivi Raj, ²D.Nagendra Kumar

¹UG Student, Department of CSE, Thiagarajar College of Engineering, Madurai, India

²Assistant Professor, Department of CSE, Thiagarajar College of Engineering, Madurai, India

¹prithiviclge@gmail.com, ²d.nagendra11@gmail.com

Abstract—The impact of technology on the world, particularly in education is increasing tremendously. The population of students in schools, colleges, and other educational institutions has been increasing over the years, which has led to an increase in the number of hands filled sheets that teachers and other working staffs must record the academic progress of the students. Schools use physical copies of answer sheets, which can be time-consuming for teachers to grade, especially if there are many students. This is an application that streamlines the process of entering marks from physical papers to websites by leveraging automation. This application is designed to cater to the needs of organizations that handle large volumes of paperwork, such as educational institutions and recruitment agencies. It fulfills this need by automating the process of entering marks from physical papers to websites, which not only saves time but also reduces the risk of errors. The application's ability to crop the required portion and convert it to text ensures that data is accurately captured and entered the target website by saving data in excel. The GUI desktop application was built using Python's Tkinter library. The application also utilizes PIL and OpenCV (Computer Vision) for image processing and cropping. As for the detection of handwritten text detection from physical answer sheets some of the APIs like Google cloud vision API, AspriseOCR (Optical Character Detection) API etc. was used. Additionally, Openpyxl was utilized for handling Excel and storing data. This application is embedded with various python library and python package and coded with python programming language. In summary, this application caters to the real-time need for digital tools that can simplify tedious tasks and improve productivity in today's world.

Keywords – Physical paper, image processing, handwritten text detection, Tkinter, automating, Open CV, Asprise OCR API.

I. INTRODUCTION

In our day-to-day life Education system plays a crucial role in shaping the future of individuals and society. Educational institutions,

particularly schools, are responsible for imparting knowledge and skills to students and preparing them for their future careers. Improving the school education system can have significant benefits for individuals and society. It can improve academic performance, promote social mobility, and prepare students for their future careers. Educational institutions should strive to provide high-quality education to all students and continuously work towards improving the education system. Storing data in educational institutions is an essential practice for managing and maintaining student and institutional information. Storing data in educational institutions is crucial for managing institutional operations, improving student outcomes, and meeting regulatory requirements.

Storing data in physical papers has been a traditional way of record-keeping and archiving information. Despite the advent of digital storage systems, there are still some advantages to storing data in physical papers. The generation of hand-filled forms for students [1] refers to the process of creating physical paper forms that students can fill out by hand. These forms can be used for various purposes, such as course registration, exam registration, and student information updates. The generation of hand-filled forms for students is still widely used in many educational institutions, despite the increasing use of digital forms and online registration systems. This is because physical forms can provide a tangible record of the student's information and can be used as a backup in case of technical issues or data loss [2]. Additionally, some students may prefer to fill out physical forms rather than digital ones, particularly if they are more comfortable with handwriting.

Text detection from handfilled forms can be useful for digitizing and archiving physical documents [3], as well as for automating data entry and analysis. This can save time and resources for educational institutions and other organizations that deal with a large volume of forms and documents. Text detection from handfilled forms [4] involves the



use of image processing and optical character recognition (OCR) technology to extract the text information filled in by the user. Text detection from hand-filled forms [1][2] can be challenging due to the variety of handwriting styles and the potential for errors. However, there are several techniques that can be used to detect and extract text from hand-filled forms.

One approach is to use image processing techniques [5][4], such as binarization and morphological operations, to enhance the contrast and clarity of the form. This can help make the text more distinguishable from the background and other markings on the form.

Another approach is to use machine learning algorithms [6], such as convolutional neural networks (CNNs), to detect and recognize the text. CNNs can be trained on a large dataset of hand-filled forms to learn patterns and features of handwriting styles, which can then be used to accurately recognize and extract text from new forms [12].

In addition, pre-processing techniques such as skew correction and noise reduction can also help improve the accuracy of text detection from hand-filled forms. Skew correction can straighten out the form to make the text lines more horizontal and smudges that may interfere with text recognition. Overall, text detection from hand-filled forms requires image processing techniques to accurately recognize and extract the text. With the advancements in OCR technology, text detection from hand-filled forms [9] has become more efficient and accurate, reducing the need for manual data entry, improving the speed and accuracy of data processing.

OCR (Optical Character Recognition) technology [7][8] is used in image processing to recognize and extract text from images. OCR text has various applications in image processing. One of the primary uses of OCR text is in document digitization. OCR text can convert paper documents into digital format, making it easier to archive, store, retrieve, and search for information.

OCR technology can be used to create accessible versions of documents for individuals with visual impairments. OCR [7] converts printed materials into electronic text that can be read by screen readers, improving accessibility.

OCR can also be used to recognize and extract text from images, such as scanned documents or photographs. This is particularly useful for identifying product names and serial numbers,

recognizing handwriting, and identifying text in license plates. OCR text can also be used for translation purposes [8]. OCR can extract text from images in one language and translate it into another, making it easier to translate signs or labels in a foreign language.

Finally, OCR can be used to extract data from forms such as invoices or receipts. This helps automate data entry [10] tasks, reduces errors, and increases efficiency. Overall, OCR text is a valuable tool in image processing that can help automate tasks [9][11], improve accessibility, and extract valuable information from images.

II. RELATED WORKS

A. Collecting Students Answer sheets

Hand-filled forms are a versatile tool used for collecting information from individuals in a variety of contexts. One common application of hand-filled forms is in the collection of survey data [1]. Surveys can be used to gather information from a large group of people about their opinions, preferences, or experiences. Hand-filled forms can be used to collect survey responses, which can then be analyzed to draw conclusions or inform decision-making. When we consider Hand-filled forms students answer papers which are corrected by teacher comes under hand filled forms which are also commonly used for grading students' marks, progress in all educational institutions like schools, colleges etc. [12] Applicants can fill out the necessary information on the form, providing the necessary details for review by the employer or institution.

In healthcare settings, hand-filled forms are used for patient intake and registration, medical history, and consent forms [2]. These forms are used to collect essential information about the patient, such as medical conditions, allergies, and contact information. This information is used to provide appropriate care to the patient and ensure that healthcare providers have access to all relevant information. Overall, hand-filled forms are an essential tool for collecting information in a variety of settings [1]. With advancements in OCR technology [7][8], the data collected on these forms can be processed more efficiently and accurately, reducing the need for manual data entry and improving the speed and accuracy of data processing.

B. Storing Students Answer sheets.

Storing students' answer sheets is an important part of the educational process, as it serves



as a record of the student's performance and progress. The storage of these answer sheets can be done either in physical or digital form. Physical storage involves keeping the answer sheets in a secure location, such as a file cabinet or locked room. The answer sheets should be organized in a way that makes them easy to find and retrieve[2]. Additionally, the storage area should be protected from environmental hazards, such as water, heat, or humidity, which could damage or destroy the answer sheets. Physical storage may also require regular maintenance, such as periodic cleaning and inspection, to ensure that the answer sheets remain in good condition. This process typically involves organizing and categorizing forms based on their content[15], such as by form type, date, or subject matter. Storing answer sheets of students is an important part of the record-keeping process in educational institutions. These answer sheets contain valuable information about the student's academic performance and progress and must be stored in a secure and organized manner for future use or reference.

The process of storing answer sheets of students typically begins with collecting the completed answer sheets from the students after an exam or test[1]. These answer sheets may be in the form of physical paper sheets or digital documents, depending on the testing format. Once the answer sheets are collected, they must be organized and categorized based on their content and context, such as by course, exam, or student ID. This helps to ensure that the answer sheets can be easily retrieved and referenced in the future. In this Desktop application we need to store the hand filled answer mark sheets of each student that should be stored in a folder. Overall, storing answer sheets of students is an important part of record-keeping [14] in educational institutions, as it enables organizations to maintain accurate and complete records of academic performance for future use or reference.

Effective record-keeping practices can help ensure the security, accessibility, and reliability of these answer sheets, supporting better decision-making and compliance with legal and regulatory requirements. In summary, storing students' answer sheets is an important aspect of the educational process, and it can be done either in physical or digital form. Regardless of the storage method used, it is important to maintain the privacy and confidentiality of the answer sheets and protect them from damage, loss, or theft.

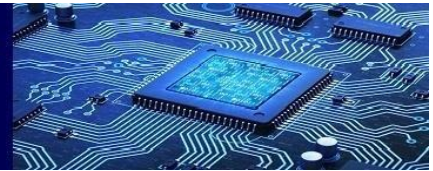
In this application, all the answer sheets of the students will be scanned and stored in papers folder for further image processing and cropping.

C. Cropping Students Answer sheets.

Cropping the mark portion of students' answer sheets is an important step in the grading process as it helps to ensure accuracy and fairness in the grading process. The mark portion of the answer sheet[2] typically contains the answers to the questions and the marks awarded by the teacher or grader. Cropping the mark portion of the answer sheet involves removing any identifying information, such as the student's name or ID number, to maintain anonymity during the grading process. This helps to prevent any bias or favoritism in the grading process, as the grader will not be influenced by the identity of the student. In addition, cropping the mark portion of the answer sheet can also help to ensure consistency in the grading process [8]. By removing any identifying information, the grader can focus solely on the answers provided and the marks awarded, without being influenced by any other factors.

Cropping the mark portion of the answer sheet can also be helpful for record-keeping purposes [6]. By separating the mark portion from the rest of the answer sheet, it can be easily stored and organized for future reference, such as in the case of appeal or review of the grading process. It is important to accurately take the required portion content from physical paper, hand-filled forms [1][2] because this information is often critical for various purposes such as students physical mark papers, employment, financial transactions, and medical records. Inaccuracies or omissions in the required portion content of hand-filled forms can lead to delays, rejection of the application, or even legal consequences. Therefore, it is crucial to ensure that the information taken from hand-filled forms is accurate and complete. Additionally, taking the required portion content accuracy can also help to ensure consistency and standardization in the information collected. This can be particularly important for hand-filled forms used across an educational institution like students answer sheets, organization, or industry etc.

Accurately capturing information from hand-filled forms also helps to maintain data integrity[8], which is important for analysis and reporting. So, if we consider the physical answer sheets of students, we can be able to analyze and maintain students' academic progress. In this application we can able to crop any portion from the scanned answer sheets of students and also we can able to crop as much as required portion from the answer sheet. the cropping which we perform in the first answer sheet will be done for rest of the all the answer sheets by storing the first answer sheet cropping coordinates by using Open Cv. Overall, cropping the mark portion of students' answer sheets



is an important step in the grading process, as it helps to maintain anonymity, ensure consistency, and facilitate record-keeping [9]. It is a best practice that should be followed by all educational institutions to ensure accuracy and fairness in the grading process.

D. Storing Marks in Excel

Storing marks in Excel can be important for various reasons, depending on the context in which the marks are being recorded. One of the most common reasons for storing marks in Excel is for academic record keeping. For teachers or academic institutions, Excel provides an efficient way to keep track of student performance over time. By storing marks in a spreadsheet, it becomes easy to sort and filter data, calculate averages and percentages, and identify trends in student performance[13]. This information can be used to create progress reports, inform instructional decisions, and communicate with parents and administrators.

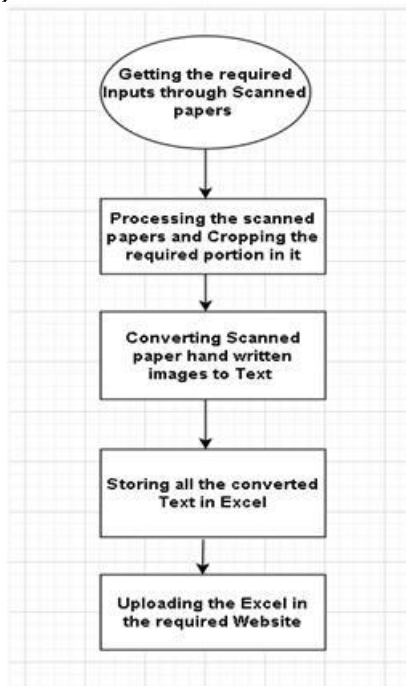


Fig.1. Flowchart of Desktop Application

By using this application, we have stored all the required cropped mark portion from multiple students answer sheets and then by using AspriseOCR (Optical Character Recognition) [6][7] API, we can able to convert cropped mark images in to text and by using openpyxl, we can store all the marks of the students in excel. Finally, storing marks in Excel can be useful for sharing data with others. Because Excel is a widely used software program, it is likely that others will be familiar with how to use it. This makes it easy to share data with colleagues, students, parents, and administrators, even if they are in

different parts of the world. By sharing data in this way, it becomes possible to collaborate on projects and initiatives that can have a positive impact on student learning and achievement. Another reason to store marks in Excel is for data analysis and reporting. When large amounts of data are collected, it can be difficult to make sense of the information without organizing it in a way that makes it easy to analyze [11].

Excel's powerful data analysis tools make it easy to generate charts and graphs that can help to identify patterns and trends in the data. This information can be used to make data-driven decisions that can improve student outcomes and inform policy decisions. In conclusion, storing marks in Excel can be an important practice for anyone involved in education or data analysis. Whether it's for academic record keeping, data analysis and reporting, or sharing data with others [15], Excel's features make it an invaluable tool for managing and analyzing data related to student performance and Fig1.shows the overall flowchart of the desktop application.

III. PROPOSED METHODOLOGY

The Proposed methodology of this desktop application is to automating data entry of students marks from the physical answer sheets. This can be done by collecting all the students answer sheets and scanning all those answer sheets and store it in a separate folder. so that we can crop the common required mark portion of first scanned answer sheet from the folder by using python modules. So by cropping the first scanned answer sheet, the rest of all scanned answer sheets of students will be cropped by using the coordinates of the first scanned answer sheet and saved in excel using python modules and packages. After saving all the students marks in excel, we can upload it to the required website. By following this proposed methodology, the educational institution can streamline its process of entering marks from physical papers into its website, saving time and effort for its staff and ensuring greater accuracy and efficiency in the process.

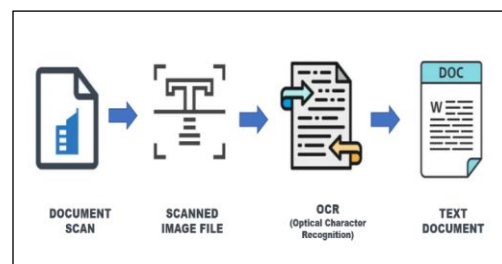


Fig.2. Overall Process



IV. APPLICATION SETUP

A. Methodology

This is a GUI desktop application built using Python that simplifies the process of entering marks from physical papers to websites by using automation. The application uses image processing and Optical Character Recognition (OCR) software to extract the marks data from scanned sheets of physical papers. The required portion of the scanned sheets containing the marks is cropped automatically by the application. The extracted text data is saved in an Excel file using the Openpyxl library for Excel file handling. The final step involves entering the data from the Excel file to the target website using automation. The application has a Graphical User Interface (GUI) that makes it user-friendly for individuals with minimal programming knowledge. The python modules and packages are used in this application are

1. Tkinter
2. PIL(Python Imaging Library)
3. Open CV (Computer Vision)
4. Openpyxl

B. Tkinter

This Desktop application is built by using Tkinter which is a standard Python library that provides a Graphical User Interface (GUI) toolkit for building desktop applications. It is a wrapper around the Tcl/Tk GUI toolkit that was originally developed for the Tcl programming language. Tkinter provides a set of widgets, such as buttons, labels, text boxes, and canvas that can be easily incorporated into a desktop application. Tkinter also provides a set of geometry managers, such as pack, grid, and place that help to position and size the widgets within the application window. All the buttons, widgets, text boxes, labels present in the desktop application are built with the help of tkinter and also it is very easy when compared with other python packages and added all other functionality with the help of tkinter. It is easy to learn and use, making it a popular choice for beginners to develop desktop applications. The library is cross-platform; meaning that applications built using Tkinter can run on different operating systems such as Windows, Linux, and Mac OS X. Overall, Tkinter is a powerful and versatile tool for building desktop applications with a user-friendly interface.

C. PIL(Python Imaging Library)

PIL (Python Imaging Library) is a popular Python library used for opening, manipulating, and saving various image file formats. This application uses PIL for opening students scanned answer sheets with various image file formats and it provides other image processing functionalities like resizing, cropping, filtering, and color conversion. PIL is a powerful tool for working with images in Python and is widely used in various fields, including computer vision, machine learning, and digital art.

D. Open CV(Computer Vision)

OpenCV is a Python library that allows you to perform image processing and computer vision tasks. This application uses Open CV for image processing which means the scanned answer sheets of the students and performing several functions. OpenCV provides a wide range of computer vision and image processing functions, such as object detection and recognition, feature extraction, image segmentation, and video analysis. It is written in C++ and has Python bindings for easier integration into Python applications. It is cross-platform, making it available on Windows, Linux, and macOS, and has a large community of contributors that constantly work on improving the library.

E. Openpyxl

Openpyxl is a Python library used for working with Excel spreadsheets. This application uses openpyxl for storing marks of students and also other text data detecting from scanned answer sheets, or any other hand filled forms etc. We can be able to change the mark entering cell in excel by changing the openpyxl part in the code. It provides functionality for reading, writing, and modifying Excel files, including support for various formats such as XLSX, XLSM, XLTX, and XLTM. Openpyxl allows users to create and manipulate Excel worksheets, rows, and columns, as well as to add charts, images, and formulas. It allows users to read, write, and modify Excel files programmatically, making it a powerful tool for automating tasks and working with data.

V. WORKING OF THE APPLICATION

This application is a desktop GUI application that is designed to automate the process of entering marks from physical papers to websites. The



GUI Desktop application runs and displays the frontend which was designed with the help of tkinter by providing buttons, widgets, labels, text boxes etc. This application has 3 buttons with 3 major functionalities. They are

- Cropping
- Storing in Excel
- Uploading in Website

These are all the functionalities available in the front end of this desktop application.

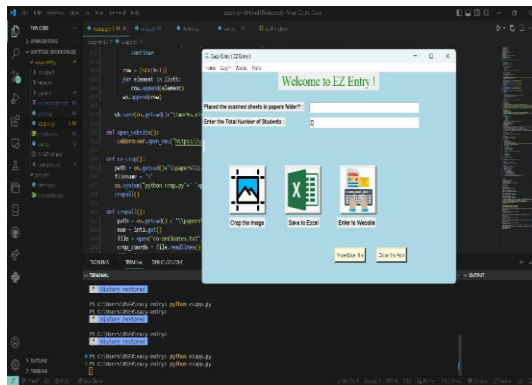


Fig.3. Frontend of Desktop Application

It uses image processing and OCR software, other OCR API like Asprise OCR API, Google cloud vision API etc. To extract the marks data as text from scanned sheets of physical papers. The required portion of the scanned sheets containing the marks is cropped automatically by the application.

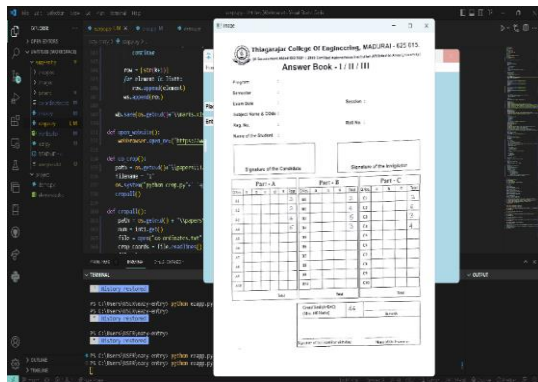


Fig.4. Processing the Scanned paper

The application uses image processing and computer vision techniques to automatically crop the required portion of the scanned sheets containing the marks data. This eliminates the need for manual cropping, making the process more efficient and accurate. The application uses the PIL (Python Imaging Library) and OpenCV (Computer Vision)

modules for image processing and cropping. The image is first loaded into the application using the PIL module, and then processed using the OpenCV module to identify the required portion of the image containing the marks data.

The application uses different techniques such as thresholding, contour detection, and morphological operations to identify the region of interest containing the marks data accurately. Once the region of interest is identified, the application uses the PIL module to crop the image and extract the required portion. The automatic cropping feature of the application ensures that only the relevant portion of the scanned sheet is processed, eliminating the need for manual cropping, and reducing the chance of errors.

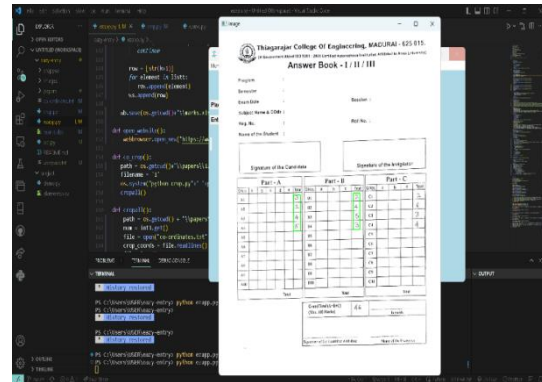
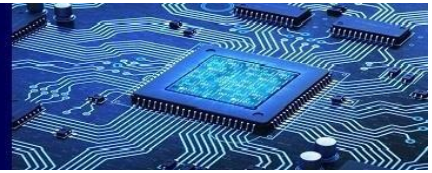


Fig.5. Cropping the Mark portion

This feature also makes the process faster and more efficient, allowing for a faster turnaround time for entering marks data. Overall, the automatic cropping feature of the application is an essential component of its functionality, which significantly simplifies the process of entering marks from physical papers to websites, making it more efficient and accurate. We can crop any part of scanned image by using this application.

The mark portion cropping is a crucial feature in this application, as it plays a vital role in automating the process of entering marks from physical papers to websites. The automatic cropping of the required portion containing the marks data eliminates the need for manual cropping, making the process more efficient, accurate, and less error prone. Without the mark portion cropping feature, the application would require manual cropping of each scanned sheet, which can be a time-consuming and tedious task, especially when dealing with many sheets.

Manual cropping can also lead to errors, as it is easy to accidentally crop the wrong portion of the image, leading to inaccurate data extraction. By



automating the cropping process, the application can accurately identify the region of interest containing the marks data, and crop only the relevant portion of the scanned sheet. This feature ensures that the marks data is extracted accurately and efficiently, leading to faster and more accurate results. In summary, mark portion cropping is an essential feature of this application, which significantly simplifies the process of entering marks from physical papers to websites, making it more efficient, accurate, and less error prone.

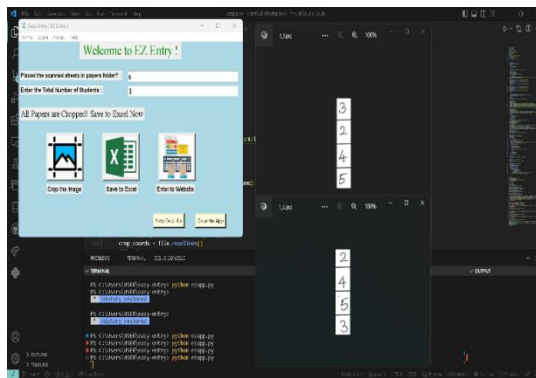


Fig.6. Cropped Mark

Converting the cropped image to text is a crucial feature of this application, as it enables the automatic extraction of marks data from physical papers, eliminating the need for manual data entry. The Optical Character Recognition (OCR) software is used in this application to convert the cropped image to text. In this application AspriseOCR (Optical Character Recognition) API is used which helps in converting cropped mark images to text.



Fig.7.OCR process

Without the OCR software, the application would not be able to extract the marks data automatically, and manual data entry would be required, which can be a time-consuming and error-prone process. The OCR software enables the

application to recognize and extract the text from the image accurately and efficiently, allowing for faster and more accurate data extraction. Furthermore, converting the cropped image to text enables the extracted data to be stored in a format that can be easily processed and analyzed, such as an Excel file. This feature is essential in academic institutions, where the marks data may need to be analyzed for statistical purposes or to identify trends in student performance. Overall, converting the cropped image to text is an important feature of this application, which significantly simplifies the process of entering marks from physical papers to websites, making it more efficient, accurate, and less error prone. Entering cropped mark data into an Excel file is an important feature of this application, as it allows for the efficient management and processing of the extracted data. The Openpyxl library is used in this application for Excel file handling, which enables the application to create, read, and modify Excel files.

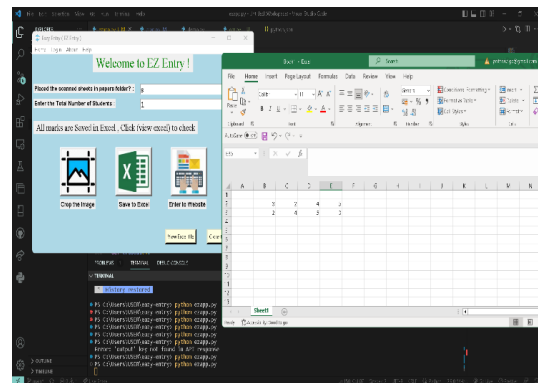
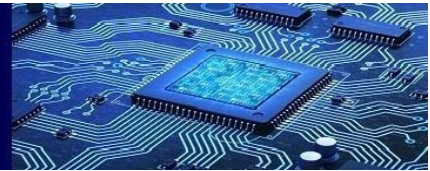


Fig.8. Marks are saved in Excel.

After getting marks in excel, we can upload the excel to our required websites. Excel is a powerful tool for performing calculations, generating charts and graphs, and organizing data, and uploading the data to a website can enable others to access and analyze the data as well. Entering cropped mark data into an Excel file is an important feature of this application, as it allows for the efficient management and processing of the extracted data. The Openpyxl library is used in this application for Excel file handling, which enables the application to create, read, and modify Excel files. Moreover, entering the cropped mark data into an Excel file provides an additional layer of accuracy and verification, ensuring that the data is correct before being entered into the target website using automation. This feature minimizes the possibility of errors during the data entry process, ensuring that the final data is accurate and reliable.



VI. CONCLUSION

The GUI Desktop application provides a streamlined and automated solution for entering marks from physical papers to websites. By utilizing image processing and OCR software, the application is able to extract marks data from scanned sheets of physical papers quickly and accurately. The extracted data is then saved in an Excel file, which can be easily imported into the target website using automation. The application is user-friendly and accessible to individuals with minimal programming knowledge, thanks to its GUI. It is built using several Python modules and packages, including Tkinter, PIL, Open CV, and Openpyxl. These modules and packages enable the application to perform complex operations with ease and efficiency. Overall, the application offers significant benefits for anyone involved in entering marks or grades from physical papers to websites. It reduces the time and effort required for manual data entry, while also improving the accuracy and reliability of the data. The application can also be customized to meet the specific needs of different users and applications.

REFERENCES

- [1] Lin WoanNing, Yap keemSiah, M .Khalid, M.Yusof“Design of an automated data entry for hand filled forms”,Intelligent systems and Technologies for the New MilleniumIEEE, September 2000.
- [2] AC ChengJohnson “Evaluating automated electronic case report form data entry from electronic health records”, Published in Cambridge University, 14th December 2022
- [3] Jeffrey C Bauer, EileenJohn, Christopher L wood, DebraPlass, Dale Richardson “Data Entry Automation Improves Cost, Quality,Performance, and Job Satisfaction in a Hospital Nursing Unit”2020.
- [4] Gorman, L.O., Kasturi, R., “Document Image Analysis, California “, IEEE Computer Society Press, 1995.
- [5] Dorransoro, J., Fractman, G., Santa Cruz, C., “Large scale neural form recognition,” Industrial Applications of Neural Networks, 1998, pp.354-362.
- [6] H belgacem, X Li , D Bianculli, L C Briand “A Machine learning approach for Automated filling of Data Entry forms” 2022.
- [7] S Ghulyani, D Jain, P Singh, S Joshi, ”A Data entry Optical Character Recognition tool using CNN(Convolution Neural Network” ,2022.
- [8] RajavelParthiban,Ezhilarasi ,“Optical Character Recognition for English handwritten text using RNN, “International conference on System, Computation, Automation and Networking , July 2020.
- [9] HV Jagadish, MageshJaya Pandian”Automating the design and construction of query forms” IEEE Transaction in Knowledge and Data engineering 2008.
- [10] Allen Cypher “Automating data entry for end users.”IEEE Symposium on Visual language and human-centric Computing 2012.
- [11] Oscar Diaz, I otaduy, G Puente, “User driven automation of web filling forms “Springer 2013.
- [12] M AI Dogether, Y AI Muallem, M Househ “The impact of Automating Laboratory requestForms on the quality of healthcare services” 2016
- [13] R Parasuraman, TB Sheridan,C.D. Wickens “A model for types and levels of human interaction with automation”,Transactions on systems, Man, and Cybernetics IEEE 2000.
- [14] Alistair Sutcliffe “Data Entry Interfaces A Human Computer interfaceDesign” 1988.
- [15] Automating data extraction from prescription document images to reduce human error L Zahray 2019