HANDWRITING-READER

By

HAOYU, SU

LONG, MA

XIAOSONG, YUAN

ZHAOYANG, WANG

Design Document for ECE 445, Senior Design, Spring 2022

TA: XIAOYUE, LI

24 MARCH 2022

Project No. 16

Contents

1. Introduction

1.1 Objective

With the economic development and social progress, the business scope of the insurance industry is expanding, and the competition within the industry is becoming increasingly intense. In the pursuit of higher revenue efficiency and cost savings, there is room for improvement in the original insurance policy entry method. The traditional insurance policy is filled by handwriting and then manually entered into the electronic system, which not only wastes a lot of labor and time costs, but also has a significant error rate.

Our solution for the transcription process of the insurance paper is a system that can read handwritten insurance. We aim to reach the accuracy of 99%. The physical part of the system is set up as a scanning device that reduces environmental interference and generates images that are passed through the interface into the processing application on the computer. Occupied with a trained machine learning model, the basic function of the system is to recognize handwritten characters. To increase identification accuracy, the system can collect address information and correct it with the actual address searched on maps, and also compare customer information such as name, phone number and address with data in database. The system can fill out the electrical insurance forms automatically and saved the scanned photos of the handwritten documents as references.

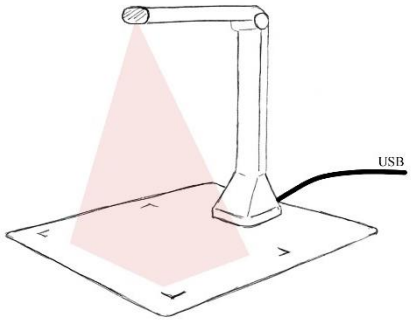


Figure 1 Scanner

1.2 Background

In the insurance industry with a large number of users and businesses, insurance companies must adopt an efficient digital business processing system to centralize business data [1], while handwritten policy documents help to build the identity of the policyholder [2], which requires that the conversion of written documents to electronic documents must be completed.

We also learned that one of the existing countermeasures for the error occurred by manual entry is the use of e signs that automatically follow during manual entry to alert operators [3]. Our system aims to solve this problem at the root by replacing manual input steps and the requirement of the ye-catching signs.

The existing policy recognition technology OCR theoretically can reach 80% correct rate [4], and the company currently uses the recognition correct rate is only 70%, which still has a reliance on manual review. Our system hopes to find a better algorithm model, supplemented by automatic correction, to minimize the error rate and reduce the amount of work required for review.

1.3 High-level Requirements List

The physical part of the system needs to be able to scan the A4 size policy and generate the image of at least 1200 dpi for input to the computer through the interface.

The application needs to be able to recognize both Simplified and Traditional Chinese characters in each area with at least 98% accuracy.

The application can search the map for the identification results of the address part, correct errors in the identification results based on the search results to reach 99% accuracy, and export the final result to excel format

2 Design

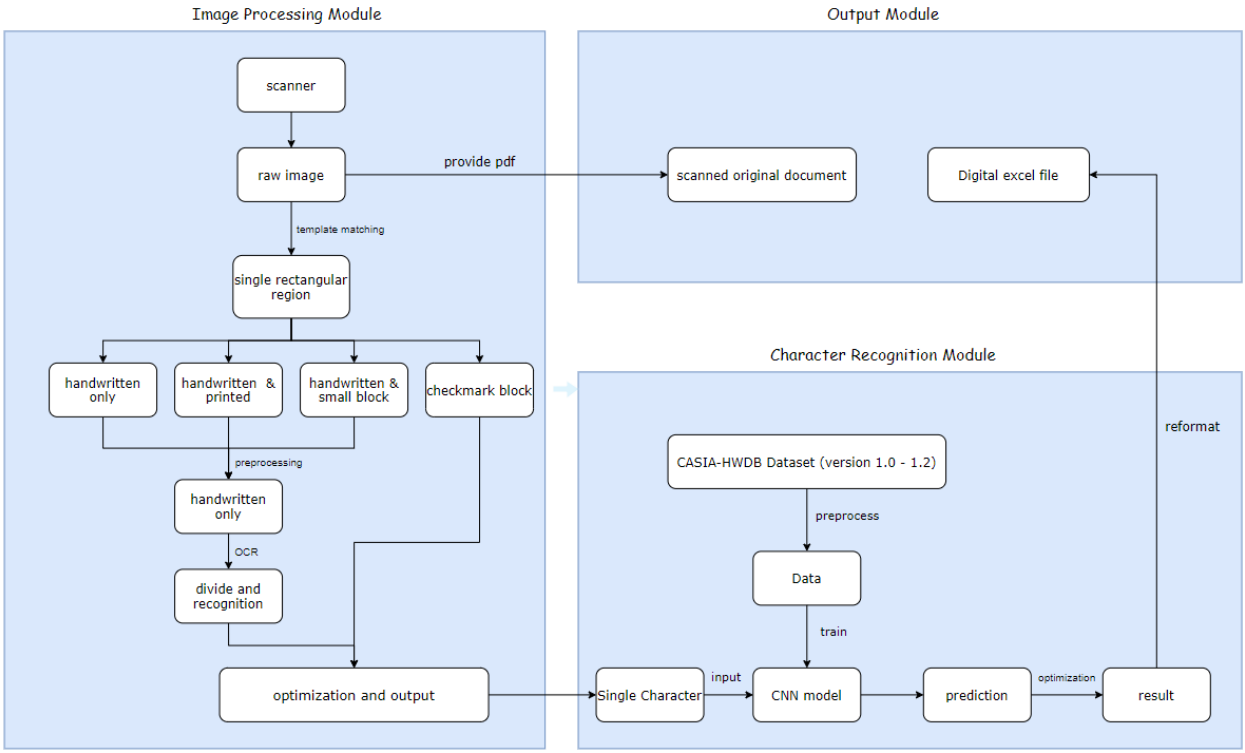


Figure 2. Block Diagram

2.1 Block Diagram

Our whole design is primarily software-based and have three modules – the Input Module, the Character Recognition Module and the Output Module. The Input Module have a scanner to scan the contract and process the scanned image. The processed image is then passed to the Character Recognition Module to do the recognition task. Finally, the digitized contract from the Character Recognition Module and the scanned contract will form the Output Module for our users to use and reference.

2.2 Image Processing Module

The input module is required to create clear and characteristic input data, particularly preprocessed images, for character recognition module through interaction of both hardware and software.

2.2.1 Scanner

The original data is handwritten contract paper, so the first step is to scan it into a digital image, not only for the copy version but also for the input data.

2.2.2 Image Processing ("Raw Image" and "Processed Image" in the block diagram)

The project initiator requires us to achieve more functions other than successful recognition to the handwritten content. Rather than using machine learning method such as NLP to extract content such as location, phone number and name through recognized texts, since the template of the contract is provided, we can use template matching method to derive handwritten sections of the raw image. Also, we can use OpenCV image processing methods such as cropping, rotation, mirroring, chroma adjustment and saturation adjustment to get clearer image.

2.3 Character Recognition Module

2.3.1 Dataset and Preprocessing

We have found a paper written by Pavlo Melnyk and his coauthors that can meet our needs to a certain extent [10]. They have established a neural network that can read single handwritten Chinese characters and the model have 97.61% accuracy on the test dataset [10]. However, after our test, it is not satisfying. The model only recognized four characters correctly among the six characters we tested. And two most important characters "市"(city) "区"(district) were not recognized correctly. We think the model is effective because it can have a high accuracy on the test dataset, the reason it has a poor performance on our test data is that the training dataset is not large enough. The authors used CASIA-HWDB 1.1 datasets containing 2,678,424 samples written by 420 and 300 people respectively [10]. Though the number of training samples is in large amount, CASIA-HWDB 1.1 only contains 3755 Chinese character classes of GB2312 level 1 [11-12]. However, GB2312 level 1 contains 6763 most frequently used Chinese characters [12], which means the range of 3755 classes is not wide enough. Now the CASIA-HWDB datasets have two more datasets, which are 1.0 and 1.2 [11]. CASIA 1.0 – 1.2 together contain 3,895,135 images including 3,721,874 Chinese characters and 173,261 symbols. All those images are classified into 7356 classes, which cover all 6763 most frequently used Chinese characters in GB2312 level 1 and some most widely used symbols [11]. With those datasets, we think that we can have a higher accuracy and a broader recognition range. And we can even find volunteers to write key characters like "市"(city) and use them to further train our model to make sure that it can read those key characters correctly. In addition, the model from the paper can only read one character at a time. Our contribution should also include separating single characters from the processed image, recognizing them and outputting the results as a paragraph. That means our model should also be

able to read punctuations. The CASIA-HWDB 1.1 dataset has already contained 171 Arabic numerals and punctuations. So, reading punctuations should not be a problem.

All images in the datasets are in size of 64 by 64 by 1 (they only have gray scale) and stored in

Item	Type	Length	Instance	Comment
Sample size	unsigned int	4B		Number of bytes for one sample (byte count to next sample)
Tag code (GB)	char	2B	"啊"=0xb0a1 Stored as 0xa1b0	
Width	unsigned short	2B		Number of pixels in a row
Height	unsigned short	2B		Number of rows
Bitmap	unsigned char	Width*Height bytes		Stored row by row

Figure 3. Format of .gnt files [11]

Figure 4. Part of CASIA-HWDB 1.1 dataset

several .gnt files. The format of .gnt file is shown in Figure 3. The "Sample size" item means the number of bytes used to store one image, and there are lots of these digitized images in one .gnt file [11]. The "Tag code" means the corresponding tag of each character in GB2312 [11-12]. It is worth noting that the two bytes used to store the tag code are in reverse order, like the instance in Figure 3 shows. Two bytes are used to store the width of the image and two more bytes are used to store the height [11]. The last item is the "Bitmap" of the image, which is the pixel value stored row by row. We extracted some images from the dataset and converged them into .png images. They are shown in Figure 4.

In the dataset preprocessing step, we will firstly extract useful information ("Tag code" and "Bitmap") from the .gnt files and store them into a single .hdf5 file. Then, we will assign x ("Bitmap") and y ("Tag code") labels for each image so that they can be used by tensorflow to train.

2.3.2 Model and Prediction

The model is supposed to be trained with machine learning method with large amounts of datasets so that it can reach an accuracy of 99%. We have searched some related papers online, most of which used CNN (Convolutional Neural Network) and other deep learning method [4-7]. Based on our research, CNN and other deep learning methods have a better performance than other machine learning methods in solving Chinese character recognition problems [8]. So, we also plan to use CNN in our model. Most of the papers can reach an accuracy around 94% and some of them can have 99% accuracy on a specific dataset [4-7, 9]. Our goal is to build a model that has 99% accuracy on the dataset of insurance contracts. We plan to use the one single pure CNN model in the Character Recognition module. Pavlo Melnyk's paper will provide reference for us because it has the highest accuracy to our knowledge [10]. However, his model is not able to meet our requirements so we will not copy it. Melnyk Net (Melnyk named his model as Melnyk Net) has 15 layers and it performed well on 3755 classes. However, according to our test, it only has around 20% accuracy on our dataset with 7356 classes. We think the reason is that it is not deep enough to handle the enlarged number of classes. We plan to build a model deeper than Melnyk Net (however, how deep is effective remains to be tested and trialed). But Melnyk Net will provide reference for when choosing activations functions and pooling methods and so on.

When do the prediction, all sized images can be used because they will be padded and then processed into 64 by 64 by 1 by our code before passing to the model. Melnyk has already realized this function in his code and it works effectively by our test [10]. So, we will just use his prediction functions and clearly reference his paper in those functions.

2.3.3 Non-ML (Machine Learning) Optimization

With the prediction of our model, we also plan to use some non-machine-learning method to do further optimization. There are three methods that we want to use:

1. Identify important information such as address, phone number, work location and so on and verify them through online database or software. Information like addresses, company names can be found online. So, with this method we can have the correct prediction even if some of the characters are not correctly recognized.

2. Scan the handwriting clearly into document so that the users can reference to the original contract if needed. Many Chinese names use rarely-used characters which may go beyond our training set. In this case the users can refer to the original document to make sure that the names are correct.

3. When a character is unseen, use the character ahead of and behind it to guess the meaning (which is frequently used when reading writings in classical Chinese) and use a proper replacement. This method can be used in address recognition because online datasets will show candidate entries according to the meaning when there is no match. This method will not be used in human name recognition.

2.4 Output module

The output module is required to form a digital recognized version of handwritten contract and apply some functions with the extracted information. The digital contract template is provided. So, it only needs to fill recognized text into its corresponding sections in the template.

2.5 Requirements and Verifications table

2.5.1 Image Scanning and Processing

Scanner

Requirement	Verification
1. Scanning/Loading into Single Page: We use a scanner to read the handwritten contract or directly load the provided scanned PDF file. The image from the scanner should be clear enough to process and the PDF file should be transformed to PNG form. 2. The template matching for each type of contract should be as precise as possible, which means it should deal with the situation that some part of the character exceeds the border. The border error could be within 5 pixels which won't cause trouble in handwritten region extraction.	A. With a fixed position and height, the scanner is used to scan A4 size paper and outputting images at a resolution of 1200dpi. B. Each page of the PDF file can be successfully converted into a PNG file and stored in the specified folder. When the handwritten region is extracted, there shouldn't remain other content or miss some part of handwritten content.

Single Rectangular Region Extraction

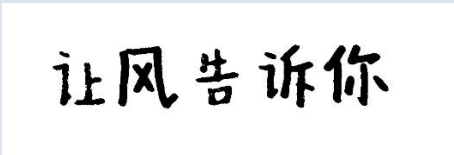
Requirement	Verification

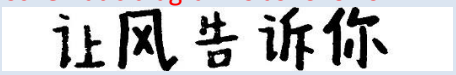
1. All the blocks that need to be identified from the whole document should be extracted. 2. The rectangular regions should be classified by the content of each block for subsequent processing.	A. As the policy format is fixed, the selection range for each block is pre-defined. Each block is extracted without the black edge of the block and ensures that everything within the range is selected and that all filled-in information is not lost.

Handwritten Content Extraction

Requirement	Verification
1. Replacement of Printed Text: In blocks where printed text and handwritten text coexisted, the printed text needs to be replaced with plain white to not interfere with subsequent cropping and recognition. **2.** Checkmark identification: checkmark is also what the policy completer manually fills in and needs to be recognized. However, checkmark recognition does not involve Chinese character recognition, so when the option is fixed, only the checkmarks need to be recognized.	A. For the text, the printed content should be all covered with white pixel and the handwritten content notes will not be erased. B. For the checkmark, the recognition result is a single selection result of multiple-choice and is consistent with the correct result.

Preprocessing

Requirement	Verification
1. Binarization of Image: we choose to use the OTSU algorithm to binarize the inputted image (Handwritten only). The output images should be divided into two parts, which means the boundaries are visible and there is no gradual changing situation. And the result images should be valid as the inputs of the cutting functions. 2. Horizontal cutting: cutting images in the horizontal direction to remove the “write edges” up and down the images. The output images should have on “write edges” in up and down directions. And the result images should be valid as the inputs of the vertical cutting function. After processing by the vertical cutting function, the image containing a single Chinese character should be valid as the neural network input.	A. We use a colorful picture as the input image of the binarization function. The output image is divided into black and white. And we can see the boundaries of the white background and the black Chinese characters. The schematic diagram is as follows:  887*296  887*296 B. We input the previous output image into the function of horizontal cutting, if the result is as out prediction, it can show that the previous binarization function is valid. And if the “write edges” are removed successfully, it can prove

	our horizontal cutting function is implemented. (But it does not mean the total success of this function, the output should be valid as the next functions' input.) The schematic diagram is as follows:  887*124 "White edges" are successfully removed.

Divide and Recognition

Requirement	Verification
1. Splitting the image into single character: We should vertically cut the input image, and the objective of this function is to split one image with multiple Chinese characters into several images, and each of them contains only one Chinese character. This module should finish this task and the output images should be valid as the input of the machine learning model.	A. We input the previous output image into the function of horizontal cutting, if the result is as out prediction, it can show that the previous horizontal cutting function is valid. The result is as follows:  Image0: 104*124 Image1: 120*124 Image2: 87*124 Image3: 111*124 Image4: 130*124 Then we input these images into the neural network to see whether they are valid.

2.5.2 Character Recognition

Data preparation

Requirement	Verification
1. Dataset should cover all GB2312 level 1 characters because they are the most frequently used Chinese characters and are frequently used in Chinese address, position descriptions and other parts of insurance contracts. 2. Dataset should include numerals and symbols because dates, addresses and parts related to money could be using numerals and the contract user could fill symbols like @ in email part and ¥ before money. 3. The preprocessed dataset should be able to be used in tensorflow without error.	A. According to the official website of CASIA dataset, the datasets we are using together can cover GB2312 level 1 [11]. B. According to the official website of CASIA dataset, the datasets we are using together can cover 171 symbols including @ and ¥ [11]. C. Train a test code (a simple CNN model) with our dataset. If there is no error, then the preprocessed dataset is in correct format.

Model Design and Train

Requirement	Verification
1. The model should accept a 2D binary image with a particular size as input and give a prediction for the character (among 7356 kinds). 2. The model should be deep and robust enough to maintain sufficient weights due to the high accuracy requirement. 3. After training with enough epochs, the validation accuracy should be above 95%.	A. Use the processed dataset (71GB), split it into training and validation set. Train the model with 50 epochs and save the weights after each epoch. If the result is still not converged, enlarge number of epochs. B. Check the validation accuracy. If it is below 95%, consider deepen the CNN model.

Prediction

Requirement	Verification
1. The pretrained model should successfully accept the single character binary image as input and give correct result (above 98% accuracy). 2. If the accuracy is not satisfied, the model should give the top 3 choices of prediction result (among 7356 kinds of Chinese characters) for optimization.	A. Use handwritten contract provided by the insurance company. After the image processing module, which create single character binary image, we will test 1000 characters and it should give above 980 correct predictions. B. If the accuracy is not satisfied, for the wrong predicted character, we will check top 3 prediction result to see if the correct character is included.

2.5.3 Output

Requirement	Verification
1. All characters predicted by the model and information indicated by recognized checkmark should be reformatted into the correct columns of the excel file provided by the company.	A. Use 5 handwritten contracts, after whole image processing and character recognition, check if the information appended into the excel file matches.

2.6 Tolerance Analysis

2.6.1 Scanner

The only physical component of our project except the computer is a scanner, which scans the contract and output the scanned image to the computer. So, the first section primarily focuses on the scanner. The candidate scanner uses USB2.0 on the computer as power supply. It has working voltage from 4.75V to 5.25V and working current is 900mA. The output voltage of USB on the computer is 5V ± 0.2V, so the voltage should meet the requirement. The output current of laptop USB is 500mA, which is lower than the required working current of the scanner. However, the output current of desktop USB can be as high as 1A, so a desktop should satisfy the requirement. We can even build a small circuit to connect two USB ports of a laptop in parallel so that a laptop can provide 1A current for the scanner. We also consulted the customer service of the scanner shop. The customer service said that though the working current of the scanner is 900mA, it could work

well with a laptop USB port.

2.6.2 Image Processing Module

The requirement for our whole system is to have 99% accuracy. However, it is very difficult for a machine learning model to have such high accuracy in a Chinese character recognition task. The prerequisite for the machine learning model to work is that the image processing module and successfully extract each handwritten character in the insurance contract. So, the requirement for the image processing module is 100% accurate to extract all handwritten characters and there is no tolerance.

2.6.3 Character Recognition Module

The accuracy of our whole system is depended on two parts, which are image processing module and character recognition module. As the whole accuracy requirement is 99% and the requirement for image processing module is 100%, the accuracy requirement for character recognition module is $\frac{\text{whole accuracy}}{\text{image processing module accuracy}} = 99\%$. There is no tolerance otherwise we cannot meet the requirement of having 99% accuracy on the whole system.

3. Cost and Schedule

3.1 Cost

According to the data showed in PayScale, the average annual salary of a project engineering graduated from University of Illinois is $75874. Assume they work 40 hours per week, we can know the estimated hourly rate of them is about $36.5.

We have 12 weeks to complete this project, and assume weekly hours worked by each member of our team is 20 hours. Therefore, the estimated salary of our team is

4*12*20*2.5*$36.5=87600$

Because our project is based on machine learning, the requirements for the hardware part are relatively low. We just need to buy an external scanner to meet the requirement of scanning the handwriting insurance papers. Our market research concluded that buying this scanner.

Description	Manufacturer	Parameters	Module	Cost	Quantity	Total Cost
Scanner	Deli	1200dpi CMOS	Input	￥399	1	￥399

Table 1 Market Research Result of the Scanner

No machine shop labor hours needed to complete the project.

3.2 Schedule

Weeks	Tasks	Member
Week1	Find and read article. Get familiar with existing code.	All
Week2	Modify and run code, can generate a correct sentence based on input character sequence. Download new dataset and do data processing.	Yuan, Wang
	Look for and purchase proper scanner. Find the tutorials of Python OpenCV package. Learn about image processing by using Python while waiting for the scanner to arrive	Ma, Su
Week3	Train model based on new datasets and analyze result.	Yuan, Wang
	Connect the bought scanner to our computers and master the scanning process. Use the image processing knowledge which we learned last week to process the insurance policy from the insurance company and complete the image cropping of the "Name" field.	Ma, Su
Week4	Reformat code framework. Append recognition result into excel file.	Yuan, Wang
	Finish cropping the images in the remaining columns and try to find ways to split the image which contains multiple Chinese characters into the image containing single Chinese character. Try to apply the method to process the images.	Ma, Su
Week5-7	Find ways to improve the model if the result doesn't reach our expect.	Yuan, Wang
	Complete the process of splitting multiple Chinese characters images into a single Chinese character image. Finish pre-processing the given data sheet by the insurance company, sort and package the processed data, and send to another group for machine-learning model test.	Ma, Su
Week6 - End of Semester	Improve the accuracy of machine learning model recognition and try to use non-machine-learning method to improve the final accuracy of Chinese character recognition.	All

Table 2 Week-by-week Schedule

4. Ethics and Safety

4.1 Ethics

There are several potential ethics and safety problems in our project. Here I will discuss these problems and give what we think are the solutions, which may not be perfect and solve these problems totally, but they are all options that our team have thought about as much as possible, and think are feasible.

Firstly, according to IEEE Code of Ethics I.4[13]., we promise we will avoid unlawful conduct in professional activities and reject bribery in all its forms.

The most serious problem is privacy data leakage and loss. Insurance company users fill in their policies with private data that they do not want others to know, such as personal phone numbers, addresses, ID numbers, etc. If the company read the filled-out insurance papers by an electronic handwriting reader, then the possibility of these private data being compromised or lost will increase. There are several reasons. First, data stored in the electronic database is more vulnerable to malicious attacks by competitors rather than data stored in physical repository. Second, due to poor account lifecycle management, the authority division and authentication identification methods are out of control, severe consequences will be resulted in, such as unequal access rights of personnel to data at secret level, high secret data flowing to low authority accounts, and secret data flowing to accounts without authority, etc. Moreover, since the data entry system is connected to the Internet, securing the data will become a difficult task under the open environment of the Internet.

Our team give our own considerations here: Even if the data is entered manually, like the insurance companies are doing now, it will eventually be stored in the insurance company's own database management system. Therefore, on the data storage side, the risk of data leakage or loss is not actually increased. Although there is indeed an increased risk on the data entry side, electronic devices could avoid some potential risks from original manual entry method. For instance, no one can guarantee that the employees responsible for entering this privacy data are reliable enough and will never leak or lose this data. And we designed the structure of the reader to be as secure as possible, ensuring that all input and output ports are under the control of the insurance company itself. Although we cannot prevent 100% data leakage and loss, we think such security is enough.

Additionally, there is a job loss problem. Inevitably, once the insurance companies decide to use the designed handwriting reader to enter insurance policies and papers, several people who originally did this work will be at risk of losing their jobs. However, our team feels that this is actually a saving of human resources, and the manpower saved can be put into more meaningful work instead of mechanically entering policies. Besides, even if the accuracy can reach 99%, the machine is not completely reliable. The insurance companies need people to reviewing the results of machine completion. The person who was responsible for entering policies can move on to complete the reviewing, which is much easier for them.

If the results of this machine do cause some people to lose their jobs, our team believes that this is a necessary sacrifice in the development of technology. By analogy, autonomous driving systems will replace the jobs of drivers and vending machines will replace the jobs of ticket takers, and we do not see the emergence of these automated machines as ethically problematic.

This is a summary of the ethical issues and our team's view. If something is not right, we are happy to accept the right comments and modify our design according to them. According to the IEEE Code of Ethics I.5[13]. “To seek, accept, and offer honest criticism of technical work, to acknowledge and correct errors, to be honest and realistic in stating claims or estimates based on

available data, and to credit properly the contributions of others;" We will publish our contact information after we have completed the development and actively accept reasonable criticism. And according to IEEE Code of Ethics III[13], our team members will monitor each other and strive to comply with ethical guidelines while developing the project of handwriting reader.

4.2 Safety

I would like to discuss the security issue in two parts, physical part and virtual part.

The virtual part is, as I said before, that there is a risk that the user's private data will be compromised or lost. Our team will refine our design to avoid these problems as much as possible. See section 3.1 for data privacy concerns.

Then, for the physical part, because we don't have any teammates who major in mechanical engineering, we will not do the mechanical part of the design ourselves, like how to arrange the scanner and the computer properly. Thus, we don't have the ability to increase the safety of the mechanics.

However, we can make our recommendations on how to use these machines safely. First of all, the operators must be trained and only after they are thoroughly familiar with these machines can they be put to work. As well, avoid using these machines in harsh environments, such as harsh heat or rooms with safety hazards. Additionally, our scanner part is external outside the whole machine, the operator needs to protect the safety of the interface. And they need to operate in strict accordance with the prescribed parameters to avoid short circuits, disconnections and other electrical connection safety hazards. Because our entire equipment is electric and will be installed in an indoor environment, a circuit problem will cause serious consequences, such as fire. Therefore, when using our equipment, electrical safety must be ensured, and when not in use, it must be stored safely in an environment out of the reach of unoccupied people. Finally, we hope that the company can invest in regular maintenance of the equipment and set up a safety officer to be responsible for the regular safety maintenance and emergency inspection and repair of the whole equipment.

References

[1] X. Chen, Y. Yao, M. Zheng, “A core business system for insurance companies based on Struts and HIBERNATE architecture,” Computer Engineering, vol. 32, no. 4, pp. 264-266+271, Apr 2006.

[2] Z. Quan, “Research on dynamic handwritten signature-based authentication,” Doctoral dissertation, University of Science and Technology of China, 2007.

[3] J. Chen, “Design and implementation of a core business system for life insurance of New China Life Insurance,” Doctoral dissertation, Beijing University of Posts and Telecommunications, 2010.

[4] Y. Jia, “Exploring the application of OCR technology in the field of document recognition,” Financial Computerizing, no. 4, pp. 48-49, 2018.

[5] Y.-Q. Li, H.-S. Chang, and D.-T. Lin, “Large-Scale Printed Chinese Character Recognition for ID Cards Using Deep Learning and Few Samples Transfer Learning,” Applied Sciences, vol. 12, no. 2, p. 907, Jan. 2022, doi: 10.3390/app12020907.

[6] R. Shashank, A. Adarsh Rai, P. Srinivasa Pai, “Handwritten Character Recognition Using Deep Convolutional Neural Networks,” Recent Advances in Artificial Intelligence and Data Engineering. Advances in Intelligent Systems and Computing, 2022, vol 1386, doi: 10.1007/978-981-16-3342-3_2.

[7] W. H. Liu, K. Ming Lim and C. P. Lee, “Visually Similar Handwritten Chinese Character Recognition with Convolutional Neural Network,” 2021 9th International Conference on Information and Communication Technology (ICoICT), 2021, pp. 175-179, doi: 10.1109/ICoICT52021.2021.9527449.

[8] Novita, Suyanto, Y. Prasti Eko, “Bonferroni Mean Fuzzy K-Nearest Neighbors Based Handwritten Chinese Character Recognition,” 2021 International Conference on Data Science and Its Applications (ICoDSA), 2021, pp. 118-123, doi: 10.1109/ICoDSA53588.2021.9617488.

[9] D. He, Y. Zhang, “Research on Artificial Intelligence Machine Learning Character Recognition Algorithm Based on Feature Fusion,” Journal of Physics: Conference Series, 2021, Ser. 2136 012060.

[10] Melnyk, P., You, Z. & Li, K. A high-performance CNN method for offline handwritten Chinese character recognition and visualization. Soft Comput 24, 7977 – 7987 (2020). Available: https://doi.org/10.1007/s00500-019-04083-3

[11] “CASIA Online and Offline Chinese Handwriting Databases,” Available: http://www.nlpr.ia.ac.cn/databases/handwriting/Online_database.html

[12] “BG 2312,” Wikipedia, Available: https://en.wikipedia.org/wiki/GB_2312

[13] Ieee.org, "IEEE IEEE Code of Ethics", 2016. [Online]. Available: https://www.ieee.org/about/corporate/governance/p7-8.html