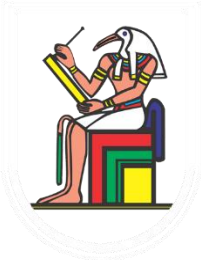


CMPN-446

Grades Auto-Filler

Project Report

Team - 0

Team Members:

Karim Yasser Negm 1180484

Seif Mohamed Emam 1180387

Arwa Ehab Mohamed 1180480

Sara Abdulsalam 1180461

Module 1 – Bubble Sheet

Pros & Cons

Pros:

- Almost all the test cases pass correctly with precise detection of bubbles

Cons:

- If multiple bubbles are selected, the choice is considered invalid. If any question has more than one correct answer, it will be considered invalid.
- If the paper has curved due to rolling, it will be almost impossible to detect the borders
- If the shading extends beyond the bubble, it will not be detected by the radius set in Hough detection

Grading Process

1) Pre-processing

The pre-processing step is the first step in the image processing pipeline. It is used to remove noise and to enhance the image. The pre-processing step is also used to prepare the image for the segmentation step.

Before applying any pre-processing however, we must first crop the image. The image is cropped to remove the background and to only include the image of the paper.

This is done by first finding all the contours in the image. It is assumed that the contour with the largest area is the contour of the paper. And the contour with the second largest area is the contour of the bubble sheet.

Once the contours are found the corners are extracted and image is warped using OpenCV's getPerspectiveTransform function.

After the image is cropped, the pre-processing step is applied. Using adaptive local thresholding, the image is binarized and to remove noise a median filter is then applied.

2) Segmentation and Analysis

Once the image is pre-processed, the segmentation step is applied. The segmentation step is used to segment the image into individual bubbles.

It can be summarized as

1. Apply circular Hough transform to find circles in the image.

2. Sort the circles from top left to bottom right and determine which circles belong to the same row.

- This is done by first finding the top left and top right bubbles in the unsorted list of circles.

- A line is then drawn between the top left and top right bubbles.

- Circles intersecting with the line are then added to the row.

- The row is then sorted from left to right.

- Repeat the process until all the circles are sorted into rows.

3. Rows are then further categorized into ID rows and answer rows.

- ID rows are rows before a large vertical gap. between the rows.

- Answer rows are all the rows after the ID rows.

4. The number of columns in the answer rows are determined by the large horizontal gap between the bubbles.

Great thanks to user S. Vogt on [StackOverflow](https://stackoverflow.com/questions/29630052/ordering-coordinates-from-top-left-to-bottom-right) for the main idea behind sorting the circles.

3) Detecting Number of Bubbles Per Row and Question

The horizontal distance between each bubble is calculated and we take double their mean as a threshold where if the distance between 2 bubbles goes beyond this threshold, we know that the choice for a specific question is done and we are detecting the bubbles of another question.

The warped image is then used again and pre-processed. Then a row is taken as a sample to calculate the average of all bubbles per row and set it as a threshold. For each row, if a bubble mean is above that threshold, then the bubble is filled. That's how a shaded bubble is detected.

4) Saving Data in Excel

Lastly, the detected choices of each question are compared to the model answer and the grades are saved into an excel sheet using csv.

Test Cases

ID

⓪ ① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨

⓪ ① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨

Name

Meow4

1 Ⓐ Ⓑ Ⓒ

2 Ⓐ Ⓑ Ⓒ

3 Ⓐ Ⓑ Ⓒ

4 Ⓐ Ⓑ Ⓒ

5 Ⓐ Ⓑ Ⓒ

6 Ⓐ Ⓑ Ⓒ

7 Ⓐ Ⓑ Ⓒ

8 Ⓐ Ⓑ Ⓒ

9 Ⓐ Ⓑ Ⓒ

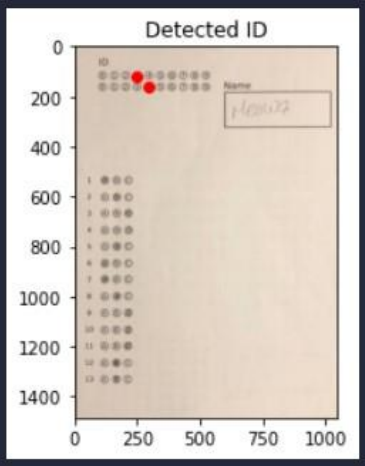
10 Ⓐ Ⓑ Ⓒ

11 Ⓐ Ⓑ Ⓒ

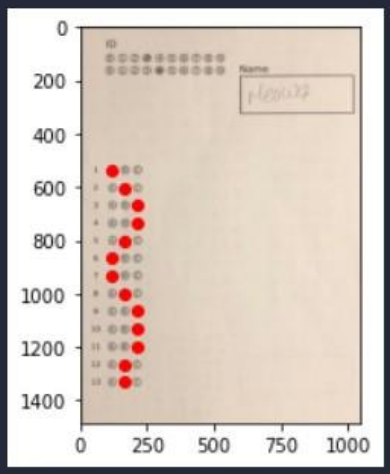
12 Ⓐ Ⓑ Ⓒ

13 Ⓐ Ⓑ Ⓒ

```
ID:  34
```




```
Id: 34
1) A
2) B
3) C
4) C
5) B
6) A
7) A
8) B
9) C
10) C
11) C
12) B
13) B
```



ID

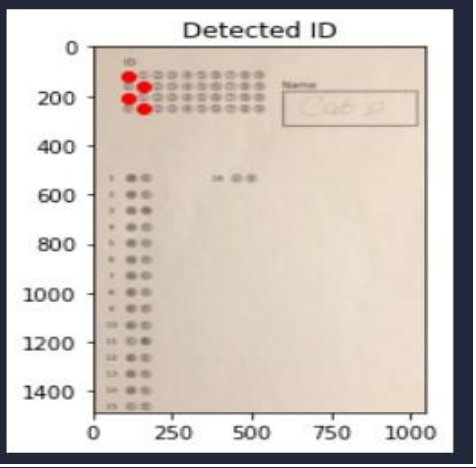
0 ① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨
0 ① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨
0 ① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨
0 ① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨

Name

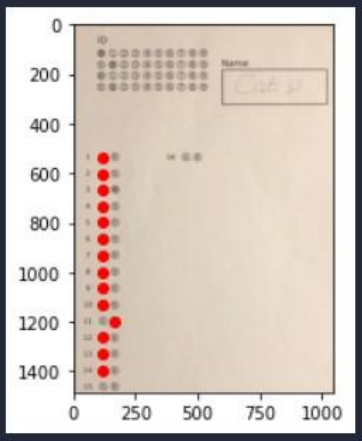
Cat p

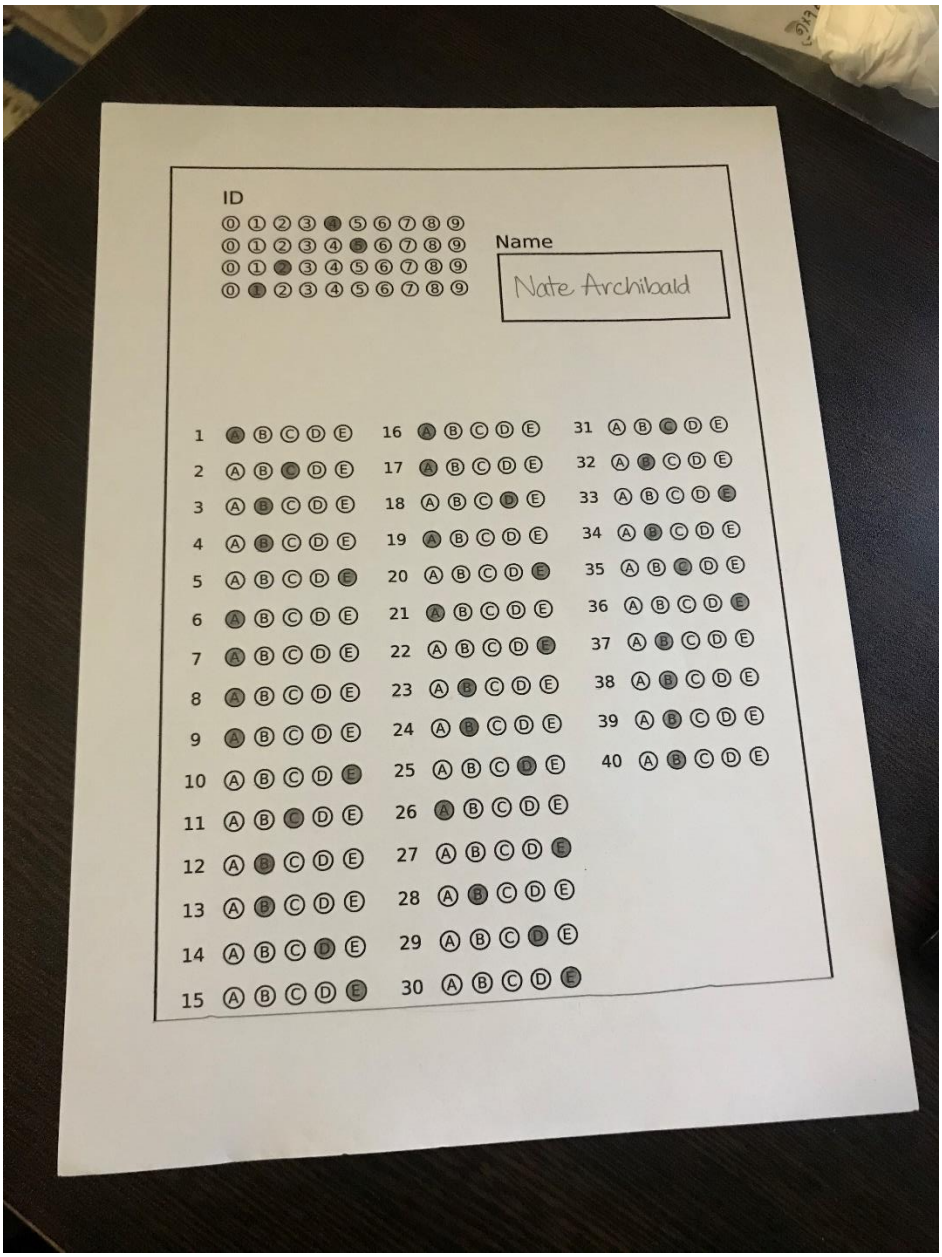
1	Ⓐ Ⓑ	16	Ⓐ Ⓑ
2	Ⓐ Ⓑ		
3	Ⓐ Ⓑ		
4	Ⓐ Ⓑ		
5	Ⓐ Ⓑ		
6	Ⓐ Ⓑ		
7	Ⓐ Ⓑ		
8	Ⓐ Ⓑ		
9	Ⓐ Ⓑ		
10	Ⓐ Ⓑ		
11	Ⓐ Ⓑ		
12	Ⓐ Ⓑ		
13	Ⓐ Ⓑ		
14	Ⓐ Ⓑ		
15	Ⓐ Ⓑ		

```
ID:  0101
```




```
Id: 0101
1) A
2) A
3) A
4) A
5) A
6) A
7) A
8) A
9) A
10) A
11) B
12) A
13) A
14) A
15) Invalid
16) Invalid
```





ID

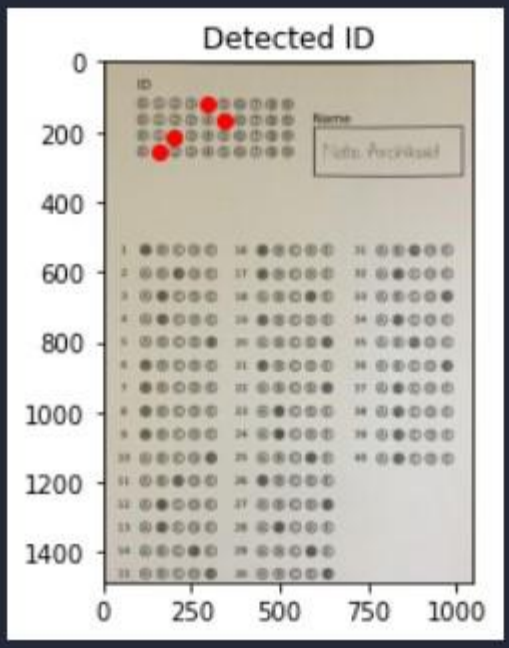
0 1 2 3 4 5 6 7 8 9
0 1 2 3 4 5 6 7 8 9
0 1 2 3 4 5 6 7 8 9
0 1 2 3 4 5 6 7 8 9

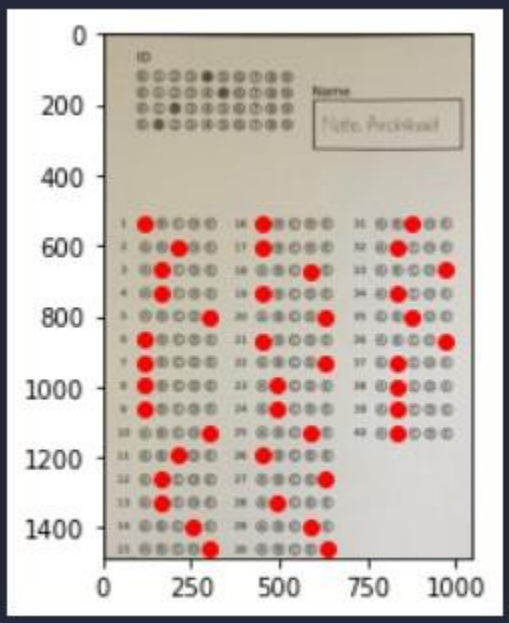
Name

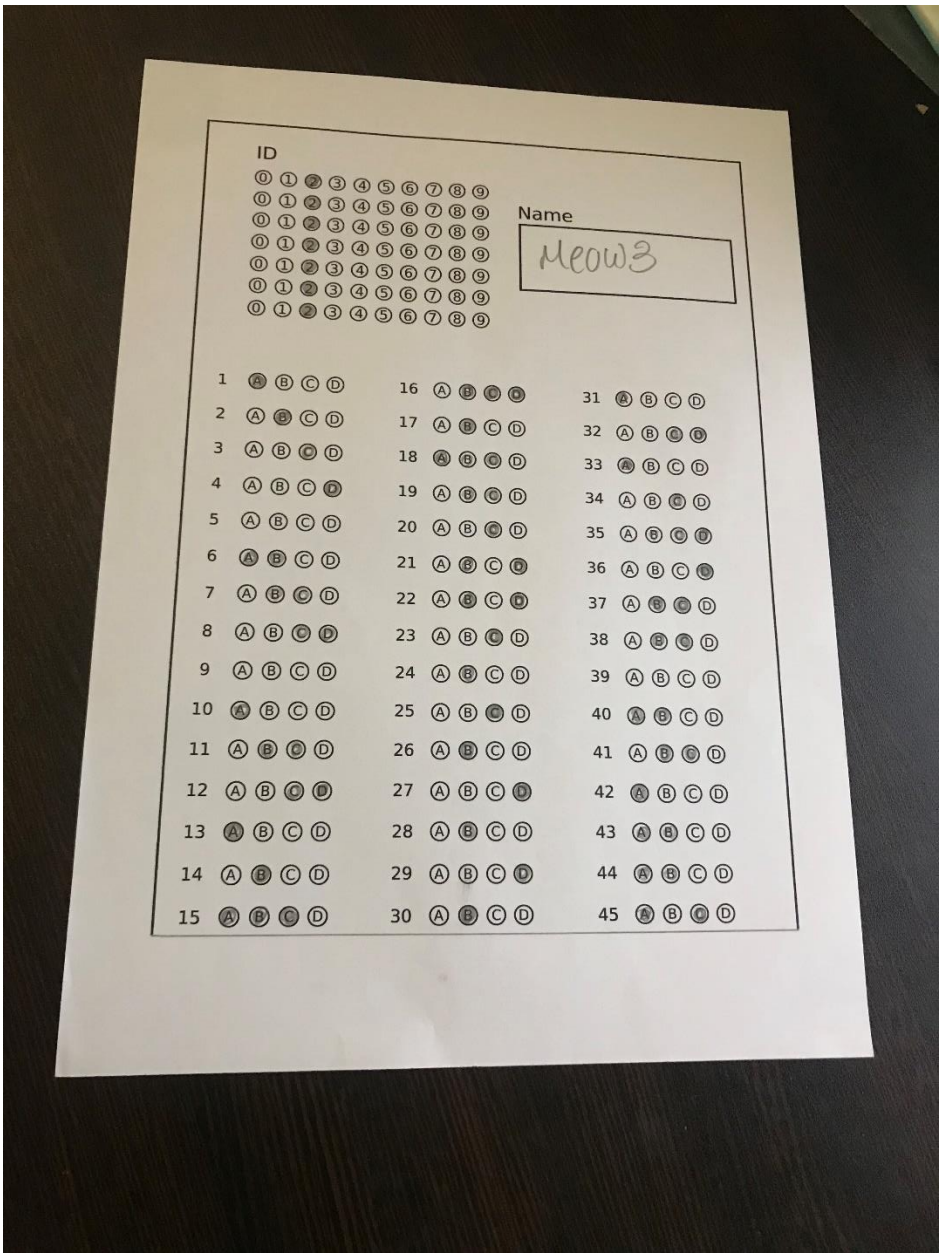
Nate Archibald

1	A	16	A	31	C
2	C	17	A	32	B
3	B	18	D	33	E
4	B	19	A	34	B
5	E	20	E	35	C
6	A	21	A	36	E
7	A	22	E	37	B
8	A	23	B	38	B
9	A	24	B	39	B
10	E	25	D	40	B
11	C	26	A		
12	B	27	E		
13	B	28	B		
14	D	29	D		
15	E	30	E		

```
ID:  4521
```


Detected ID
0
200
400
600
800
1000
1200
1400
0
250
500
750
1000


0
200
400
600
800
1000
1200
1400
0
250
500
750
1000



ID

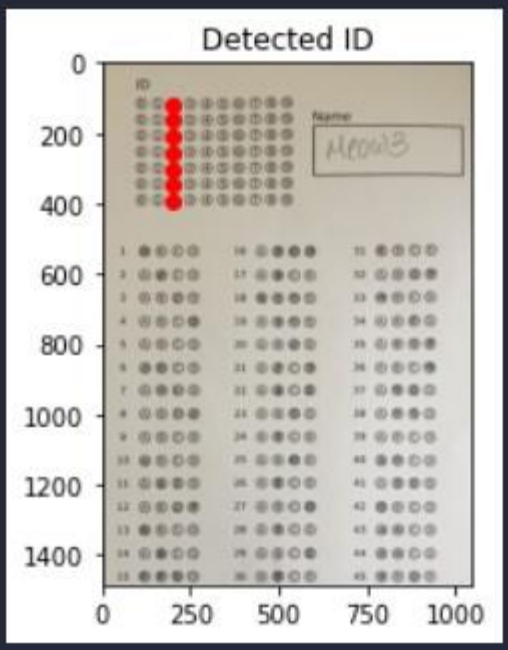
0 1 2 3 4 5 6 7 8 9
0 1 2 3 4 5 6 7 8 9
0 1 2 3 4 5 6 7 8 9
0 1 2 3 4 5 6 7 8 9
0 1 2 3 4 5 6 7 8 9
0 1 2 3 4 5 6 7 8 9
0 1 2 3 4 5 6 7 8 9

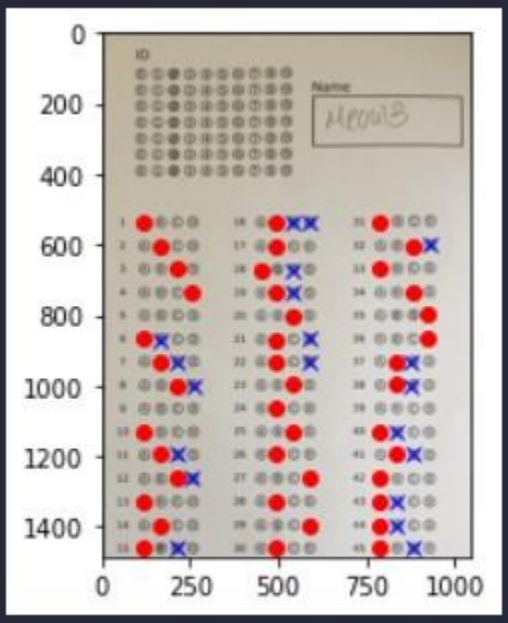
Name

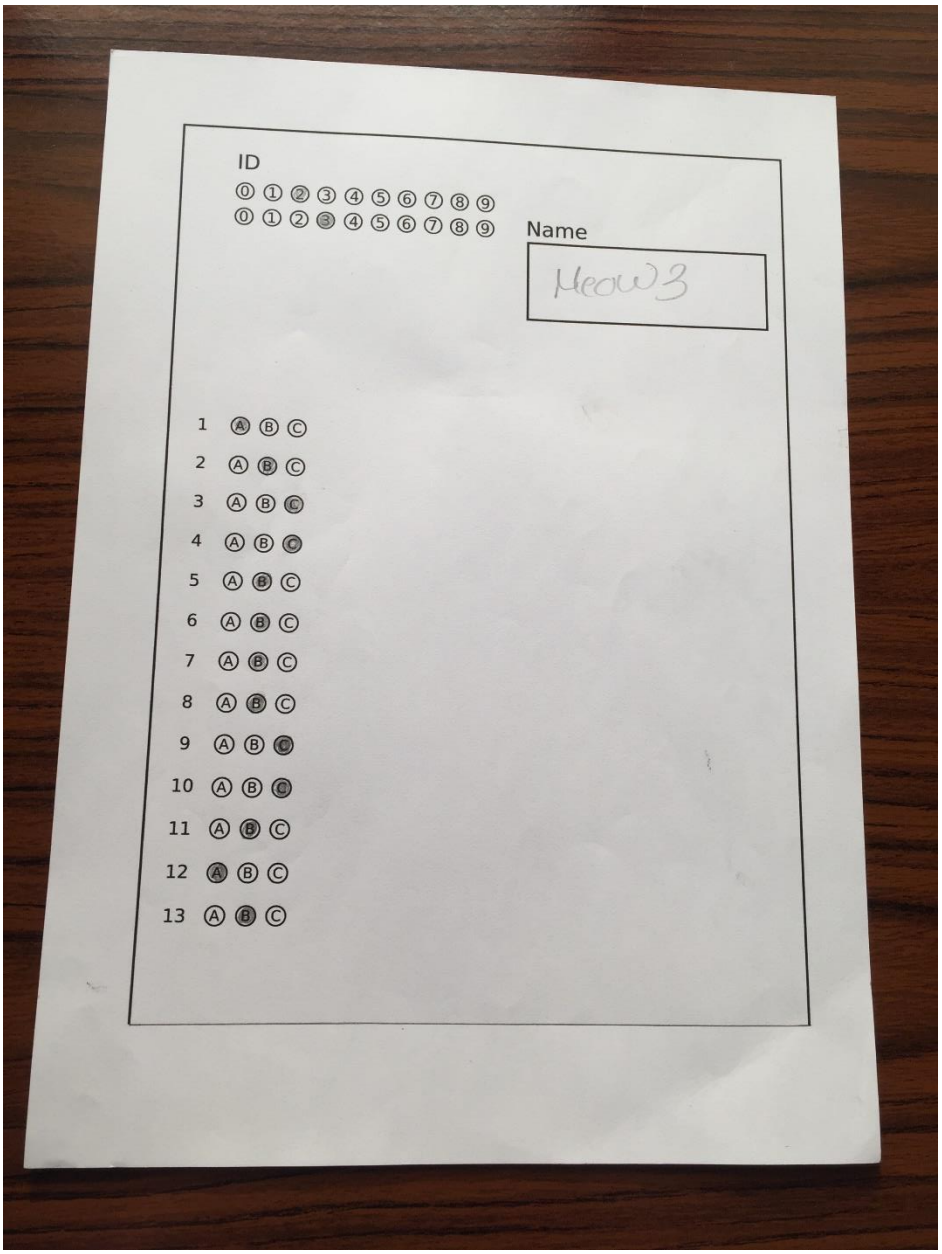
Meow3

1	A	16	A	31	A
2	B	17	B	32	B
3	C	18	C	33	C
4	D	19	D	34	D
5	A	20	A	35	A
6	B	21	B	36	B
7	C	22	C	37	C
8	D	23	D	38	D
9	A	24	A	39	A
10	B	25	B	40	B
11	C	26	C	41	C
12	D	27	D	42	D
13	A	28	A	43	A
14	B	29	B	44	B
15	C	30	C	45	C

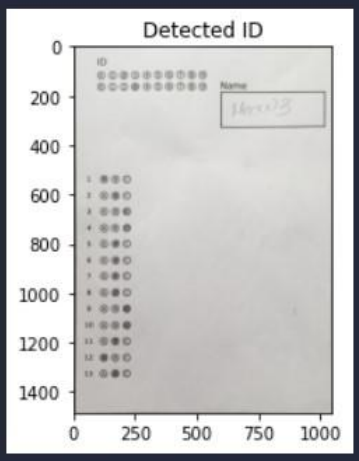
```
ID:  2222222
```


Detected ID
0
200
400
600
800
1000
1200
1400
0
250
500
750
1000

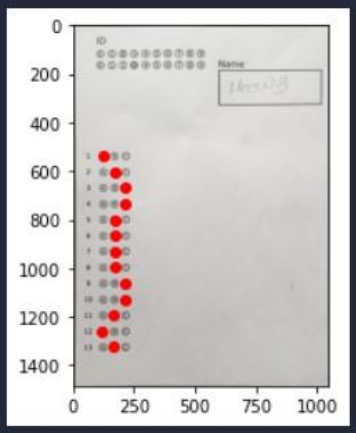

0
200
400
600
800
1000
1200
1400
0
250
500
750
1000


ID
0 1 2 3 4 5 6 7 8 9
0 1 2 3 4 5 6 7 8 9
Name
Meow3
1 A B C
2 A B C
3 A B C
4 A B C
5 A B C
6 A B C
7 A B C
8 A B C
9 A B C
10 A B C
11 A B C
12 A B C
13 A B C

```
Error: Some bubbles are not filled properly.
ID:  Error
```




```
Id: Error
1) A
2) B
3) C
4) C
5) B
6) B
7) B
8) B
9) C
10) C
11) B
12) A
13) B
```



Sample 1

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	ID	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Score
2															
3	ID_34	0	1	0	0	1	0	0	1	0	0	0	1	1	5
4															
5	ID_89	0	1	0	0	1	0	0	1	0	0	1	0	1	5
6															
7	ID_54	0	1	0	0	1	0	0	1	0	0	1	0	1	5
8															
9	ID_78	0	1	0	0	1	0	0	1	0	0	1	0	1	5
10															
11	ID_76	0	1	0	0	1	0	0	1	0	0	1	0	1	5
12															
13	ID_90	0	0	1	0	0	0	0	0	1	0	1	0	0	3
14															
15	ID_98	0	0	0	1	0	0	1	0	0	1	0	0	1	4
16															
17	ID_56	0	1	0	0	1	0	0	1	0	0	1	0	1	5
18															
19	ID_Error48	0	1	0	0	1	1	0	0	1	0	0	0	1	5
20															
21	ID_12	0	1	0	0	1	0	1	0	0	0	0	0	1	4
22															
23	ID_45	0	0	0	0	1	0	1	1	1	0	1	0	1	6
24															
25	ID_Error9	0	1	0	0	1	1	1	1	0	0	1	0	1	7
26															
27	ID_01	0	1	0	0	0	0	0	1	1	0	1	0	1	5
28															
29	ID_43	0	1	0	0	1	0	0	1	0	0	1	0	1	5
30															
31	ID_69	0	1	0	0	1	0	0	1	0	0	1	0	1	5

Sample 2

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	ID	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Q16	Score
2																		
3	ID_9998	0	1	0	1	0	1	0	0	0	1	0	0	0	1	0	1	6
4																		
5	ID_0101	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1
6																		
7	ID_2189	0	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	4
8																		
9	ID_5555	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1
10																		
11	ID_2222	0	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	14
12																		
13	ID_0887	0	1	0	1	0	0	0	0	0	0	0	0	0	0	1	0	3
14																		
15	ID_0991	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	2
16																		
17	ID_Error	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	0	7

Further samples are too wide to fit in the report.

Module 2: Table construction

Pros:

Our model works well with accepted accuracy and can substitute a human with some monitoring.

It can work under some sort of noisy or distorted image and with different lighting of captured image.

Can work under shadowed pictures.

Cons:

With some sort of highly distorted image the accuracy decreases as it can read or interprets the symbols wrongly.

Also curved pictures have some problems with accuracy.

Process:

First, we started with pre-processing the image which includes:

Binarization with gaussian adaptive thresholding.

Then finding contours to get the largest contour-the paper in our case- in other words in this step we remove the background from the image.

Second, we extract the table itself

By finding the contours in the processed image, contours will have the table shape.

Third, we extract the cells and find their position

We do this by searching for all vertical and horizontal lines using erosion then dilation(opening). Then we get the intersection between these two outputs.

Fourth, now we have the table and the cells positions,

We loop for every column and for every row to extract the data.

For horizontal and vertical lines we find contours and count this contour.

And we put their position in an array along with empty cells to not enter the knn prediction.

This part we used KNN algorithm, we made our own dataset that we took them from samples of sheets, cropped them and done some pre-process on them. After that we trained the knn –fit- to our desired symbols.

Now its time to predict the symbol in each cell, we do this by passing a cropped image (the cell) to the knn function then it predicts it and categorize it. Its accuracy is around 80%.

For names and ids we used OCR.

Finally we used pandas library to write the table in an excel sheet.

:

Fall-2021 Class list for Lecture Marketing (GENN326) Location : 20105 - 45 - الجيزة الرئيسي - odd Time : Tuesday(11:13)

Code	Student Name	English Name	1	2	3
1180236	احمد معتز لطفي احمد	Ahmed Motaaz Lotfy Ahmed	1	III	
1180333	حبيبة عصام حسب الله توفيق عمران	Habiba Essam Hassaballah Tawfik Omran	2	□	
1180128	سعد الدين محمد سعد محمد	Saad El-din Mohamed Saad Mohamed	3	?	
1180255	عبد الله محمد جلال السحيمي	AbdAllah Mohammed Galal El-Suhaimi	4	□	
1180274	علا ايمن عبدالفتاح المغربي	Ola ayman abdelftah elmaghraby	5	=	
1180056	علي شريف علي حسب الله	Ali Sharif Ali Hasb Allah	6	I	
1180041	عمر محمد فتحي شلقامي شعراوي	Omar Mohamed Fathy Shalkkamy Shaarawy	7	✓	
1180606	فاطمة عصام محمد جاب الله	Fatma Issam Mohamed Gaballah	8	✓	
1180456	فرح اسامه زين الدين محمد	farah ossama zein elden	9	✓	
2200022	محمد السيد احمد عبداللطيف	Mohamed Elsayed Ahmed Abdellatif	0	IIII	
1180552	محمد خالد محمد ابراهيم شمس	Mohamed Khaled Mohamed Ebrahim	0	≡	
1180207	محمد عبداللطيف محمد المصري	Mohammed Abdul Latif Mohamed El Masry	9	?	
1180045	مصطفى نضال مصطفى عزت حافظ	Mostafa Neddal Mostafa Ezzat Hafez	1	?	
1180212	منار سمير كامل ابوهواش	Manar Samir Kamel AboHawash	2	□	
1180155	منه الله هشام حسن احمد ابو النصر	Mennatallah Hisham Hassan Ahmed Abu-Nasr	3	□	
1170343	وليد علاء فتحي انيس	Waleed Alaa Fathy Anees	4	II	
1180172	ياسمين عمرو فاروق محمد رشاد	Yasmin Amr Farouk Mohamed Rashad	5	?	

Fall-2021 Class list for Lecture Marketing (GENN326) Location : 20105 - 45 - الجيزة الرئيسي - odd Time : Tuesday(11:13)

Code	Student Name	English Name	1	2	3
1180236	احمد معتز لطفى احمد	Ahmed Motaaz Lotfy Ahmed	2	✓	5
1180333	حبيبة عصام حسب الله توفيق عمران	Habiba Essam Hassaballah Tawfik Omran	3	✓	3
1180128	سعد الدين محمد سعد محمد	Saad El-din Mohamed Saad Mohamed	4	✓	3
1180255	عبد الله محمد جلال السحيمى	AbdAllah Mohammed Galal El-Suhaimi	5	2	1
1180274	علا ايمن عبدالفتاح المغربى	Ola ayman abdelftah elmaghraby	6	3	—
1180056	على شريف على حسب الله	Ali Sharif Ali Hasb Allah	7	□	□
1180041	عمر محمد فتحى شلقامى شعراوى	Omar Mohamed Fathy Shalkkamy Shaarawy	8	□	□
1180606	فاطمة عصام محمد جاب الله	Fatma Issam Mohamed Gaballah	9	?	✓
1180456	فرح اسامه زين الدين محمد	farah ossama zein elden	0	?	✓
2200022	محمد السيد احمد عبداللطيف	Mohamed Elsayed Ahmed Abdellatif	1	✓	✓
1180552	محمد خالد محمد ابراهيم شمس	Mohamed Khaled Mohamed Ebrahim	2	✓	?
1180207	محمد عبداللطيف محمد المصرى	Mohammed Abdul Latif Mohamed El Masry	5	3	?
1180045	مصطفى نضال مصطفى عزت حافظ	Mostafa Neddal Mostafa Ezzat Hafez	9	4	?
1180212	منار سمير كامل ابوهواش	Manar Samir Kamel AboHawash	6	?	□
1180155	منه الله هشام حسن احمد ابو النصر	Mennatallah Hisham Hassan Ahmed	0	?	□
1170343	وليد علاء فتحى انيس	Waleed Alaa Fathy Anees	4	□	□
1180172	ياسمين عمرو فاروق محمد رشاد	Yasmin Amr Farouk Mohamed Rashad	3	□	5

```
6 18
[0.] [0.]
[0.] [0.]
[0.] [0.]
[1.] [1.]
[1.] [1.]
[0.] [0.]
[0.] [0.]
[1.] [1.]
[1.] [1.]
[1.] [1.]
[1.] [1.]
[0.] [0.]
[0.] [0.]
[0.] [0.]
[2.] [2.]
[1.] [1.]
[1.] [1.]
[1.] [1.]
[['Q1' 5 5 5 3 3 0 0 4 4 5 5 2 4 4 4 0 0]
 ['Q2' 5 2 2 1 4 0 0 5 5 5 1 4 'R' 0 0 0 5]]
```

```
6 18
[1.] [1.]
[1.] [1.]
[0.] [0.]
[0.] [0.]
[0.] [0.]
[2.] [2.]
[0.] [0.]
[1.] [1.]
[1.] [1.]
[2.] [2.]
[['Q1' 3 0 1 0 3 1 5 5 5 4 2 'R' 5 0 0 1 'R']
 ['Q2' '' '' '' '' '' '' '' '' '' '' '' '' '' '' '' '' '']]
```