

Medicine Identification Application for Visually Impaired People

Angel Negi¹, Aishwarya Bhure², Dhanashree Patil³, Akshita Maskara⁴, Madhuri Bhalekar⁵

School of Computer Engineering & Technology, MIT World Peace University, Pune, Maharashtra, India

angelpnegi@gmail.com¹ aishwaryabhure99@gmail.com² dhanup6068@gmail.com³
akshitamaskara@gmail.com⁴ madhuri.bhalekar@mitwpu.edu.in⁵

Abstract: Reading is a fundamental part of our lives. Information and text is present everywhere in magazines, in newspapers, in receipts, in letters, bank documents etc. But text present on medicine and its cover is something which is important and needs to be crosschecked every time we buy it. While buying medicines, people tend to make sure that they are buying a correct one by cross checking the medicine name, expiry date, manufacturing date and the MRP so that they do not get fooled. It is one of the easiest tasks to perform for normal people as they can read. But what about the people with visual impairment? How will they be able to get their medicine details?

The proposed application tries to imitate this job for them and helps them better understand their medicines. The application scans the medicine content side and uses the advanced technologies like Tesseract optical character recognition (OCR), Computer Vision, Text to Speech technology to extract the required information and makes it available to the user in the form of audio. The application helps visually impaired people to independently carry out this task without the hindrance of a third person.

Keywords: OCR (Optical Character Recognition), Computer Vision, Text-to-Speech, Tesseract OCR.

1. INTRODUCTION

According to the survey conducted by World Health Organization (WHO) in 2010, 1181.4 million is the total population count of India and out of which 152.238 million are the people who suffer with blindness, low vision, and visual impairment. As an individual loses vision, comes the loss of independence. Most of the time, they need a third person to assist them. When it comes to written text, they can only come to know about it through audio.

Today's world is the world of digitalization and technology is advancing rapidly. Artificial Intelligence, Computer vision, etc are some of the advanced technologies used today which can be really useful to build applications that can extract the text which then can be converted into an audio using text to speech technology. This can be used on the medicines and get the relevant text extracted followed by an audio output.

Optical Character Recognition (OCR) is the most widely used technology for text recognition and extraction. It was created back in 1914 to help people with visual impairment to have written text read out for them. Since then, tremendous advancements have been made in this technology and several issues like speed, accuracy, size, etc have been improved.

The proposed application makes use of Tesseract OCR which is freely available on internet these days to recognise the text and pytsx3 python text to speech library for converting it into an audio output.

II.LITERATURE SURVEY

In [1], author Snigdha presents a system that is able to detect text from any documents including medical reports and recognizes the medicine that can assist the visually impaired without the hindrance of a third person and motivates them to live long. The system helps these people in terms of mobility and can also verify the medicine details. The proposed application is implemented using a smartphone and an android application. Users can start the application by shaking the phone. Information is stored in the MySQL database through the TOMCAT server. After scanning, the text detector detects the medicine and passes it on to the server (TOMCAT). It checks and matches the medicines and also at what time to be taken. At the right time to take the medicine, the server sends the notification and finally after receiving the information, it is converted into speech.

In [2], author Shirly Edward proposes a text to speech device for the visually impaired which involves Raspberry Pi. The system architecture consists of two main modules: Image correction module and voice processing module. Once the Raspberry Pi's camera captures the image, as a part of image correction, gray scaling and binarization is performed. Then the image processing module makes use of tesseract OCR and targets the text. The user captures the image via GPIO pin. Once the text is extracted, TTS correction and voice module converts it into speech. Flite and espeak which are supported by Raspberry Pi are used for this.

In [3], the authors proposed an application for the visually impaired. It provides complete assistance in medicine intake scenarios using label reading. The system even has the ability to remind the patients when to take medicines. Using visual features, medicine boxes are detected. Edge detection, colour reduction are some of the techniques used. Each medicine is registered with a sound file and an image before the application starts. And finally when they scan the medicine box using a camera, it tells them whether they have picked the correct one or not.

In [4], Authors Samruddhi and Revati propose a real time hand held object detection system to assist the visually impaired. The proposed system reads the text from a camera captured image. The framework comprises three sections namely - scene capture, data processing and audio output. The data processing unit tries to find the text patterns. The MESR algorithm is used to detect the blob in the images. After OCR recognises the text, Microsoft speech software development kit plays the audio output. The paper points out that high performance can be achieved using MESR and OCR for text detection and recognition.

In [5], authors propose a medical assistive system for identification of medicines by the visually impaired. The system works in two parts: 1) Medicine pill and 2) Medicine box.

Medicine pills are identified based on their features. Sobel edge detector morphological operations are used for the same. Once a person picks a pill from the box, its label is extracted using OCR and is matched with the templates that are available on the database. Then TTS technology will communicate with the person to identify whether he has picked the correct one or not.

III.SYSTEM ARCHITECTURE

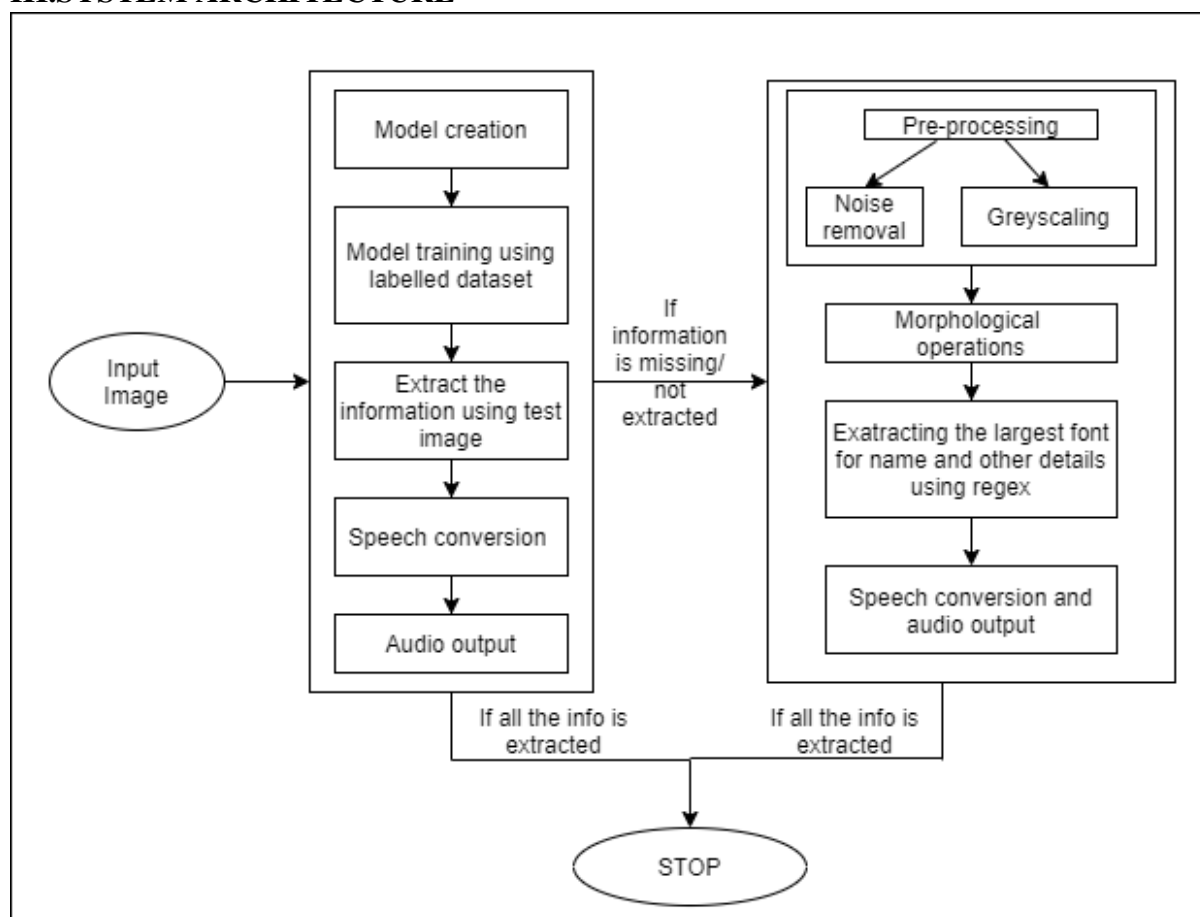


Figure 1: Detailed system architecture

Figure 1 gives a detailed glimpse of the proposed system architecture. The architecture comprises several components/modules which are explained in detail in the methodology section.

III.METHODOLOGY

3.1 Dataset Creation:

The dataset consists of 250 images of medicine strips containing all the required information. The size, shape, text, font color of every medicine image is different. The dataset has been labelled marking four fields as name, manufacturing_date, expiry_date and MRP. Figure 2 gives a glimpse of labelled images.



Figure 2: Labelled Images

3.2 Model Creation:

The model used in the project is created using an API called Nanonets. Model has two major building blocks:

- Text detection
- Text recognition

Text Detection:

YOLO is a Single Shot Detector, which means that it predicts both the boundary box and the class at the same time. When dealing with speed and accuracy there’s a trade-off. YOLO is fast but since it detects two things at the same time, it has comparatively lower accuracy rate than Regional-Based Detector.

Abbreviated for You Look Only Once, YOLO is true to its name. It is a real time object detection network which uses Darknet-53 as it’s feature extractor.

The first step is to collect the data and augment them if required. With a sufficient amount of data, images have to be annotated which infers that boxes are to be created on the image and each box is given the class it belongs to.

Training of the data is done by darknet-53 which can be cloned using the command git clone <https://github.com/pjreddie/darknet.git>. As it runs the iteration, the one with maximum mAP (Mean Average Precision) score is to be used further to extract the data from the images.

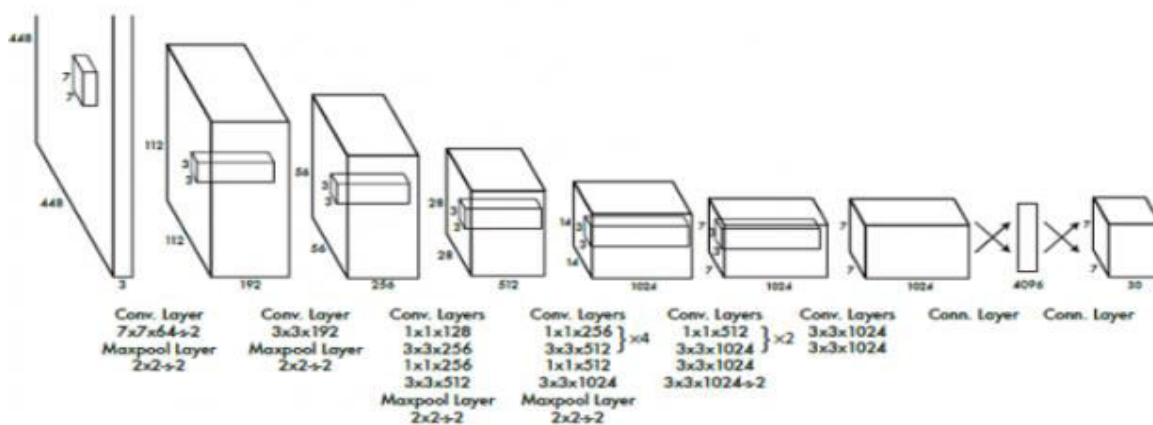


Figure 3: YOLO Architecture ⁶

Text Recognition:

For the process of text recognition, Nanonets uses Tesseract 4 OCR. Once an image is passed into YOLO. It detects the required text regions and crops them out from the image. Later, those regions are passed to tesseract one by one. Tesseract reads them, and information is stored.

If in case, the model fails to extract all the required information or some part of it, the image is processed using a step by step methodology explained in 3.3.

3.3 Following is the step-by-step methodology used:**Step 1: Image input**

Initially feed a system with an image. The inputted medicine image must be as clear as possible so that it gets easy for the tesseract OCR to extract text. Basically, clearer the image, better the output.

Step 2: Image Pre-processing

Image pre-processing involves performing some basic operations like: Noise removal and grey scaling the image. For best OCR results, the image should be as sharp as possible. Like the edges, correct skew angle, sharp character edges, no noise presence, etc. Noise in images is the random discrepancy in brightness or colour information. Hence, noise removal is the first thing that is performed as a part of image pre-processing.

Step 3: Performing Morphological operations

After step2, two morphological operations opening and closing are applied to make images sharp. Both the operations involve two basic operations: dilation and erosion. Opening is erosion followed by dilation. It removes any narrow connections and lines between two regions. It helps in removing small objects from an image but at the same time preserves the shape and size of larger objects in the image. It makes the image thin. It is represented by the formula:

$$(A \circ B) = (A \ominus B) \oplus B$$

Closing is the reverse of opening that is dilation followed by erosion. It fills up the narrow black holes or regions in an image at the same time it preserves the shape and size of the objects in an image. It makes the image thick. It is represented by the formula:

$$(A \circ B) = (A \oplus B) \ominus B$$

Step 4: Extracting texts

Tesseract OCR is used to extract all the contents of the content side. To extract the name, function is used to extract the text with the biggest font as the font size of the medicine label is bold and large in most of the cases. For extracting other information like the dates and MRP various regexes are applied. The final output obtained is in dictionary format.

Step 5: Final Output

The text that is generated in the previous step in the dictionary form is now fed to the pyttsx3 (python text to speech) library for further converting it into an audio output. The generated output is in the form name is (name of the medicine), MRP is (Rs. Amount) etc.

IV. RESULT AND ANALYSIS

The dataset is divided into two parts: training and testing. 230 images have been labelled and the model is trained. Rest of the unlabelled images are used for testing the system. The training accuracy of the model achieved is 75.55%. Figure 3 and figure 4 depicts the training progress and confusion matrix for the same.

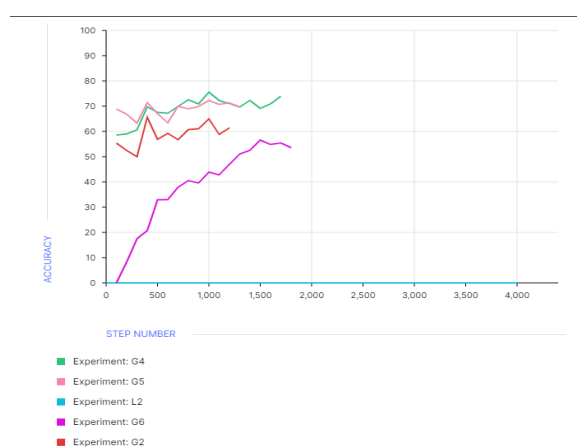


Figure 4: Training progress graph

Field level metrics

Precision: **0.76** Recall: **0.76**F1 Score: **0.76**

| | Predictions: Incorrect | Predictions: Correct |
|-------------------|------------------------|----------------------|
| Actual: Incorrect | 1 | 5 |
| Actual: Correct | 5 | 16 |

Figure 5: Confusion matrix

The figure below is the output in the form of dictionary after extraction of all the relevant details required this dictionary is then converted into audio output for the visually impaired people to hear.

```
{'Name': "'Dapagliflozin', 'M R P': 142.0, 'Manufacturing Date': 'november 2020', 'Expiry Date': 'september 2022'}
done
```

Figure 14: Dictionary Output

The model when subjected to new images, in this case 25 test images, gave an overall accuracy of 74% and individual parameter accuracies are as shown below in figure 14.

Name: Name was successfully extracted from 15 images from a total of 25 images i.e accuracy 60%

MRP: Mrp was successfully extracted from 21 images from a total of 25 images i.e accuracy 84%

MFG: Mfg was successfully extracted from 22 images from a total of 25 images i.e accuracy 88%

EXP: Mfg was successfully extracted from 16 images from a total of 25 images i.e accuracy 64%

Overall accuracy considering all parameters achieved is 74%.

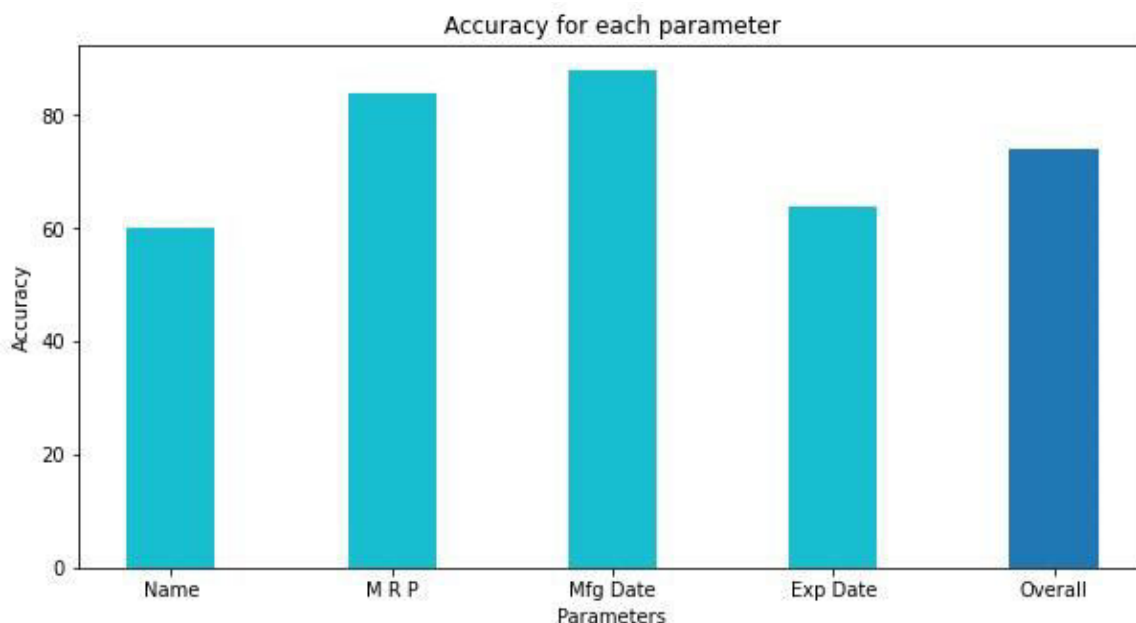


Figure 15: Accuracy percentage

Performance of the logic on all test images is shown through figure 7, which gives the number of parameters model could extract for every image.

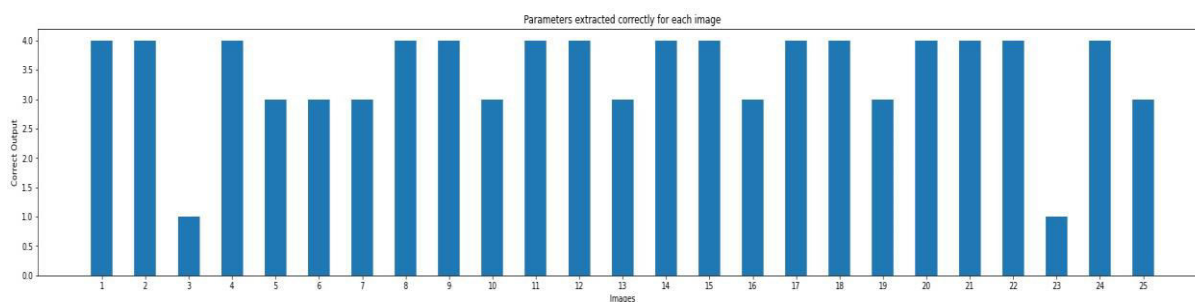


Figure 16: Performance on every image

V.CONCLUSION AND FUTURE SCOPE

The developed system is successfully able to extract the required labels of the medicines: name, price, manufacturing date and expiry date. The model is built using nanonet api. If in case the model fails to extract any of the given information, various logics and techniques such as regex, morphological operations etc are used. The accuracy achieved is ____.

As a part of future study, the system accuracy can be increased by increasing the size of the database. The larger the dataset, the better the accuracy. Some preprocessing techniques other than the explored techniques like opening, closing, thresholding can also be studied for proper text extraction. The scope can further be extended to medicine bottles.

REFERENCES

1. Snigdha Kesh, Ananthanagu U, 2017, Text Recognition and Medicine Identification by Visually Impaired People, *International Journal of Engineering Research & Technology (IJERT) ICPCN – 2017* (Volume 5 – Issue 19)
2. Edward, Shirly. (2018). Text-to-Speech Device for Visually Impaired People. *International Journal of Pure and Applied Mathematics*. 119
3. Dhanajay Chaudhari (2018) Voice Based Application as Medicine Spotter For Visually Impaired *International Research Journal of Engineering and Technology (IRJET)*.
4. S. Deshpande and R. Shriram, "Real time text detection and recognition on hand held objects to assist blind people," *2016 International Conference on Automatic Control and Dynamic Optimization Techniques (ICACDOT)*, 2016
5. Vasavi, S., Swaroop, P.R.S. & Srinivas, R. Medical assistive system for automatic identification of prescribed medicines by visually challenged from the medicine box using invariant feature extraction. *J Ambient Intell Human Comput* (2019)
6. Joseph Redmon, Santosh Divvala, Ross Girshick & Ali Farhadi. You Only Look Once: Unified, Real-Time Object Detection.