A Project Report

on

"Smart Electronic Health Records using OCR"

Submitted to the

Savitribai Phule Pune University

In partial fulfillment for the award of the Degree of

Bachelor of Engineering

in

Information Technology

by

Alam Kathat (Seat No.: B150228507)

Aman Gupta (Seat No.: B150228510)

Atul Kumar Tiwari (Seat No.: B150228520)

Sanju Baloria (Seat No.: B150228552)

Under the guidance of

Prof. Yuvraj Gholap

Department of Information Technology

Army Institute of Technology

Dighi Hills, Pune-Alandi Road,

Pune 411015.

2021-2022

CERTIFICATE

This is to certify that the project based seminar report entitled "Smart Electronic Health Records using OCR" being submitted by Alam Kathat (Seat No: B150228507), Aman Gupta (Seat No: B150228510), Atul Kumar Tiwari (Seat No: B150228520), Sanju Baloria (Seat No: B150228552) is a record of bonafide work carried out by him/her under the supervision and guidance of Prof. Yuvraj Gholap in partial fulfillment of the requirement for **BE (Information Technology Engineering) - 2015 Course** of Savitribai Phule Pune University, Pune in the academic year 2021-2022.

Date: 09/05/2022

Place: Pune

Prof. Yuvraj Gholap
Seminar Guide

Dr Sangeeta Jadhav
Head of the Department

Dr. B. P. Patil
Principal

External Examiner :

ACKNOWLEDGMENT

I am highly indebted to my guide Prof. Yuvraj Gholap for his guidance and constant supervision as well as for providing necessary information regarding the seminar report and also for her support in completing the seminar report. I would like to express my special gratitude and thanks to Project Coordinator Prof. Gajanan Walunjkar for giving me such attention and time.

This acknowledgment would be incomplete without expressing our thanks to Prof. Sangeeta Jadhav, Head of the Department (Information Technology) for her support during the work.

We would like to extend our heartfelt gratitude to our Principal, Dr.B. P. Patil who provided a lot of valuable support, mostly being behind the veils of college bureaucracy.

We would also like to thank and appreciate team Veritas and specially Mr. Vaibhav joshi in showing keen interest in our project and helping us to build a Production level project .

We would also like to express our gratitude towards our parents and friends for their kind co-operation and encouragement which helped us in completion of this report. Our thanks and appreciations also go to our colleagues in developing the Project report and people who have willingly helped us out with their abilities.

(Alam Kathat) (Aman Gupta)

(Atul Kumar Tiwari) (Sanju Baloria)

Abstract

Healthcare is ever developing field focusing on creating more effective and efficient process of treatment. In era of smart technology, Machine learning and big data analytic are playing huge role in boosting the growth of health-care system. But it faces a big problem of data poor quality and quantity. India healthcare system mainly depend on paper-based prescription and report, it does not have needed data for better implementation and precision in machine learning technology. Our Paper focus on developing machine learning algorithm to create digital record from old paper-based record using Convolutional Recurrent neural network algorithm for Image processing and creating new record using voice to text conversion using Hidden Markov Model. The record created could further be analyzed to created better quality data. The newly created data could increase scope and precision of healthcare machine learning models.

Contents

Contents

List of Figures

List of Tables

Chapter 1

INTRODUCTION TO DIGIHEALTH

1.1 Introduction to Project

Since Data has been most important part of healthcare it boosts the growth and development of the field. Need of data is more than ever before in the world of advancing technology. Since ages data in healthcare is in form of prescription and reports given by doctor to patient mostly vocally or hand-written on fragile object like paper. By digitalizing the data hospitals can easily maintain the files and records of patients. The aim of this project is to create digital record from old paper-based record using Convolutional Recurrent neural network algorithm for Image processing and store in the private and secure database. The record created could further be analyzed to created better quality data like it can be used in analysing the disease and the effect of particular medicine on a particular disease and like side effects of a particular medicine etc.

1.2 Motivation behind project topic

As we usually see that most the times the doctor use to write the observation of the patients diseases on the paper which is sometimes lost by the patient due to this the data of the observation is lost.So Due to this reason we decided to create a Digital Record System which stores the record digitally and due to this there is zero chance of data lost.

1.3 Aim of the Project

The aim of this project is to create a user friendly system which record the observation and then convert that recorded observation into to the text by using some machine learning algorithms and then store that data into private and secure database.

1.4 Objectives of the work

Objective of project is to create digital prescription from paper based prescription using Optical character recognition for specialized treatment and further research purpose. Also to create User interface for easy and fast access of digital prescription to user. Data security is also a major aspect of project to protect personal information and prescription of user.

Chapter 2

LITERATURE SURVEY

2.1 RELATED WORK

The early related work in storing medical data in electronic format includes both OCR and voice to text conversion. The OCR was based mainly upon two neural networks:-

- Convolutional Neural Network
- Recurrent Neural Network.

But the problem with both neural networks was that of accuracy. In case of printed text either of model is capable of giving accurate results but here we are dealing with medical data with doctors handwriting, the accuracy reduced with either of them. As we are dealing in medical data, accuracy needs to be given utmost importance.

To increase the accuracy and to take benefits of both the neural networks in paper [1] combination of both the neural networks is used. CRNN was used in paper [1]. An image is fed as input to the CNN and after passing it through various layers a character is identified. RNN is used to handle sequential data because it considers current input as well as past inputs, as it memorizes previous inputs due to its internal memory. RNNs utilises different data points in a sequence to generate more accurate results.

The proposed CRNN consists of 13 convolutional layers accompanied by 3 Bidirectional Long Short Term Memory (LSTM) layers. ReLU activation function is applied to the convolutional layer to get a rectified feature map of the image. The images are inputted and splitted into windows or sub-images, keeping height of sub-image equal to that of text line image. From the feature maps generated by the last convolutional layer, the vectors are extracted and

passed on to the LSTM layer. Then outputs of both forward and backward bidirectional LSTM layers is combined together, instead of doing it at the end of each layer.Due to this weights are also optimized faster.

2.2 APPROACHES

E-prescription is a computer-generated prescription which could be filled, transferred and stored digitally.

E-prescription is important aspect of digitalization of Healthcare and solve many existing problem in Indian healthcare system such as:

- Secure and Remote access of prescription in case of emergency.
- Digital authentication of prescription to impose drug regulation.
- For future ML analysis on report for better treatment and educational purpose.
- Increase convenience for patient, pharmacist and doctor

There are many ways to create a E-Prescription.

- Through digital creation It most used technology for creation of e prescription. But it is newer technology and newer prescription can be created but older prescription canât be digitalized by the user. There exist no medical history for user from before enrolment date in application.
- Through Scan storage
  - Storing scan files of prescription is possible for user and application, but scan storage cost large amount of space in system due to high amount and quality of file which could increase cost of storage of prescription.
  - A Scan file not always better since a unreadable image will not be helpful as well as cant be used for further medical analysis and research purposes.
- Through Image OCR (Selected Solution)
  - Storing OCR output of the prescription image is very space efficient and user friendly. A OCR output is in text format so can be understandable by common user and can be stored with minimum storage requirement.
  - OCR output is also computer understandable and can be used in further

2.3 STATE OF THE ART ALGORITHMS

2.3.1 Convolutional Neural Network

Convolution Neural Networks or covnets is an algorithm which can take in an input image, assign importance (learnable weights and biases) to various aspects/objects in the image and be able to differentiate one from the other.

A convolutional neural network can have tens or hundreds of layers that each learn to detect different features of an image. Filters are applied to each training image at different resolutions, and the output of each convolved image is used as the input to the next layer. The filters can start as very simple features, such as brightness and edges, and increase in complexity to features that uniquely define the object.

Layers used to build ConvNets A covnets is a sequence of layers, and every layer transforms one volume to another through a differentiable function. Types of layers: Layers functionality is explained by taking an example of running a covnets on an image of dimension 32 x 32 x 3.

- Input Layer: This layer holds the raw input of the image with width 32, height 32, and depth 3.
- Convolution Layer: This layer computes the output volume by computing the dot product between all filters and image patches. Suppose we use a total of 12 filters for this layer weâll get output volume of dimension 32 x 32 x 12.
- Activation Function Layer: This layer will apply an element-wise activation function to the output of the convolution layer. The volume remains unchanged hence output volume will have dimension 32 x 32 x 12.
- Pool Layer: This layer is periodically inserted in the covnets and its main function is to reduce the size of volume which makes the computation fast reduces memory and also prevents overfitting. Two common types of pooling layers are max pooling and average pooling. If we use a max pool with 2 x 2 filters and stride 2, the resultant volume will be of dimension 16x16x12.

2.3.2 Recurrent Neural Network

Recurrent Neural Network(RNN) are a type of Neural Network where the output from previous step are fed as input to the current step. In traditional neural networks, all the inputs and outputs are independent of each other,

but in cases like when it is required to predict the next word of a sentence, the previous words are required and hence there is a need to remember the previous words. Thus RNN came into existence, which solved this issue with the help of a Hidden Layer. The main and most important feature of RNN is Hidden state, which remembers some information about a sequence.

RNN have a âmemoryâ which remembers all information about what has been calculated. It uses the same parameters for each input as it performs the same task on all the inputs or hidden layers to produce the output. This reduces the complexity of parameters, unlike other neural networks.

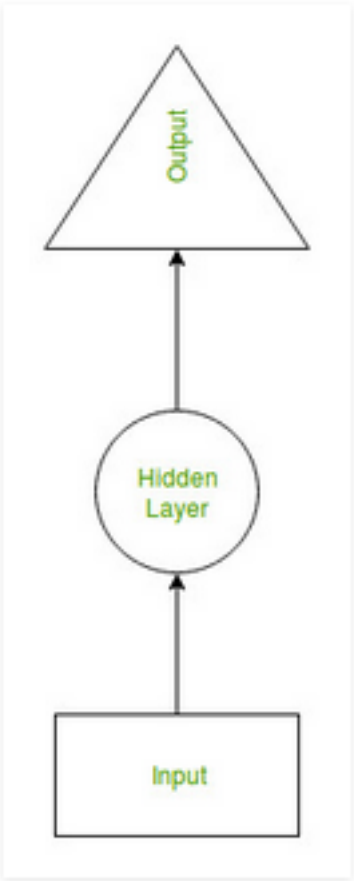


Figure 2.1: Recurrent Neural Network

2.3.3 Convolutional Recurrent Neural Network

Convolutional Neural Network Recurrent Neural Network together constitute this network model, hence having benefits of both network models. The simple feed forward neural network only considers current input and cannot memorise past inputs, so it cannot process the sequential data. In CNN pixels of image is fed as the input and then it is fed to input layer which accepts the input in the form of arrays. After this it is passed on to hidden layer which carry out feature extraction. Convolutional, ReLU and Pooling layers are important hidden layers. At last there is a fully connected layer that recognises the object in the image. The RNN can process sequential data as it takes into account current input as well as past inputs (Due to added feature of internal memory it can remember its past inputs). RNNs utilises different data points in a sequence to generate more accurate results. They basically takes inputs and reuse the activations of previous nodes or subsequent nodes in the sequence to get the output.

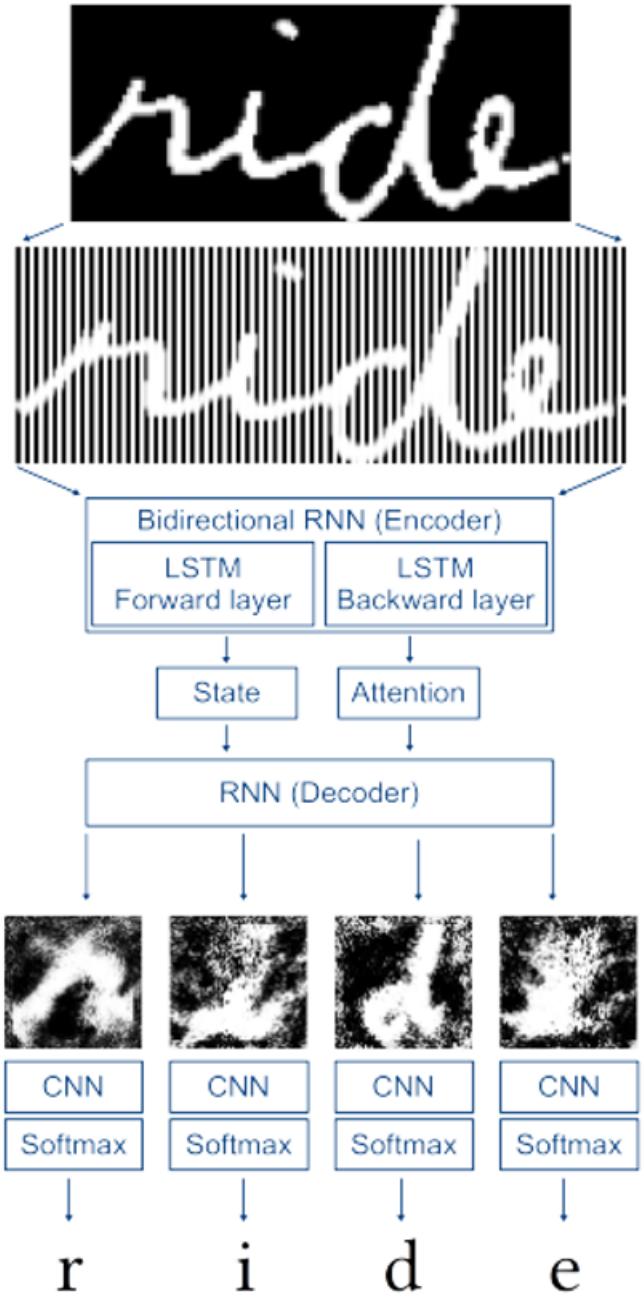


Figure 2.2: Convolutional Recurrent Neural Network

2.4 ANALYSIS

SNo.	Title of Paper	Methodology	Year
1	Studying Hospital Process Management System Concerning Information Quality	About Data Quality in field of healthcare and its effect.	2009
2	Semantic Medical Prescriptions â Towards Intelligent and Interoperable Medical Prescriptions	About benefits of E-prescription and how can it help in increasing medication effectiviness.	2013
3	Smart Analytics And Predictions For Indian Medicare	About analysising and diagnosing using ML based on previous records.	2018
4	Big Data Security in Healthcare	Discuss Advantage and concern about Digital healthcare system.	2018
5	Knowledge Discovery, Analysis And Prediction in Healthcare using Data Mining And Analytics	Suggesting alternative medication and analysing data to improve healthcare.	2016
6	Applying Internet of Things and Machine-Learning for Personalized Healthcare: Issues and Challenges	Creating personalized medication plan based on daily routine and habits.	2019
7	Optical Character Recognition for English Handwritten Text Using Recurrent Neural Network	OCR using Recurrent Neural Network	2020
8	Medical Handwritten Prescription Recognition Using CRNN	OCR using RNN and CNN both.	2019
9	A Proposed Framework for Recognition of Handwritten Cursive English Characters using DAG-CNN	Recognition of handwritten cursive handwriting using CNN.	2019
10	High-Performance OCR for Printed English and Fraktur using LSTM Networks	Using LSTMs, recognition of handwriting.	2013
11	An Algorithmic Approach for Text Recognition from Printed/Typed Text Images	Recognition of handwritten cursive handwriting using CNN.	2018

Table 2.1: Analysis of various works.

Chapter 3

PROBLEM STATEMENT

Medical data is very important for analysis using ML algorithms to predict any outbreak of disease such as Covid-19, etc. But in Indian healthcare system lot of data is present in papers which cannot be used for analysis leading to wastage of data. Moreover patients also face problems in conserving those medical slips. Problem statement is to bring this data in electronic format and create a centralized healthcare for easy access of medical data to both patients and doctors.

Chapter 4

REQUIREMENT SPECIFICATIONS

4.0.1 Web Technologies

The technologies we are going to use to build the user interface and to connect our models are as follows:

- HTML,CSS
- React JS Framework
- Django

4.0.2 Libraries/Framework/Database

Some resources are going to be used to manipulate the data input.

- Anaconda/Jupyter Notebook
- PyTesseract
- Numpy
- TensorFlow
- OpenCV
- Pandas
- Matplotlib
- Some additional Python Library

Chapter 5

SYSTEM PROPOSED ARCHITECTURE

Our Goal of the Digihealth to create textual format output from scan or capture image of prescription to server the need of digitizing the medical history of patient from paper-based prescription to easily readable and reachable format. The Created output could be used further medical purposes ranging from specialized treatment for patient to Further study of disease which can be helpful patient as well medical field. With advancement of medical field the data can be used for research purpose as well as form create Artificial Intelligent doctor to assist doctor.

Basic Architecture Proposed consist of collecting or receiving medical prescription from patient as well as doctor in scanned or Image capture format. After the receiving the image it goes through many layer of Machine Learning models to remove noise and preprocess the image. After preprocessing the image further goes to main Combined model based on CRNN and Tesseract to get the OCR output in textual format. The output further goes to dictionary for word completion and grammatical correction .Then the output is stored in Database with many other details stored regarding the prescription and patient. The Digital prescription can be accessed by doctors and patient for better treatment and used in Medical Advancement.

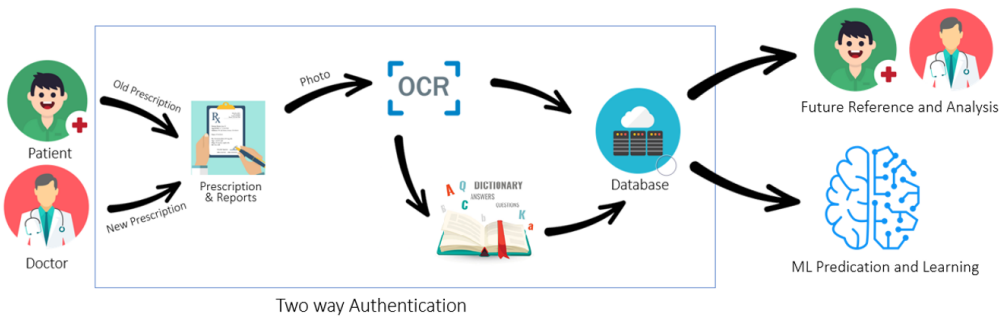


Figure 5.1: Basic Architecture

Chapter 6

HIGH LEVEL DESIGN OF THE PROJECT

6.1 Use Case Diagram

The proposed model of the project.Given an input of the prescription and then the data is get extracted from the image and then goes through the various machine learning operation like preprocessing segmentation,feature extraction,Neural network.At the end the extracted data after applying various operations passes to the database with the help of Rest Api.

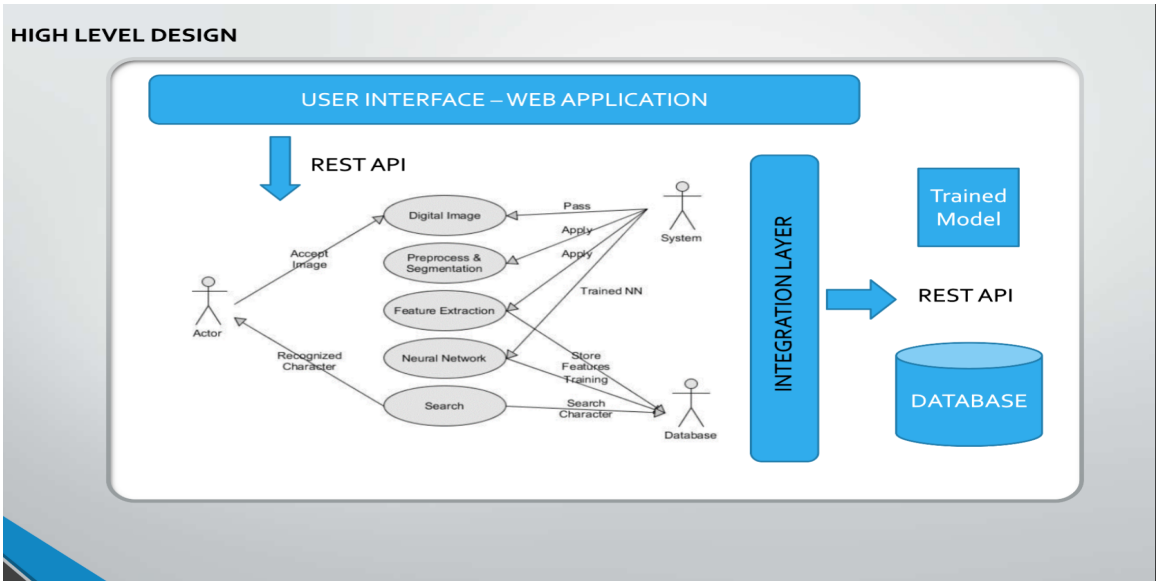


Figure 6.1: Use Case Diagram

6.2 Class Diagram

The Class diagram characterize the classes of the application.We end up with seven classes(User,Captureimage,procesimage etc) to build and run the application.This class diagram describe the relationship among the classes and show their attributes and methods.

Figure 6.2: Class Diagram

6.3 Activity Diagram

In the proposed activity diagram model.We input the document image(Prescription image) then it go through the processing and the text extraction can be done then the extracted data passes to the segmentation process,then the character images are label from the text,then feature extraction can be done on it and at the end the classifier model and the label process will be apply on it.

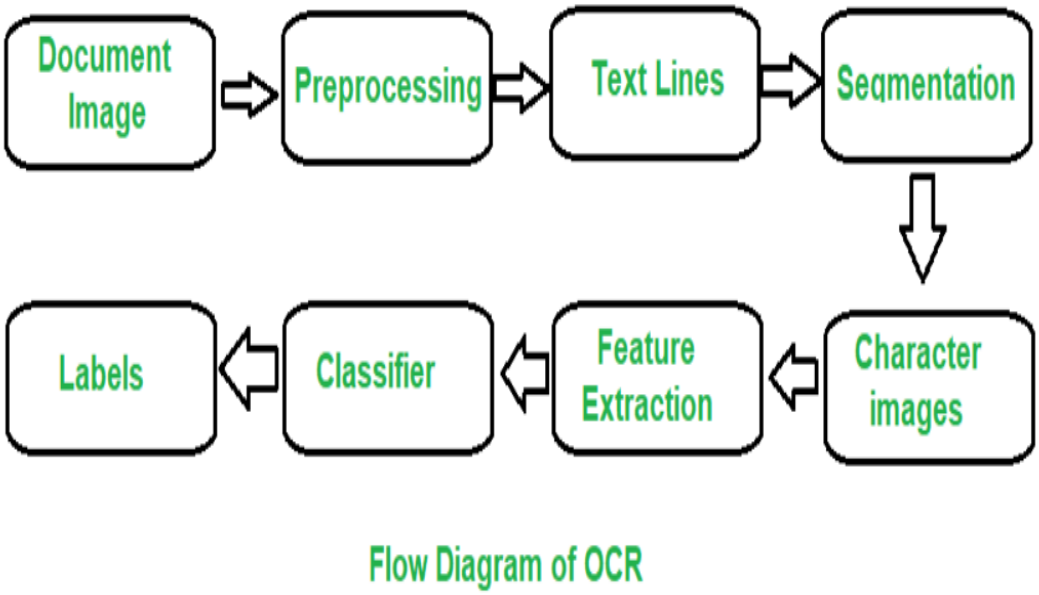


Figure 6.3: Activity Diagram

6.4 Sequence Diagram

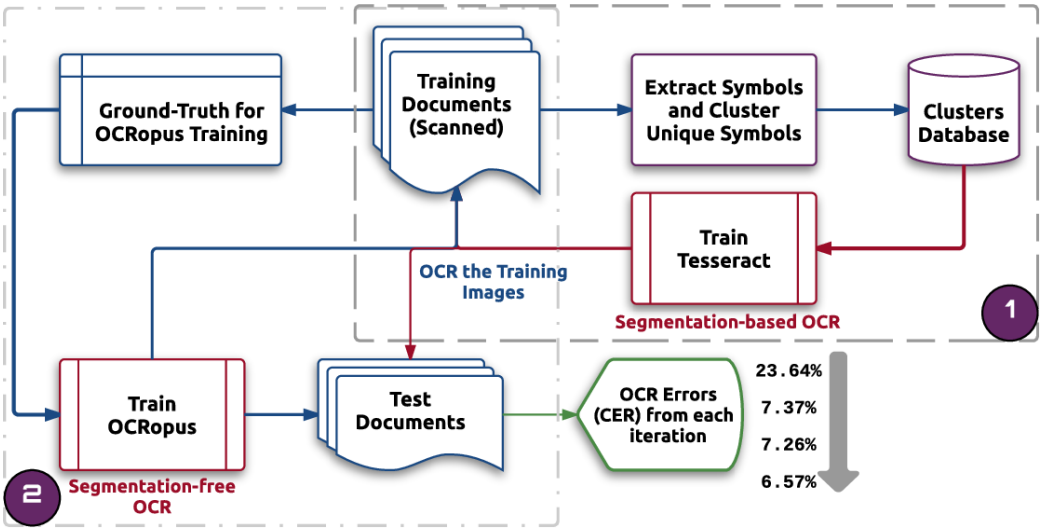


Figure 6.4: Sequence Diagram

Chapter 7

SYSTEM IMPLEMENTATION

In this section, we briefly outline the workflow of our method referred to as Smart OCR. Our Goal of the project to create textual format output from scan or capture image of prescription. To create the output we collect the prescription form our application User Interface which are provide by patient and doctors. The image goes through initial step of Page detection /background noise removal here we process and crop the image to removal maximum amount of noise to improve efficiency of model. Then the image is sent to Word detection model here the image is segments to remove internal noise and sent only needed textual image for OCR. After Segmentation the image goes through preprocessing to improve image quality and OCR capability. Finally the image goes to through Optical character recognition here we have used combination of Convolutional recurrent neural network and Tesseract Model to give output with highest precision and accuracy.

The Output from OCR is stored in Database with additional data about prescription and patient. Patient/ Doctor can access the prescription through UI depolyed at web.The data is protected using two-factor authentication to protect from any data breach.

7.1 Page Detection

Page Detection Module can be also be called noise removal module. In this module we process the image to remove unwanted background segments. We are using Grey scale converstion and Canny edge Detection method to find image border between prescription and unwanted background. Initially we Median blur the image to minimize the effect of textual content on edge detection, creating a image with highlighted different edge. The threshold value for blur is applied to remove unwanted details during edge detection. Then we add black border to create contour for the image edge. After the process module find main edge border for the image, We add 5px offset to contour to prevent any corner data loss. The image cropped after proccess according to final border after offset.

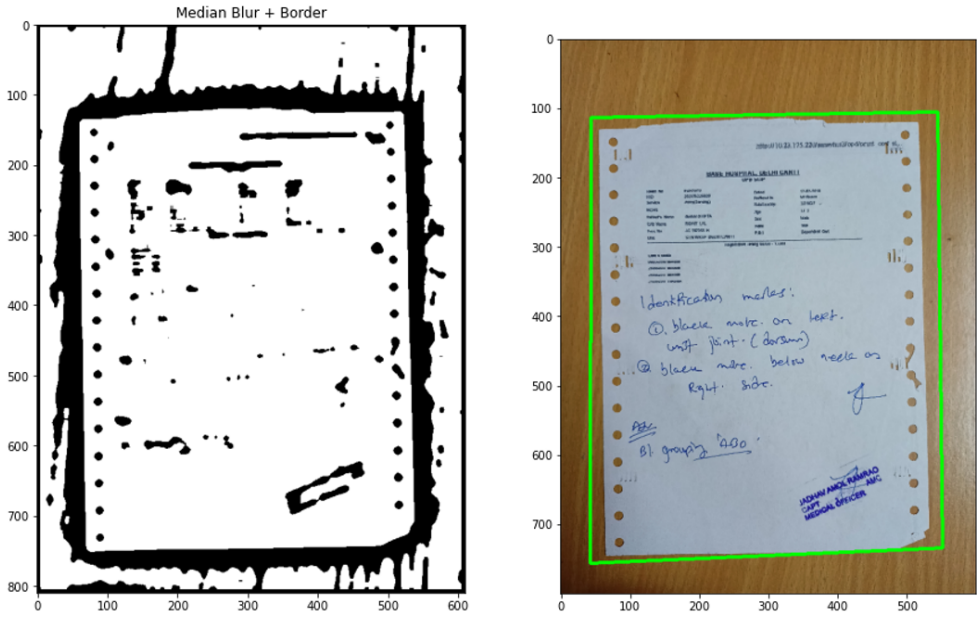


Figure 7.1: Page Detection

7.2 Word Detection

Word Detection module is basic model to segment large image from page detection model with multiple textual paragraph into small word segments to increase accuracy of OCR model. Here we 2 different technique firstly sober operator and Watershed algorithm to segment image in image. In Sober Operator we initially blur image a little to find change is continuity of word and creating segments based on the finding in contours.In watershed algorithm we initially remove basic noise to reduce interference to the algorithm. Then we convert image by labeling each pixel to 0 or 1 based on pixel value in comparison to threshold value. After the process we create contour based on values. The image segments are created based on the contour.After combining the segments from both the process the segment are restructured and regrouped based on size and contour. The final result gives a group of segment to send for OCR.

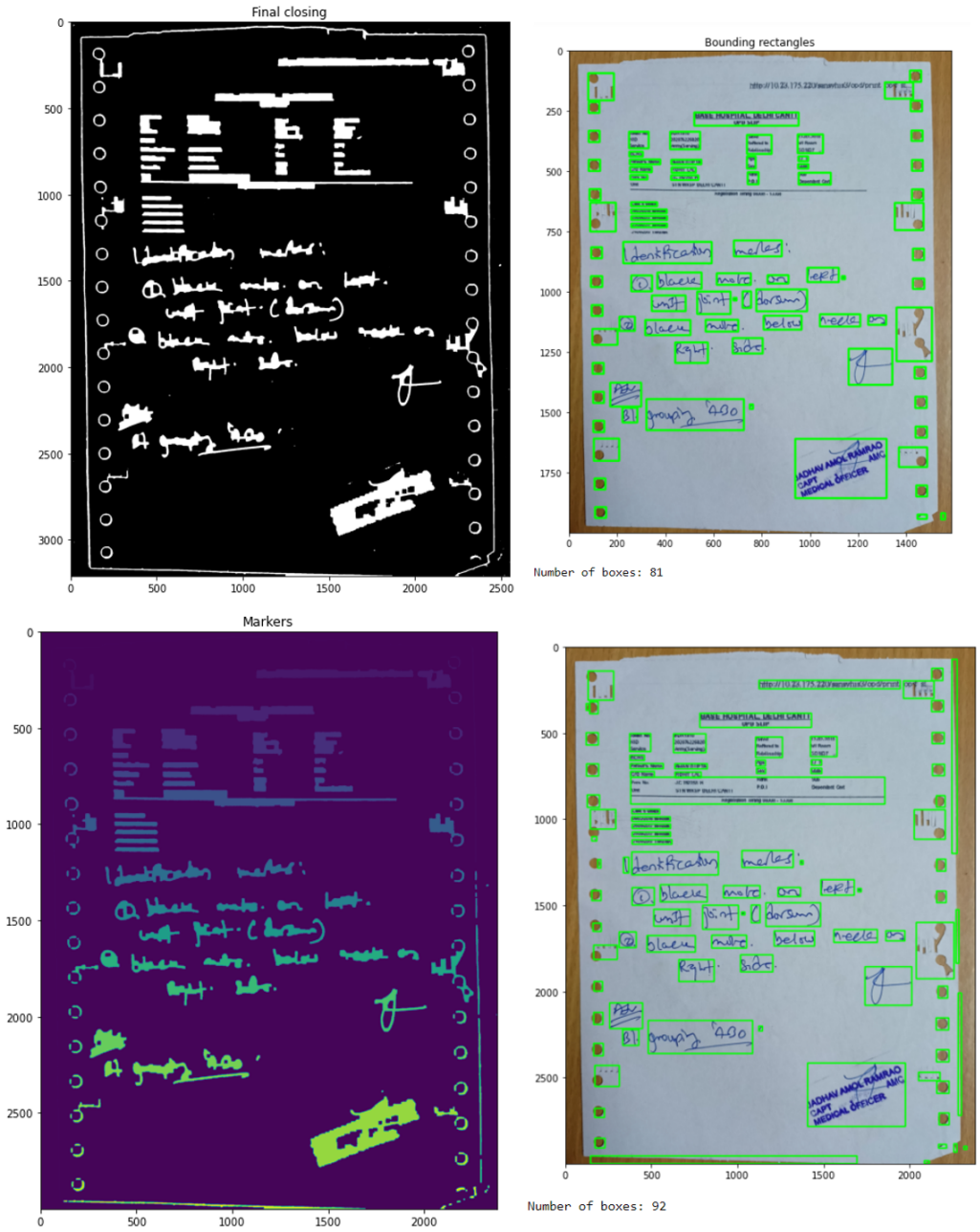


Figure 7.2: Word Detection

7.3 Word Preprocessing

Data preprocessing is a process of preparing and improving the raw data and making it suitable for further OCR Models.Here we used Word detection output segment individually as input to process the image. Initially we imported preprocessing library and used them to preprocess and normalize the image. Normalization can be explained as process of changing the values of numeric columns in the dataset to use a common scale, without distorting differences in the ranges of values or losing information. We applied Thresholding of image, Histogram Equalization, Using filtering on normalized image, Data augmentation using elastic transform to image for preprocessing.

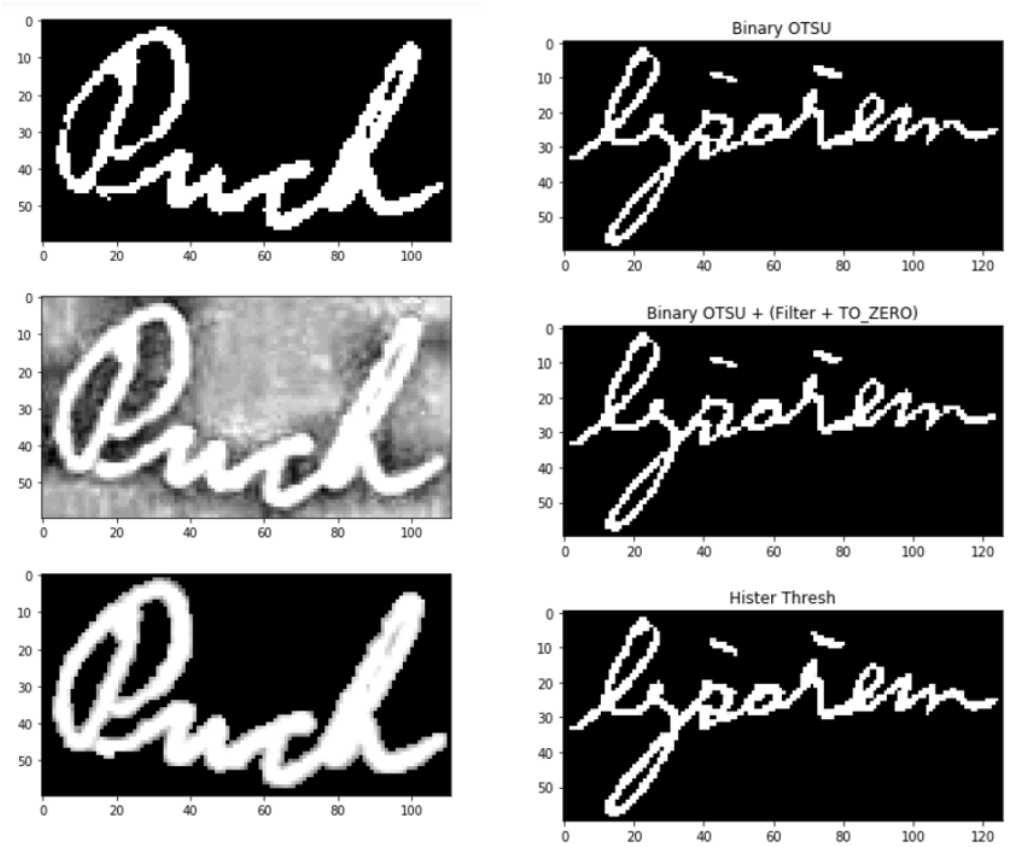


Figure 7.3: Word Preprocessing

7.4 Optical character recognition

Optical character recognition is Main module for our project. We train the CRNN model which is combination of 2 major machine learning models CNN and RNN. The CRNN model is trained upto 140000 iteration to reach a model accuracy of 0.941. We also added tesseract model for printed text which has accuracy of more than 0.95 for recognizing printed text. Both model are combined based on Precision for each word segment from previous word detection module.

```
Kernel interrupted, saving model...
Accuracy 0.941
```

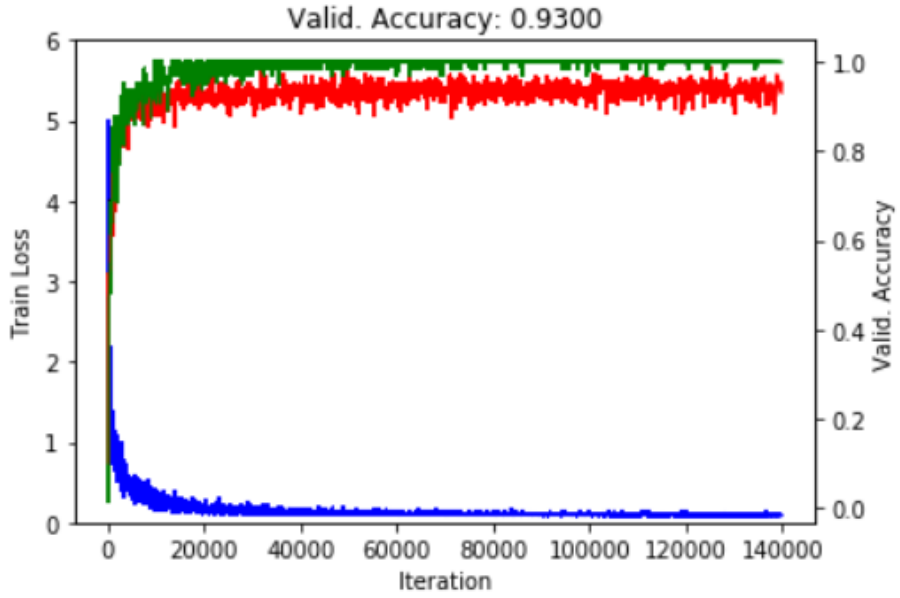


Figure 7.4: CRNN Training

7.5 Web access

We have created a web page where user can upload an image at a time, and the OCR of image would be done in background and user can see the text and at the same time the text would be saved in database. To implement this we have used Django framework. We created a template, which is displayed to the user and created a function which takes an image as input, and passes

that image to the ML codes as input and receives the OCR text generated by ML codes and renders the same in front end to the user and also creates an object in the database for the same image.

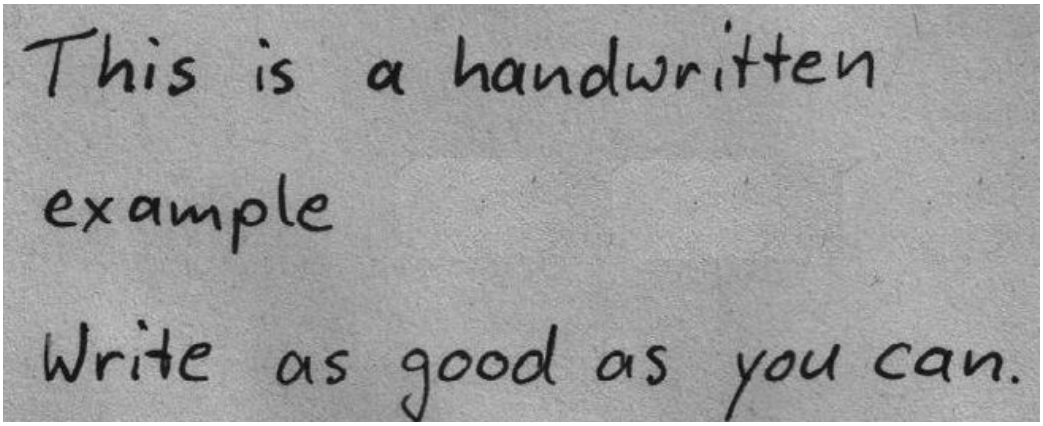


Figure 7.5: Input image

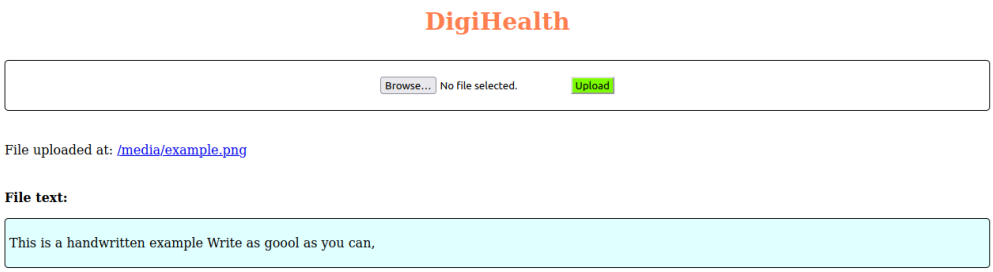


Figure 7.6: Output image

Chapter 8

System implementation-code

8.1 System implementation-code

8.1.1 Cloning the repository

Following steps are there:-

- !git clone –recursive https://github.com/Veritasosrb/veridical.git
- Go to Branch "AIT-2022-DigiHealthOCR"

8.1.2 Installing Environments

Install the following softwares:-

- Python
- Anaconda
- Jupyter notebook

After this run the following:-

- Run Anaconda Powershell
- Run commands:-
  - "conda create –name ocr-env –file environment.yml"
  - "conda activate ocr-env"

8.1.3 Running Modules

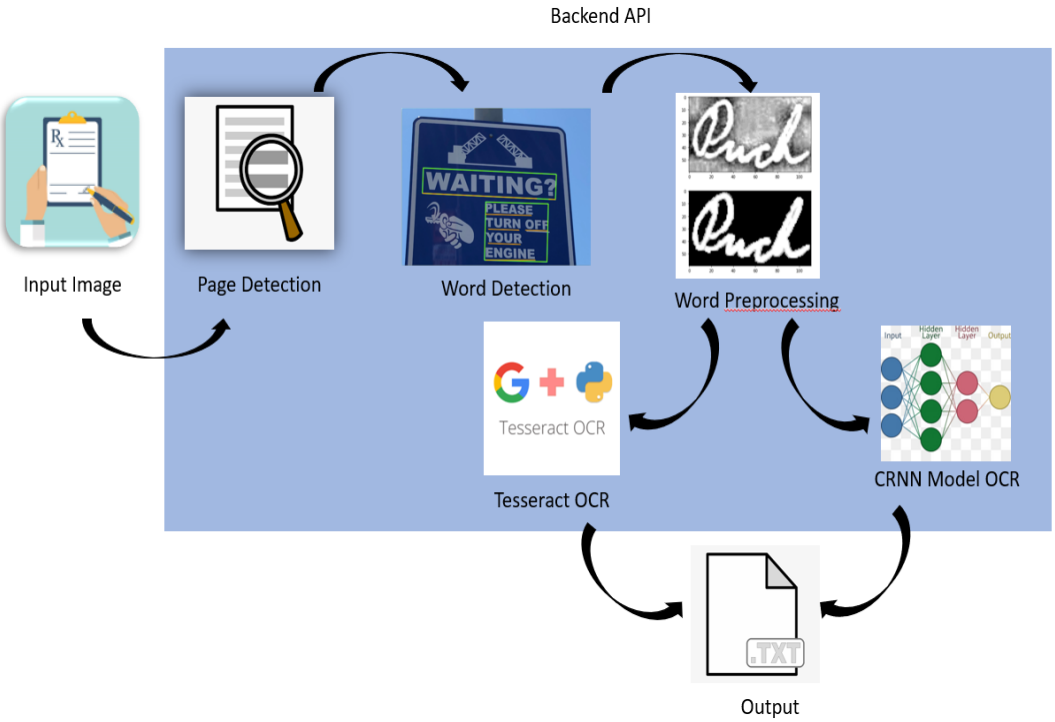
Start Jupyter Notebook with ocr-env environment.
Open Notebook Folder , It contain all executable python notebook modules

1. "page_detection.ipynb" conatins Page Detection Module
2. "word_detection.ipynb" conatins Word Detection Module
3. "word_preprocessing.ipynb" conatins Word Preprocessing Module
4. "ocr_evaluator.ipynb" conatins Model Accuracy calculator Module
5. "OCR_CRNN.ipynb" conatins OCR Module using CRNN Model
6. "OCR_CTC.ipynb" conatins OCR Module using CTC Model
7. "Tess.ipynb" conatins OCR Module using Tesseract Model

8.1.4 Running Web Server

1. Install Django using Command - "py -m pip install Django"
2. Go to Web folder and run Command prompt
3. Run Command "python manage.py runserver"
4. Open Browser and Open URL "http://localhost:8000/ocr/upload-doc/"
5. Upload Image to the page for OCR to Receive the Output

8.2 Flow Diagram



Project Workflow consist of collecting medical prescription from patient as well as doctor in scanned or Image capture format through web user interface. After the receiving the image it goes through many layer of Machine Learning models to remove noise and preprocess the image. After preprocessing the image further goes to main Combined model based on CRNN and Tesseract to get the OCR output in textual format. The output further goes to dictionary for word completion and grammatical correction .Then the output is stored in Database with many other details stored regarding the prescription and patient.The output can be accessed through web user interface.

Chapter 9

Test Results

Test Result on a Random Image through Django API using OCR function which is using CRNN and Tesseract Models.

Figure 9.1: Testing Output

Chapter 10

Experimental Results and Result Analysis

Comparison between Accuracy of CRNN, CTC and Combined model. Final is a combined model is combination of Tessereact and CRNN give an Accuracy of 75.431 .

Model	Accuracy	Total Words	Accurate Word	Time Taken
Combined Model	75.431	48660	36704	107.3 sec
CRNN	60.69	1623	850	39.88 sec
CTC	30.1	1623	490	22.8 sec

Table 10.1: Model Accuracy

```
STATS: CTC
--- 22.8 seconds ---
Correct/Total: 490 / 1623
CERacc: 69.6242 %
Letter Accuracy: 30.191 %
Letter Accuracy without Correction: 24.6457 %
Word Accuracy: 0.0 %
Word Accuracy without Correction: 0.0 %

STATS: CRNN
--- 39.88 seconds ---
Correct/Total: 850 / 1623
CERacc: 60.6901 %
Letter Accuracy: 52.3722 %
Letter Accuracy without Correction: 44.9784 %
Word Accuracy: 0.0 %
Word Accuracy without Correction: 0.0 %
```

Figure 10.1: Accuracy CRNN and CTC

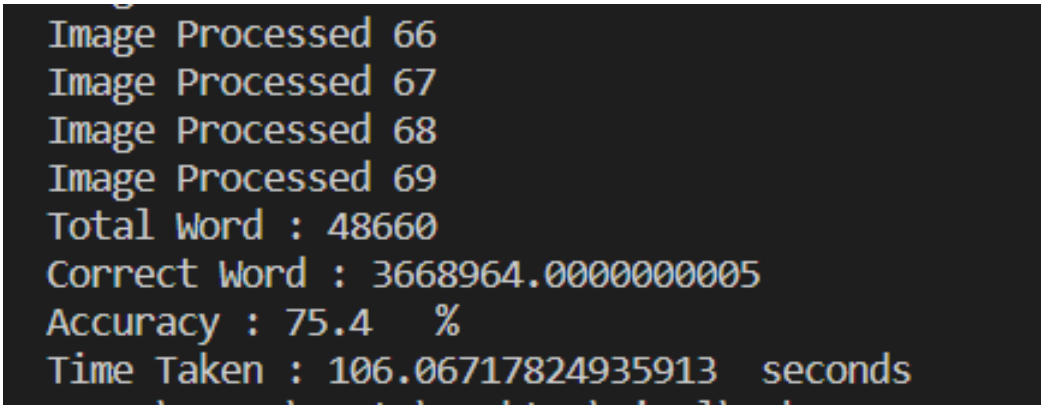


Figure 10.2: Accuracy Combined Model

Chapter 11

CONCLUSION

Most of the medical data in the Indian healthcare system is present in paper form, whether it is doctorâs prescription, patients report, etc. Since the population of india is huge and thus the medical data is also big. Now, with advancement in machine learning and artificial intelligence the data can be analysed properly and something can be made out of that data that could be beneficial to whole mankind.We are trying to bring data in electronic format using OCR.But doing OCR of doctors handwriting is not an easy task. Most of the time doctors use abbrevations to write prescriptions, their handwriting is also not so ligible.

Seeing all these factors we are using a combination of both RNN and CNN to make benefits out of both the neural networks. We will be also using external APIs to match the output of machine with the actual existence of such medicine or treatment.Moreover, bringing the patients all the medical data in one common platform will also bring relief to patients and doctors. As patients wonât have now to store all the previous slips and reports, etc. For doctors it will be easier to see patients past records and treat patients effectively.

Chapter 12

Appendix

12.1 References

Bibliography

[1] Aman Gupta , Yuvraj Gholap , Atul Kumar Tiwari , Alam Kathat, Sanju Baloria *Aspects of E-Prescription: Creation and Security*

[2] Ali Khalili,Bita Sedaghati. *Semantic Medical Prescriptions â Towards Intelligent and Interoperable Medical Prescriptions.*

[3] Sagar Kawthankar,Eram Ansari,Raj Joshi,Silviya DâMonte. *Smart Analytics And Predictions For Indian Medicare.*

[4] Sudipta Chandra, Soumya Ray, R.T.Goswami. *Big Data Security in Healthcare.*

[5] Akshay Raul, Atharva Patil, Prem Raheja, Rupali Sawant. textitKnowledge Discovery, Analysis And Prediction in Healthcare using Data Mining And Analytics.

[6] Farhad Ahamed,Farnaz Farid. *Applying Internet of Things and Machine-Learning for Personalized Healthcare: Issues and Challenges.*

[7] Roger Achkar, Khodor Ghayad, Rayan Haidar, Sawsan Saleh, Rana Al Hajj. *Medical Handwritten Prescription Recognition Using CRNN* in *IEEE Conference, 2020.*

[8] R.Parthiban, R.Ezhilarasi, D.Saravanan. *Optical Character Recognition for English Handwritten Text Using Recurrent Neural Network* in *IEEE Conference, 2020.*

[9] Florin Daniel Anton, Silvia Anton. *A Patient Data Management System for Medical Services and Training* in *IEEE Conference.*

[10] Baki Koyuncu, Hakan Koyuncu. *Intelligent Hospital Management System (IHMS)* in *IEEE Conference, 2015.*

[11] Rumia Sultana, Rajesh Palit . *A Survey on Bengali SpeechâtoâText Recognition Techniques.*

[12] Shobha Bhatt, Anurag Jain, Amita Dev. *Acoustic Modeling in Speech Recognition:A Systematic Review.*

[13] Ayushi Trivedi,Navya Pant, Pinal Shah,Simran Sonik and Supriya Agrawal. *Speech to text and text to speech recognition systems-A review.*

[14] A. Vithya Vijayalakshmi, Dr. L. Arockiam. *Hybrid security techniques to protect sensitive data in E-healthcare systems.*

[15] Kissi Mireku, Kingsford Zhang, Fengli, Gbongli Komlan. *Patient Knowledge and Data Privacy in Healthcare Records System.*

[16] Suman K. Saksamudre, P.P. Shrishrimal, R.R. Deshmukh. *A Review on Different Approaches for Speech Recognition System.*

[17] Smita B. Magre, Ratnadeep R. Deshmukh,Pukhraj P. Shrishrimal. *A comparative study on feature extraction techniques in speech recognition.*

12.2 Paper Published

https://www.ijert.org/aspects-of-e-prescription-creation-and-security

Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 11 Issue 02, February-2022

Aspects of E-Prescription: Creation and Security

Aman Gupta
Army Institute of Technology
Pune, Maharashtra

Prof. Yuvraj Gholap
Army Institute of Technology
Pune, Maharashtra

Atul Kumar Tiwari
Army Institute of Technology
Pune, Maharashtra

Alam Kathat
Army Institute of Technology
Pune, Maharashtra

Sanju Baloria
Army Institute of Technology
Pune, Maharashtra

***Abstract*:- Healthcare is ever developing field focusing on creating more effective and efficient process of treatment. In era of smart technology, Machine learning and big data analytics are playing huge role in boosting the growth of healthcare system. But it faces a big problem of lack of data in electronic format. India healthcare system mainly depend on paper-based prescription and report, it does not have needed data in electronic form for its better use and prediction. Our Paper focus on developing machine learning algorithm to create digital record from old paper-based record using Convolutional Recurrent neural network algorithm for Image processing and creating new record using voice to text conversion using Hidden Markov model. The record created could further be analyzed to created better quality data. The newly created data could increase scope and precision of healthcare machine learning models.**

1 INTRODUCTION

Healthcare is most basic requirement of a person. India is leading medical and pharmaceutical hub of the world. Indian healthcare system is advance but still not digital. Our Healthcare still depends on paper based physical record rather than digital records. With Technically advancing world, machine learning has become major aspect of development in healthcare system. Machine learning is field of data analytic and requires large amount of high quality of data for better precision. Since Healthcare is very critical field and even small mistake can cost a life it requires precision of nearly 100 percentage. The required high precision could be gain by high quantity and quality of data. The paper propose method to tackle problem of bringing data in electronic form. OCR will bring all the paper based data and voice to text conversion will allow doctors to quickly diagnose the patient and at the same time bring data in electronic format. Later on this data will be used to predict lot of things with growing machine learning advancements.

2 HEALTHCARE DATA

Data has been most important part of healthcare it boosts the growth and development of the field. Need of data is more than ever before in the world of advancing technology. Since ages data in healthcare is in form of prescription and reports given by doctor to patient mostly vocally or handwritten on fragile object like paper. It provides a great insight of disease and patient. There was no copy or centralized data with hospitals. The fragile prescription would get lost or deforms beyond readability and restoration. Since medicine were given of prediction basis a lack of data posses a threat of disaster ranging from allergies to death of patient but I well maintained patient data can even predict upcoming disease and help doctor creating personalized treat for the patient for faster and better recovery.

Today hospital store basic information about patient but still work mainly work with paper-based perception patient record. The perception and reports held by patient are vulnerable to deformation and cannot be used by doctor for analysis. Hospital database can be maintaining but basic information is not considered as a good quality of data even if there is a large quantity on based [1] data quality is more important than data quantity since quantity without quality could lead to hazardous results. It explained that we can increase the quality by storing detailed information of patient Diagnosis report, Hospitalization and recovery process. The detailed high-quality information could help doctor in solving later complications and a detailed study of many such cases could lead to more efficient process of dealing with same disease later.

There are also growing numbers of Hospital providing detailed records of the patients. But No centralized database is shared between them which could lead to communication gap between hospital. With growing technology and advancement in Big Data and Machine learning we can predict, detect disease and Train machine to help under low medical staff. Data can also play an unimaginable role in curing disease by understanding treatment and running stimulation for creating more efficient method to cure it.

2.1 Healthcare Prediction

With growing technology and data production, the time can never be better to use machine learning and Artificial intelligence to predict and analysis the Healthcare problems. The Prediction can vary based kind of data, Quality of Data and Quantity of Data.

- According to [2] prediction could be based on prescription could be used to analysis the proper effect and effectiveness of medicine. These assessments could be used to prescribe better suited medicine based on previous data making treatment more efficient and faster, it also reduces the time used for trial-and-error method. The result can also help pharm company to reconfigure medicine based on effect and effectiveness.
- According to [3] machine learning could help by predicting of disease based on report and precious data, Like Image of skin when compare with previous record to predict whether it is skin cancer or fungal

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 11 Issue 02, February-2022

infection before it is too late to operate. [2] also discussed about predicted patient disease based on symptoms and there severeness which could help doctor in a case of outbreak. According [3] the process

Figure 1: Applications of machine learning [4]

increases efficiency of hospital while increasing the effectiveness of treatment.

- The [4] focus on benefits and problems with fully digitalized healthcare system. It provides many benefits of big data in healthcare like predicting epidemics, improving quality of life, improving efficiency of treatment, Genome sequencing to identify disease etc. It also showed problem like Medical identity Theft, Insurance fraud, Data leak and incorrect medication due to inaccuracy. It also tired to suggest solution to privacy by introducing Dynamic map reduce, 2-way authentication and column based encryption.
- The [5] discussed about using data mining for suggesting alternative medicine to patient for high-cost medicine and also analysing medical requirement need of lower income group to subsidize them for better help. [5] also focused provide insurance and other facility on demand rather than wild guess to reduce waste of fund provided by Government.
- According to [6] everyone has different daily routine, eating habit, physical and mental condition but are provide with same medication. [6] discuss about personalizing medication based on pattern and previous records using machine learning which could make treatment effective and efficient. It also suggests about lifestyle advice to increase effect of medication for to reduce treatment time.

3 SPEECH TO WORD CONVERSION (SPEECH RECOGNITION)

Speech recognition refers to ability of a program/machine to identify words or phrases from a audio or voice. Exploration of Speech recognition has first started in 1950s,Audrey was the first speech recognition system Although it can recognize only digits then IBM

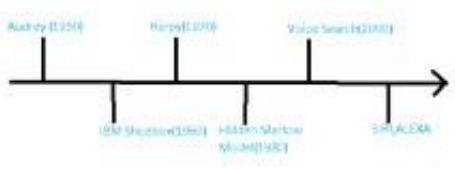


Figure 2: History of speech recognition

shoebox in 1960 after that HARPY in 1970 then HIDDEN MARKOV MODELS and now we have Alexa, Siri.

3.1 Signal Pre-Processing

Before going in to the practical view of speech- to–text systems we should know the basic of signal processing first. In this we have:

- Sampling the signal :- As Analog signal are memory hogging because of having unlimited no of samples and due to which processing the is computationally expensive. Due to which sampling comes into picture. Sampling is a process in which analog-signal is transformed into a digital-signal .
- Extraction Technique for an audio signal - In this technique from the speech signal we extract feature and later give them as a input to the model. Different types of extraction techniques are :– Linear Predictive coding (LPC) :-

  In speech analysis technique Linear-Predictive-coding is among the most effective technique, it also provide a accurate estimates about the parameter of speech and also used for encoding quality-speech at low-bitrate [11]. In Linear Predictive coding the first step is Frame blocking after that the next step is windowing this is done to minimise signal-discontinuities. After Windowing the Auto correlation analysis is done on samples and then the last step that is LP analysis which is based on Levinson-Durbin.

  - Mel-frequency cestrum co-efficient(MFCC) : - Mel frequencycestrum-co-efficient (M F C C) technique is based on human-auditory perception system and is one of the strong techniques. In MFCC to input signal certain steps is applied that is Framing, Windowing, Discrete Fourier transform and Mel Filter Bank Algorithm
  - Dynamic Time Warping

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 11 Issue 02, February-2022

3.2 Acousic decoder

In this paper [12] it is discussed in detail.

- Acoustic Models :- In Speech-to-text(STT) Acoutstic-Model is a fundamental part . In this the corelation between acoutics

Aspects of E-Prescription : Creation and Security

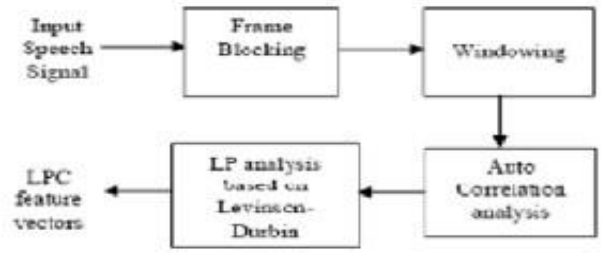


Figure 3: Block diagram of linear predicting coding [16]

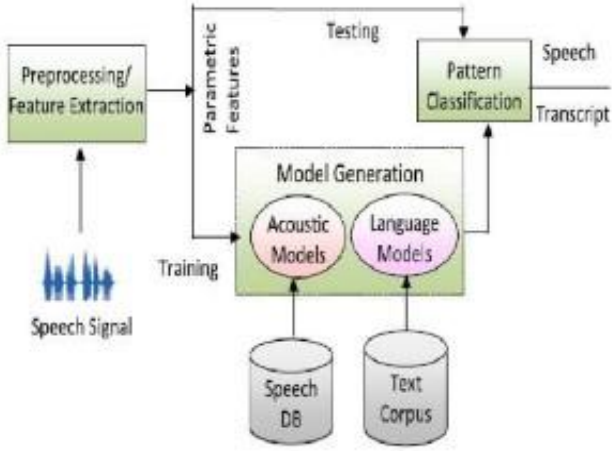


Figure 4: Architecture of speech recognition [17]

information and pheonetics is created. This is done by training which establish a connection between pheonetics(Basic Unit Of Speech) and acoustic - Observations .

- Language model :- In this model we find the probability of occurrence of a word after a certain sequence of word. In case of acoustically confused spoken utterance language models are responsible for making decision by incorporating syntactical and semantic constraints. It also distinguishes word or phrases that has similar sound.
- Pattern Classification :- In this method unidentified pattern is matched with the exiting sound-reference pattern and then similarity is computed between them . At the time of testing after completion of training, pattern is classified for the recognition of speech . Different methods or approaches for classification of patterns are:-
  - Template
  - Knowledge
  - Neural Network
  - Statistical

3.3 Speech to Word Conversion Methods

- Hidden Markov Model :- It is one of the best model for mod-
  elling discrete-state-operations.In STT or Speech-Recognition
  it is proved to be very effective. In this series or observation - series is made from infinite- state of a process at each time a transition is generated and for each transition probability is also calculated. . This model is best for real-time speech-to-text(STT) conversion for mobile users.
  Parameters on which it depend is-
  - Recognition accuracy
  - Recognition Speed
- Cuckoo Search Optimization (using Artificial Neural Network Classier):- Automatic speech recognition with this optimisation method is used for improving recognition, removing unwanted noise and better performance .
  Steps involved are -
  - The first step is pre-processing which is the crucial portion of speech to word, which is done to delete preventable waveform from signal. To remove background noise signals are passed through high-pass filter.
  - Acoustic features are extracted from the speech signal. they are of 2 types:-

a) Mel - Frequency - Cep - Strum - Coefficients (M F C C)

b) Linear - Predictive - Coding - Coefficients (L P C C)

  - Classification: The classifier used in this is ArtificialNeural-Network. The neural-network is a three layered classifier with n-input nodes,k-output nodes 1 hidden node.

4 OPTICAL CHARACTER RECOGNITION

Most of the medical data is in the physical pages. To bring that data in electronic form we will use OCR. We are making an OCR for hand written English characters. The challenging part here will be that doctors use lot of abbreviations and mostly write in cursive writing which is difficult to understand. We have also discussed how can we tackle this challenge. The [7] presented the study of using CRNN that is Convolutional Recurrent Neural Network.

4.1 CRNN Introduction

Convolutional Neural Network Recurrent Neural Network together constitute this network model, hence having benefits of both network models.

The simple feed forward neural network only considers current input and cannot memorise past inputs, so it cannot process the sequential data. In CNN pixels of image is fed as the input and then it is fed to input layer which accepts the input in the form of arrays. After this it is passed on to hidden layer which carry out feature extraction. Convolutional, ReLU and Pooling layers are important hidden layers. At last there is a fully connected layer that recognises the object in the image.

The RNN can process sequential data as it takes into account current input as well as past inputs (Due to added feature of internal memory it can remember its past inputs). RNNs utilises different data points in a sequence to generate more accurate results. They basically takes inputs and reuse the activations of previous nodes or subsequent nodes in the sequence to get the output.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 11 Issue 02, February-2022

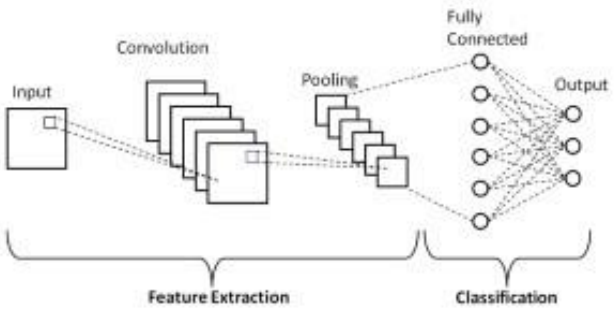


Figure 5: Basic convolution neural network [7]

The proposed CRNN consists of 13 convolutional layers accompanied by 3 Bidirectional Long Short Term Memory layers. The ReLU activation function is applied over convolutional layer to get a rectified feature map of the image.

The images are inputted and splitted into windows or sub-images, keeping height of sub-image equal to that of text line image. From the feature maps generated by the final convolutional layer, the vectors are extracted and passed on to the LSTM layer. Then outputs of both forward and backward bidirectional LSTM layers is combined together, instead of doing it at the end of each layer.Due to this weights are also optimized faster.

4.2 Training and Validation

In this paper[7] CRNN is applied using python. The dataset contains clearly written handwritten letters as well as doctor's prescription for training and testing. Output contains spaces, punctuation and numbers.

Since doctors most of the times write more than one prescription very closely as shown in the figure below, the initial challenge is to divide prescriptions alone into small segment. Since program won't be able to detect each prescription individually so each prescription was manually inserted line by line. This brought more precise result with reduced computational time.

In paper[7], since prescriptions were small in size so after splitting these were given 340 * 60 dimensions. Sliding window of width '3' is taken so that 3 or more characters of input image can be scanned at a time. The height of the window was adjusted to 64 pixels. It should be noted that the total number of feature vectors generated from every sliding window(SW) is crucial to give sufficient sampling for image.

From each window, sixteen feature vectors are generated from feature maps produced by last convolutional layer and passed onto observation sequence.

More than 9,750 training samples should be processed, so as to validate data. Also in the paper[7] the initial parameters were defined like epoch(=0), batch size(=10) and learning parameter(=0.0005). Number of image processed at every iteration is termed as Batch size. To properly train and provide neural network, input prescription slips of different doctors is taken. Output format is text for each input image.

After completion of training, a log file is generated containing details of each epoch that is training error, training loss, etc.

4.3 *Results*

In this paper [7] it is observed that training machine with input of clearly hand written characters and doctors prescription. The experiment achieved accuracy of around 95 percent. Still algorithm needs to work on reading hard paragraphs and split them into individual characters. The algorithm will not produce accurate result if during training only clear characters are used and for testing we use doctor's prescription and vice-versa is also true.

5 DATA SECURITY AND ACCESS

Creating a centralized database is a very important task for any system for fast and easy access to all data from a specific reliable source to further process making the whole infrastructure fast. But creating a centralized database come at a cost of security and access problem of its own. Since all the data is in a centralized any cyberattack on database may be disastrous and any security breach in security could lead to privacy hazard. Most of possible problem can be solved before and during creation of database by implementing some small strategies and implantation.

- Database has been an important part of healthcare since very long time, doctor have been referring to them for treatment analysis, treatment planning and comparing different patient's medical prescription for better service and treatment. It has also been useful to patience for understanding their treatment and keeping record of all previous report. A centralized database could improve the ability of doctor to treat a patient by getting all related information to patient at same place, as well providing all related case as well in same place making planning and research to the treatment easy and efficient. As for the patient he can access all his data any place any time at one place. A centralized database can even reduce cost of data management by resource sharing and reducing redundancy in data making the process of treatment more cost friendly.
- Centralized database for healthcare is important for modernizing the system while making it fast and efficient. But centralizing database have its own share of problem, cyber attack are very common occurrence to centralized system with no or improper data security algorithms. The most common problem considered is cyber attack to database to steal, append or change information for personal or criminal uses. Since healthcare database have large amount of personal and important data which could be stolen for selling data criminally or returning for ransom fees. There are many ways to deal with the problem most common way could for encryption and creating a Disaster Recovery plan.

– Encryption and decryption of data is most simple and effective way for securing the while only allowing access to original users. Encryption and decryption are usually done using private and public keys which are created using multiple mathematical algorithms. Here we encrypt our data using the key and sent it to respective user where it is decrypted for further use. Here the data remains totally secure throughout

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 11 Issue 02, February-2022

the transaction. Encryption is very effective way against cyber-attacks, Aspects of E-Prescription : Creation and Security

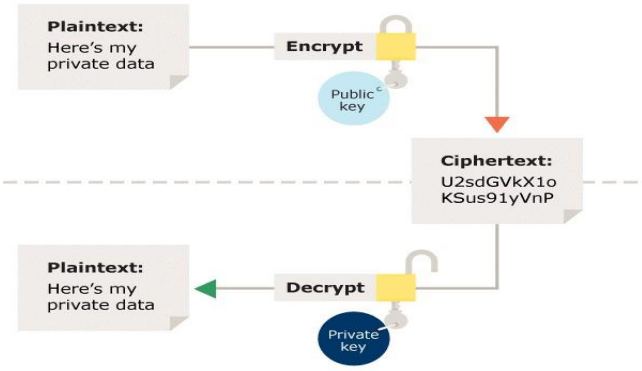


Figure 6: Encryption using Keys [14]

malware and ransomware. It also promotes high data integrity during transactions.

- Disaster Recovery Plan is also very important for data security from any ransomware attacks or physical attack to the server. Since Healthcare database store high amount of very important for treatment and losing them is not an option. So creating a Recovery plan help in case of any database failure. Disaster Recovery plan is very useful for all kind of ransomware and unseen disaster to database systems.

• Having Centralized Healthcare Database with good cyber security is important but useless if user does not have sufficient knowledge or authentication security. Have a secure authentication system and identification is more important than the Database system since without proper authentication system the data can endangered with very simple cyber-attack like brute force login attacks etc.

- System can use unique id for identification e.g., Aadhar card in case of Indian healthcare database. These simple identification ids are easy to memorise in case of healthcare emergence and unique to every user to stop identity conflicts.
- Identification system can help identifying the user but does not provide any kind of security. Authentication system is backbone for also secure user logins. It provide defence against all basic cyber-attacks. Multifactor authentication is the most common way to provide authentication to user in secure systems. Here 2 or more authentication mediums are used for providing more accurate and secure login by decreasing chance of any brute force login attacks to the user. Two factor authentication is most common way-out multifactor authentication, here we can use Password and OTP from mail or mobile to create a secure as well as fast authentication for the user.
- Authentication is an important process for protect the system from outside cyber-attacks, Buts protection against internal cybercrime are also important, there can be cases of illegal data sharing outside organization. Healthcare have higher risk to these types of attacks, pharmacist or patience can exploit if given open access to private information. Deciding information access to each user and giving access to private information for certain period is best possible way stop these cases. In Healthcare systems Private should only be directly accessed by emergency services, Doctor should be given right to have full access to nonprivate information for treatment and research works, whereas pharmacist and other patient could be given permission on consent of user for fixed time period. Best way implements the process could be done using tokens, a token-based system verifies every action based on token access provided where token can have basic condition to followed when ever been used, in our case a time-based token can be used and where access time is limited to certain user or private data. This could significantly decrease internal cyber crime and risk to the private data.

6 CONCLUSIONS

With advancement in ML and AI, more and more information could be processed and productive results can be generated. Hence storing healthcare data in electronic format becomes important to fight any outbreak and to find best suited treatment of diseases. Data could be collected through reports and handwritten characters using OCR, but since doctors handwriting in not easily readable for machine, we can use voice to text conversion for all such data to maintain accuracy and correctness of data.

Since data collection is personal and private we need secure database management to store it. Also the data collected could used for various medical researches and predictions. The data collected could be further used for personalized treatment of the patients.

REFERENCES

[1] Maria Ikram, Yan-ping Zhao, Ying Su. *Studying Hospital Process Management System Concerning Information Quality.*

[2] Ali Khalili,Bita Sedaghati. *Semantic Medical Prescriptions – Towards Intelligent and Interoperable Medical Prescriptions.*

[3] Sagar Kawthankar,Eram Ansari,Raj Joshi,Silviya D'Monte. *Smart Analytics And Predictions For Indian Medicare.*

[4] Sudipta Chandra, Soumya Ray, R.T.Goswami. *Big Data Security in Healthcare.*

[5] Akshay Raul, Atharva Patil, Prem Raheja, Rupali Sawant. textitKnowledge Discovery, Analysis And Prediction in Healthcare using Data Mining And Analytics.

[6] Farhad Ahamed,Farnaz Farid. *Applying Internet of Things and Machine-Learning for Personalized Healthcare: Issues and Challenges.*

[7] Roger Achkar, Khodor Ghayad, Rayan Haidar, Sawsan Saleh, Rana Al Hajj. *Medical Handwritten Prescription Recognition Using CRNN* in *IEEE Conference, 2020.*

[8] R.Parthiban, R.Ezhilarasi, D.Saravanan. *Optical Character Recognition for English Handwritten Text Using Recurrent Neural Network* in *IEEE Conference, 2020.*

[9] Florin Daniel Anton, Silvia Anton. *A Patient Data Management System for Medical Services and Training* in *IEEE Conference.*

[10] Baki Koyuncu, Hakan Koyuncu. *Intelligent Hospital Management System (IHMS)* in *IEEE Conference, 2015.*

[11] Rumia Sultana, Rajesh Palit . *A Survey on Bengali Speech–to–Text Recognition Techniques.*

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 11 Issue 02, February-2022

[12] Shobha Bhatt, Anurag Jain, Amita Dev. *Acoustic Modeling in Speech Recognition:A Systematic Review.*
[13] Ayushi Trivedi,Navya Pant, Pinal Shah,Simran Sonik and Supriya Agrawal. *Speech to text and text to speech recognition systems-A review.*
[14] A. Vithya Vijayalakshmi, Dr. L. Arockiam. *Hybrid security techniques to protect sensitive data in E-healthcare systems.*
[15] Kissi Mireku, Kingsford Zhang, Fengli, Gbongli Komlan. *Patient Knowledge and Data Privacy in Healthcare Records System.*
[16] Suman K. Saksamudre, P.P. Shrishrimal, R.R. Deshmukh. *A Review on Different Approaches for Speech Recognition System.*
[17] Smita B. Magre, Ratnadeep R. Deshmukh,Pukhraj P. Shrishrimal. *A comparative study on feature extraction techniques in speech recognition.*

12.3 Report Plagiarism

Project Report v3

ORIGINALITY REPORT

28%	27%	14%	14%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.ijert.org Internet Source	12%
2	www.geeksforgeeks.org Internet Source	5%
3	Submitted to Indian Institute of Technology, Madras Student Paper	4%
4	Submitted to Queensland University of Technology Student Paper	1%
5	Submitted to The Robert Gordon University Student Paper	1%
6	Submitted to Somaiya Vidyavihar Student Paper	<1%
7	"Data Science and Intelligent Applications", Springer Science and Business Media LLC, 2021 Publication	<1%
8	www.semanticscholar.org Internet Source	<1%

12.4 Sponsorship Project Completion Certificate

Veritas Software Technologies India Private Limited,
EON Free Zone, 0-02 Floor, Wing-4, Cluster A, Plot No.1,
Survey No.77, MIDC Knowledge Park,
Kharadi, Pune – 411014, India.
Corporate Identification Number: U72200PN2015FTC154978

24 May 2022

To Whom It May Concern,

This is to certify that **Alam Kathat** has participated in the Veritas university students outreach programme. He has successfully completed "Smart Electronic Health Records using OCR" which commenced 7/27/2021 and was conducted over a duration of 10 months

Their project consisted of "To read the handwritten documents and reports by doctors using OCR and store the data in database for further use in healthcare and research purposes" For any further questions, please reach Huzefa Vapra at 9579122289 or huzefa.vapra@veritas.com

Best Regards,

Sincerely,
For Veritas,

Leucadia Milly Sandeep
Pronouns: *she, her, hers*
India HR Leader (Business Partner & Talent Acquisition)
Human Resources
Veritas Software Technologies India Pvt.Ltd

Veritas Software Technologies India Private Limited,
EON Free Zone, 0-02 Floor, Wing-4, Cluster A, Plot No.1,
Survey No.77, MIDC Knowledge Park,
Kharadi, Pune – 411014, India.
Corporate Identification Number: U72200PN2015FTC154978

24 May 2022

To Whom It May Concern,

This is to certify that **Aman Gupta** has participated in the Veritas university students outreach programme. He has successfully completed "Smart Electronic Health Records using OCR" which commenced 7/27/2021 and was conducted over a duration of 10 months

Their project consisted of "To read the handwritten documents and reports by doctors using OCR and store the data in database for further use in healthcare and research purposes" For any further questions, please reach Huzefa Vapra at 9579122289 or huzefa.vapra@veritas.com

Best Regards,

Sincerely,
For Veritas,

Leucadia Milly Sandeep
Pronouns: *she, her, hers*
India HR Leader (Business Partner & Talent Acquisition)
Human Resources
Veritas Software Technologies India Pvt.Ltd

Veritas Software Technologies India Private Limited,
EON Free Zone, 0-02 Floor, Wing-4, Cluster A, Plot No.1,
Survey No.77, MIDC Knowledge Park,
Kharadi, Pune – 411014, India.
Corporate Identification Number: U72200PN2015FTC154978

24 May 2022

To Whom It May Concern,

This is to certify that **Atul Kumar Tiwari** has participated in the Veritas university students outreach programme. He has successfully completed "Smart Electronic Health Records using OCR" which commenced 7/27/2021 and was conducted over a duration of 10 months

Their project consisted of "To read the handwritten documents and reports by doctors using OCR and store the data in database for further use in healthcare and research purposes" For any further questions, please reach Huzefa Vapra at 9579122289 or huzefa.vapra@veritas.com

Best Regards,

Sincerely,
For Veritas,

Leucadia Milly Sandeep
Pronouns: *she, her, hers*
India HR Leader (Business Partner & Talent Acquisition)
Human Resources
Veritas Software Technologies India Pvt.Ltd

Veritas Software Technologies India Private Limited,
EON Free Zone, 0-02 Floor, Wing-4, Cluster A, Plot No.1,
Survey No.77, MIDC Knowledge Park,
Kharadi, Pune – 411014, India.
Corporate Identification Number: U72200PN2015FTC154978

24 May 2022

To Whom It May Concern,

This is to certify that **Sanju Baloria** has participated in the Veritas university students outreach programme. He has successfully completed "Smart Electronic Health Records using OCR" which commenced 7/27/2021 and was conducted over a duration of 10 months

Their project consisted of "To read the handwritten documents and reports by doctors using OCR and store the data in database for further use in healthcare and research purposes" For any further questions, please reach Huzefa Vapra at 9579122289 or huzefa.vapra@veritas.com

Best Regards,

Sincerely,
For Veritas,

Leucadia Milly Sandeep
Pronouns: *she, her, hers*
India HR Leader (Business Partner & Talent Acquisition)
Human Resources
Veritas Software Technologies India Pvt.Ltd