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

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
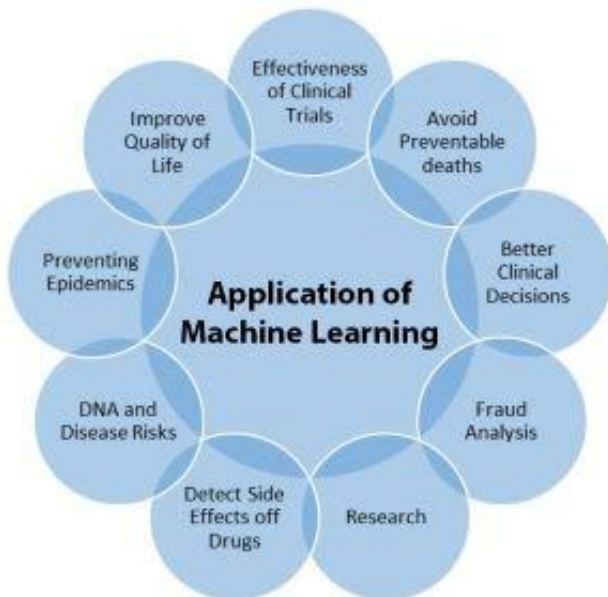

Figure 1: Applications of machine learning [4]

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

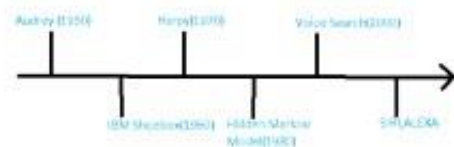the first speech recognition system Although it can recognize only digits then IBM

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

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

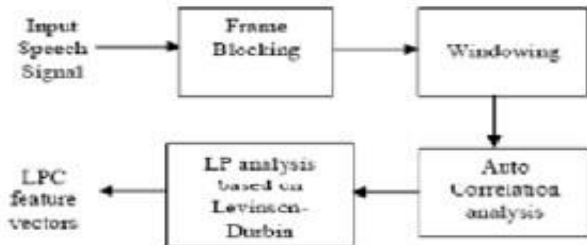Aspects of E-Prescription : Creation and Security

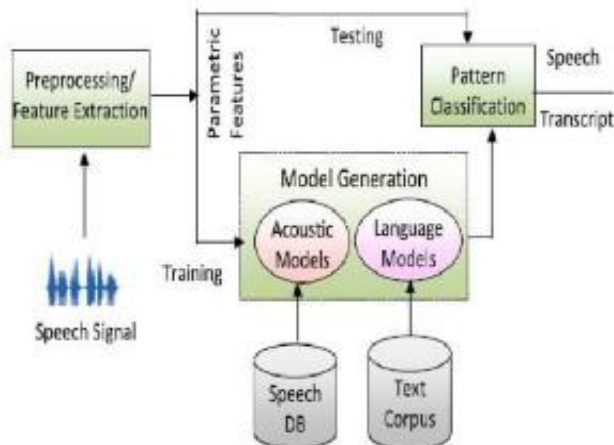Figure 3: Block diagram of linear predicting coding [16]

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.
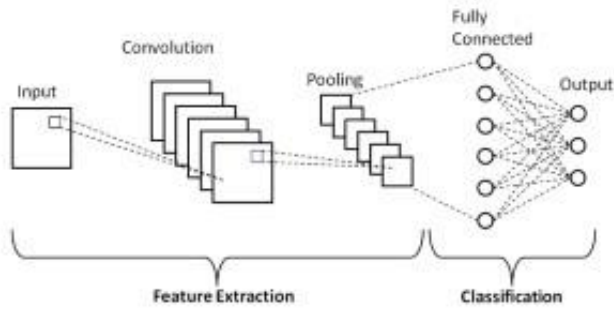
Figure 5: Basic convolution neural network [7]

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

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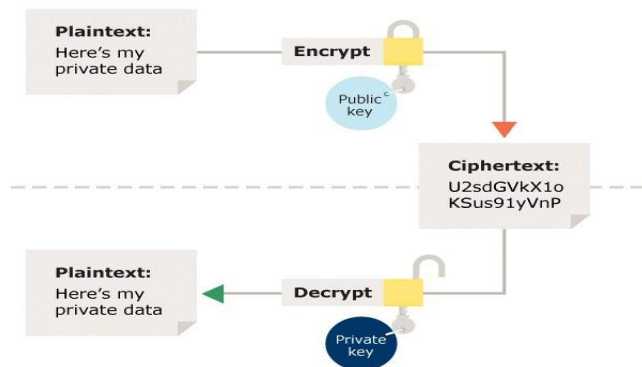

Figure 6: Encryption using Keys [14]

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. textit*Knowledge 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.*

[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.*