Other Technical Details

Project Title: Development of Prediction and Response System to Monitor Communicable Diseases in Real-Time using Deep Learning

1. Origin of the Proposal:

Building a fortified healthy society is the main objective of the government and other health organizations. In recent years a very good amount of progress has been made in improving the health of millions of people. However, in the case of a few communicable diseases such as malaria and tuberculosis, progress has slowed or not in good progress. In addition to the existing communicable diseases, now the entire world is facing problems with a dangerous novel corona (COVID-19) virus. As per the World Health Organization (WHO) [1], reports infectious diseases are still a major cause of deaths and ill health. The diseases most commonly associated with emergencies include diarrheal diseases, acute respiratory infections, measles, malaria, tuberculosis etc. Taking the example of novel corona (COVID-19), from 2019 onwards, despite being avoidable or preventable, nearly 200 countries' population was at risk of the disease. The WHO is also highlighted the importance of communicable disease alert and response for mass gatherings [5].

The goal of this project is to design a web and mobile-based prediction and response system to predict using deep learning learning algorithms, track, visualize with a geo-tag, prevent and treat the outbreaks of communicable disease in real-time. This enables the government or health organizations with a focus on vulnerable groups such as children, senior citizens and pregnant women.

This project addresses the gap between data science and healthcare computing. Addressing the gap is an effective way to bring the domains of data science, machine learning, and computing together in a way that can have long-term benefits to scale up the health data analysis.

Based on the aim of the project, the objectives are grouped into two broad classes (i) Back-end for data acquisition phase (ii) Front-end for disease prediction, track, visualize with a geo-tag, and dissemination system. The Detailed objectives of each class are given as follows:

1.1 Back-end side objectives

The back-end side of our project will concretely meet the challenge by the pursuit of the following research objectives:

- Implementation of a fast and scalable deep learning algorithms to ingest and pre-process communicable diseases data.
- Design and development of a computational framework consisting of post-processing and statistical analysis tools, to integrate the communicable diseases data from multivariate input sources.

1.2 Front-end side objectives

The objectives for the development of front-end disease prediction and data dissemination system are as follows:

- To design and develop a prototype framework for the generation of information from the processed data using the latest data visualization tools.
- Design and development of Web Application and Mobile APP, for the better dissemination of the knowledge.

2. Review of the status of Research and Development in the subject

The developments which are happening in the area of data analysis have good potential to apply it for healthcare data analysis. The research and development in the proposed area in the national and international arena have been discussed briefly in the following paragraphs.

2.1. International Status:

Some of the existing initiatives, the United Nations Development Programme (UNDP) in Africa [3], initiated a new mobile technology in data collection and prediction of future transmissions. The World Bank with the help of UNDP and the global fund [4], introduced a real-time monitoring system using mobile or tablet devices to monitor the malaria disease. The utilization of this technology is now helping the government or health organizations to predict, track, visualize, prevent and treat malaria outbreaks in real-time. This system contributed a remarkable amount of decrease in the malaria spread.

2.2. National Status:

In India, having a limited medical facility in rural areas and not good road connectivity and long distances need to travel between the communities to gather information, it is a very laborious, time taking and complicated task. Without accurate, quality, and timely data, it is very difficult to report where outbreaks are occurring or to predict very useful patterns of future transmission. This has very serious consequences when decisions need to be taken on which location should be targeted to prevent the upcoming danger by educating the public or to initiate treatment campaigns and how often the resources are allocated. A well designed communicable disease prediction and response system helps to initiate a health emergency.

In the national scenario, some work has been done in this direction like the AarogyaSetu [6] for the COVID-19 data only, but there a need for an integrated framework for the prediction, tracking, visualizing with a geo-tag and prevention.

Recently initiated digital India program aims to transform India into a digitally empowered society and knowledge economy by leveraging IT as a growth engine of the India economy, this project is in tune with digital India in the healthcare domain.

2.3. Importance of the proposed project in the context of the current status:

From above studies, it is observed that the proposed research project on design web and mobile-based prediction and response system to predict using machine learning algorithms, track, visualize with a geo-tag, prevent and treat the outbreaks of the communicable disease in

real-time, is much useful for the general public for taking the necessary steps in their life particularly people living in rural India and also useful for the government to take necessary steps to alert and prevent the spread of the combinable diseases.

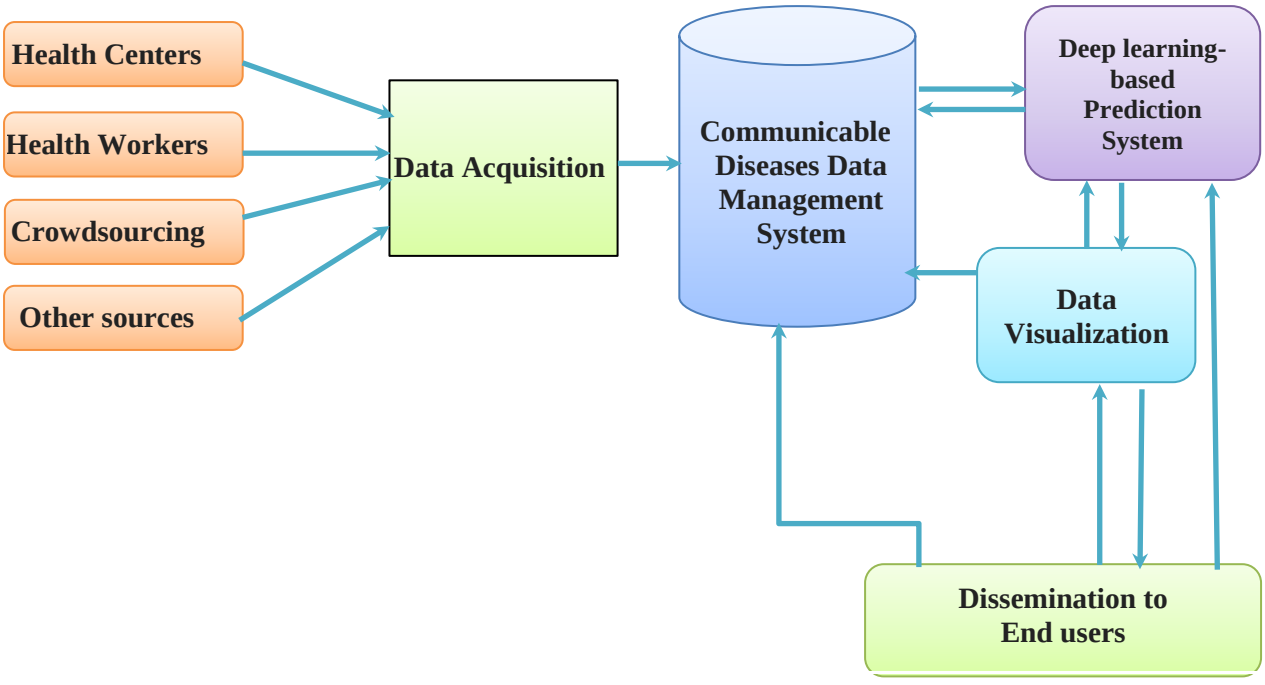
2.4. If the project is location specific, the basis for selection of location be highlighted:

For this project, we have selected the Telangana state as a study area. But in the initial stage of this project, we will do the experiments in the Warangal urban and rural areas.

3. Work Plan:

3.1. Methodology:

The overall methodology framework for the proposed research is shown in Figure 1. This framework consists of four major components viz. (i) Data Acquisition (ii) Back-end Data Management System (iii) Prediction and Visualization System (iv) Disseminating of the Knowledge to the end-users. The detailed explanation of the activities involved in each component is explained in the subsequent sections.



3.1.1 Data Acquisition

Ingestion and pre-processing of the multi-source communicable diseases data analysis will be carried out in a fast and scalable data framework. There four major data sources for this system which include: (i) Data collected from the Health organizations which are under the control of the government. (ii) Health workers, who working in the rural areas and serving in rural India. (iii) The collection from the crowdsourcing, in this the general public will provide data (iv)other sources like the non-profitable organizations. The architecture of this framework will be based on scalable machine learning. The ingestion workflows will be implemented, including a dedicated routine to convert the data to a format suited for further computation.

One of the major challenges of collecting ensemble online data from multiple input sources is to deal with the different start formats and ensemble sizes. Routines to handle the different formats and sizes of the data will be implemented.

3.1.2 Communicable diseases Data Management System

In this project, the communicable diseases data management and computing system usually need to handle a variety of heterogeneous data collected from multiple sources. In many of the cases, these systems are required to quickly answer the user's instant queries, for example, visualization of the affected areas, tracking of the affected persons and prediction of the future transformations. Without an efficient data management method, it is very difficult to organize the data received from multiple heterogeneous data sources. A typical workflow in the communicable diseases data management would involve operations like data representation in the form of the graphs and its maintenance; this is a frequently used data format for the representation.

3.1.3 Prediction System

A typical workflow of the model for the multi-source communicable diseases data analysis would involve operations like (i) Predicting the number of people is going to affect in the next few months or weeks, even the short term predictions like in the next 24 hours. (ii) Estimating the number of locations is going to affect in the specific time. (iii) The model is needed to train with a rich set of training data collected from different sources

The Deep Learning is the best approach and becoming popular for predicting patterns from the large data. In this prediction system we recast the popular DL methods like CNN [7], Autoencoder [8], Bidirectional LSTM[9] etc.

3.1.4 Data Visualization

User interfaces are of critical importance when it comes to the dissemination of knowledge obtained from the prediction system. Computer scientists are developing a number of open-source web-based visualization tools and apps to provide online data to users in the most effective way. Many researchers highlighted the benefits of the data visualization tools and the challenges involved in big data visualization. We will resort to custom implementations of state-of-the-art data analysis tools for web and mobile applications on the front-end side of this project, in order to enhance the scientific and societal value of the online data analysis and tailor the results to the specific needs of end-users. The methodology for the front-end prediction dissemination system consists of two phases:

Phase 1: Communicable diseases data visualization:

Communicable diseases data collected from different sources will be in the form of simple messages, database records, and flat files. The visualization of the input data and the predicted results will use standardized and state-of-the-art technologies within HTML5, PHP, and Android.

Phase 2: Knowledge Dissemination system:

A prototype web portal will be developed for the dissemination of the knowledge gained in the previous phase of the system. This system will be responsive and use graphs and text to communicate the knowledge. As advancement to the above web portal, a prototype mobile app will be developed for the benefit of the smart device users to the dissemination of the knowledge gained in the previous phase of the system.