SYSTEM TO PREDICT DISEASE FROM EARLY PHASE DATA

B.E. PROJECT SYNOPSIS

BY

Rubal Gupta
4ps13cs100

Vishal Soni
4ps13cs124

Ashwini MJ
4ps13cs027

Krishna Keshav
4ps13cs065

MENTORED BY

B.S. Mamatha
Assoicaiate professor
Computer Science Dept

PES College of Engineering, mandya

Computer Science and Engineering

2016-2017

INTRODUCTION

There are many extremly common and communicable diseases which requires affected patients to wait at the early phase before starting treatment, Some of the commonly existing examples are Breast cancer, Cholera, Typhoid, Tuberclosis and many more which eventually causes millions of deaths worldwide due to lack of accessibility to healthcare and awareness. Our project is one step towards many awareness programs being held worldwide to make people aware if they are prone to it according to thier current medical condidtions. We will be taking one particular among many such diseases and will develop a system that would predict the probability of occurence of diseases in a particular human body.

Problem Statement :

Problem Statement involves if there is any way to predict extremly and commonly existing diseases such as particular type of Cancer, Cholera, Typhoid, Tuberclosis etc. from current medical condidtions of a patient.

Our Solution:

Our solution towards the problem is that – yes, it is possible to develop an algorithm to predict such particular type of disease using Machine Learning. Since Machine learning is defined as “Making machine or a system capable of self-learning using few given data sets and provide results for input data”. We would like to quote “Machine Learning” as specialized field of technology and “Analysis of a particaular Disease” as a specialized field of study.

Feasibility study

The first phase of development of a system involves design and analysis of an algorithm to predict disease as efficiently as possible using data-sets with given resullts. We have been able to categorize our problem into “regression problem” and “classification problem”.

Using Regression technique, we provide data sets with already obtained results to the system. After machine learns from enough data we can provide it with input data and obtain a result as function of contionuous values which is probability of occurrence of disease in one particular human body.

Using classification problem technique, when the system has learnt enough we try to predict the results as a discrete values. Ex: 0 and 1. In our case, this could be as simple as “Positive” for highly exposed to one particular disease and “Negative” if users don't have to worry about that particular disease.

Such are only advancement towards development of an algorithm. Complete system requires implementation of an algorithm, storage for data-sets, retrieval of data-sets, accessibility towards system for common users and user interface of the system.

Methodolgy

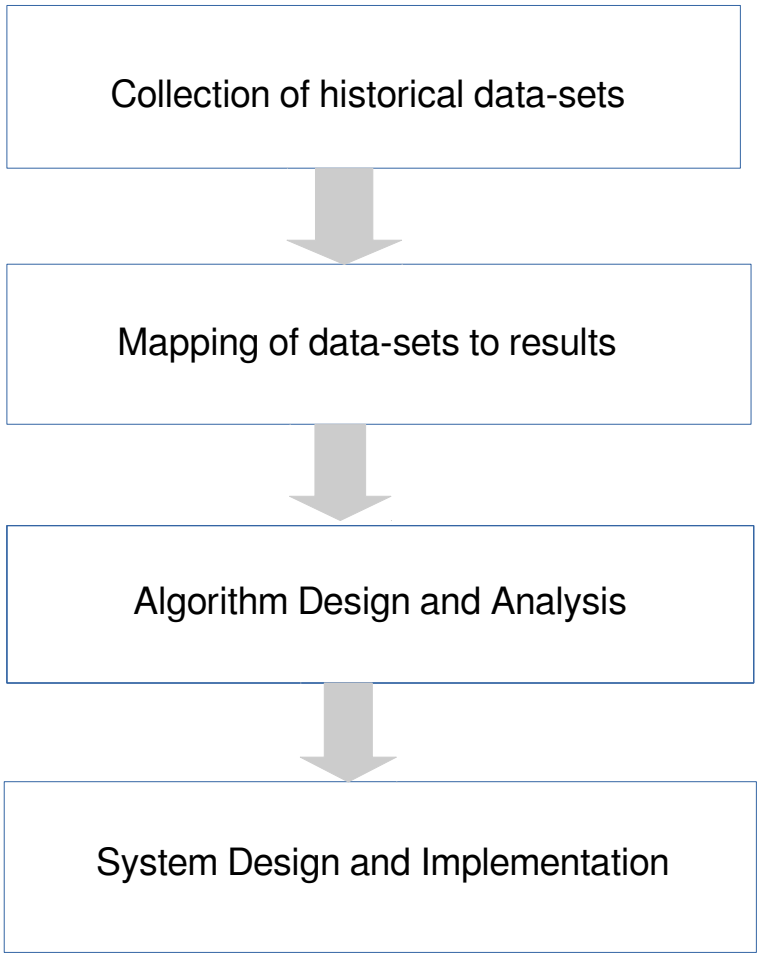

Fig:- Flowchart of Methodology

Hardware/Software Components and Technoligies

- A software to implement algorithms, Ex – Matlab, Octave.
- An IDE to program.Ex – Vim, Code-Blocks, Eclipse.
- Programming Language – Python
- Storage – Amazon Web Services or Local Database.
- A system capable of compiling a software.
- Version Control System – GIT

Non-Technical Requirements

- Historical records of patients with results.
- Study of Medical analysis on a selected disease.

Existing Systems

- There exist a popular tool named “PREDICT” which was focused on predicting Breast Cancer specifically. It's use cases have been successsfull in recent days.

Bibliography

- Machine Learning by Andrew Ng – Stanford University
- List of Diseases to be selected from - http://www.yourarticlelibrary.com/diseases/top-13-common-communicable-diseases-found-in-india/34984/