

# Seminar I

## Predicting the Unpredictable

### A General Disease Prediction System

Hitesh Mishra & Riya Choudhary

Silicon Institute of Technology , Sambalpur

January 3, 2024

# Table of Contents

- Introduction
- How it Works
- Types
- Diabetes Prediction System
- Heart Disease Prediction System
- Benefits
- Challenges
- Future Developments
- Conclusion

# Abstract

Now-a-days, people face various diseases due to the environmental condition and their living habits. So the prediction of disease at earlier stage becomes important task. But the accurate prediction on the basis of symptoms becomes too difficult for doctor. The correct prediction of disease is the most challenging task. To overcome this problem data mining plays an important role to predict the disease. Due to increase amount of data growth in medical and healthcare field the accurate analysis on medical data which has been benefits from early patient care. With the help of disease data, data mining finds hidden pattern information in the huge amount of medical data. We proposed general disease prediction based on symptoms of the patient. For the disease prediction, we use K-Nearest Neighbor (KNN) and Convolutional neural network (CNN) machine learning algorithm for accurate prediction of disease.

# Motivation

- Most hospitals today employ some sort of hospital information systems to manage their healthcare or patient data.
- These systems typically generate huge amounts of data which take the form of numbers, text, charts and images.
- Unfortunately, these data are rarely used to support clinical decision making. There is a wealth of hidden information in these data that is largely untapped. This raises an important question: “How can we turn data into useful information that can enable healthcare practitioners to make intelligent clinical decisions?” This is the main motivation for this research

# Introduction

- The General Disease Prediction System is an innovative technology that uses advanced artificial intelligence algorithms to predict the occurrence of diseases in individuals.
- This system is designed to analyze various factors such as age, medical history, lifestyle habits, and environmental conditions to provide accurate predictions

## How it Works

- The General Disease Prediction System works by collecting and analyzing large amounts of data from various sources such as medical records, genetic testing, and environmental sensors.
- The system then uses machine learning algorithms to identify patterns and correlations between different variables to accurately predict the likelihood of developing certain diseases.

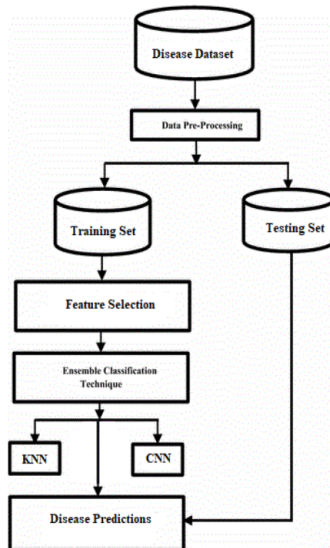
## Prediction System

# General Disease Prediction System

- Artificial Intelligence made computer more intelligent and can enable the computer to think.
- AI study consider machine learning as subfield in numerous research work. Different analysts feel that without learning, insight can't be created.
- There are numerous kinds of Machine Learning Techniques like Unsupervised, Semi Supervised, Supervised, Reinforcement, Evolutionary Learning and Deep Learning.
- These learnings are used to classify huge data very fastly. So we use K-Nearest Neighbor (KNN) and Convolutional neural network (CNN) machine learning algorithm for fast classification of big data and accurate prediction of disease.
- Because medical data is increasing day by day so usage of that for predicting correct disease is crucial task but processing big data is very crucial in general so data mining plays very important role and classification of large dataset using machine learning becomes so easy.



# System Architecture



## Types

# Types of Prediction System

- Diabetes Prediction System
- Heart Disease Prediction System

# Diabetes Prediction System

- Diabetes is one of deadliest diseases in the world. It is not only a disease but also a creator of different kinds of diseases like heart attack, blindness, kidney diseases, etc. Diabetes Mellitus is defined as a group of metabolic disorders mainly caused by abnormal insulin secretion and/or action.
- Insulin deficiency results in elevated blood glucose levels (hyperglycemia) and impaired metabolism of carbohydrates, fat and proteins.
- There are two major clinical types :
  - Type 1 diabetes (T1D)
  - Type 2 diabetes (T2D)
- According to the etiopathology of the disorder. T2D appears to be the most common form of diabetes, mainly characterized by insulin resistance.
- The main causes of T2D include lifestyle, physical activity, dietary habits and heredity, whereas T1D is thought to be due to autoimmune destruction of the Langerhans islets hosting pancreatic beta cells.

## Working Principle

- Machine learning is the scientific field dealing with the ways in which machines learn from experience. The purpose of machine learning is the construction of computer systems that can adapt and learn from their experiences.
- We have developed a system using data mining which has the ability to predict whether the patient has diabetes or not.
- Furthermore, predicting the disease early leads to treating the patients before it becomes critical. Data mining has the ability to extract hidden knowledge from a huge amount of diabetes-related data.
- The aim of this research is to develop a system which can predict the diabetic risk level of a patient with a higher accuracy.
- This research has focused on developing a system based on three classification methods namely, Support Vector Machine, Logistic regression and Artificial Neural Network algorithms.

# Heart Disease Prediction System

- Heart disease, alternatively known as cardiovascular disease, encases various conditions that impact the heart and is the primary basis of death worldwide over the span of the past few decades.
- It associates many risk factors in heart disease and a need of the time to get accurate, reliable, and sensible approaches to make an early diagnosis to achieve prompt management of the disease.
- Data mining is a commonly used technique for processing enormous data in the healthcare domain.
- Researchers apply several data mining and machine learning techniques to analyse huge complex medical data, helping healthcare professionals to predict heart disease.
- This project will be presenting various attributes related to heart disease, and the model on basis of supervised learning algorithms as Naïve Bayes, decision tree, K-nearest neighbor, and random forest algorithm.

## Working Principle

- For predicting heart attack, significantly 15 attributes are listed and with basic data mining technique other approaches e.g. ANN, Time Series, Clustering and Association Rules, soft computing approaches etc. can also be incorporated.
- The outcome of predictive data mining technique on the same dataset reveals that Decision Tree outperforms and some time Bayesian classification is having similar accuracy as of decision tree but other predictive methods like KNN, Neural Networks, Classification based on clustering are not performing well.
- Decision Tree and Bayesian Classification further improves after applying genetic algorithm to reduce the actual data size to get the optimal subset of attribute sufficient for heart disease prediction.

## K-Neighbors Classifier Pseudo code

```
KNeighborsClassifier(X,y, x) // X: training data, y: class labels of X, x: unknown sample
i=1
do
Calculate distance  $D(X_i, x)$ 
while( $i \leq m$ )
Compute set Z having indices for the k smallest distances  $D(X_i, x)$ .
Return label with  $\{y_i \text{ where } i \in Z\}$ 
```



## Features

# Features

- Accurately predicts the spread of diseases.
- It has the ability to analyze large amounts of data in real-time.
- Predict future outbreaks of diseases.
- It has ability to integrate with other healthcare systems, such as electronic health records and hospital information systems.
- This enables it to access a patient's medical history and genetic makeup, which can be used to predict their likelihood of developing certain diseases.
- User-friendly interface that makes it easy for healthcare professionals to use.
- The system provides visualizations and dashboards that allow users to quickly and easily interpret complex data.
- It also has built-in analytics tools that enable users to conduct further analysis and refine their predictions over time.

## Applications

# Applications

- The general disease prediction system has a wide range of applications in the healthcare industry. One of the primary applications is in disease surveillance, where it can be used to monitor and track the spread of infectious diseases.
- The system can also be used to predict outbreaks of diseases based on historical data and current trends.
- Another application of the system is in personalized medicine. By analyzing a patient's genetic makeup and medical history, the system can predict their likelihood of developing certain diseases and recommend personalized treatment plans.
- This can lead to better patient outcomes and more effective use of healthcare resources.

## Future Developments

# Future Developments

- As technology continues to advance, the General Disease Prediction System will likely become even more accurate and effective.
- New developments in wearable devices, genetic testing, and artificial intelligence will enable the system to provide even more personalized and precise prediction

## Literature Review

# Literature Review

Ref. No	Author(s)	Year	Publisher	Objective	Proposed Model	Impact
1	Dhiraj , Gajanan , Ektaa	2010	IEEE	Design a Disease prediction model using dataset for disease prediction	Disease prediction by machine learning over big data from healthcare communities	The proposed general disease prediction based on symptoms of the patient. For the disease prediction, we use K-Nearest Neighbor (KNN) and Convolutional neural network (CNN)
2	B.Qian,X. Wang,N. Cao,H. Li	2012	Springer	A relative similarity based method for interactive patient risk prediction	Heart Disease Prediction model	A prototype heart disease prediction system is developed using three data mining classification modeling techniques.
3	IM. Chen,Y. Ma, Y. Li, D. Wu	2017	IEEE	A Prediction model to predict diabetes in female patients	Diabetes Prediction Model for female patients	A prototype diabetes prediction model is developed which using decision trees and random forest algorithm to predict the missing values. The model has an accuracy score of 74.4



## Conclusion

# Conclusion

- The General Disease Prediction System represents a major breakthrough in the field healthcare, offering numerous benefits for patients and healthcare providers alike.
- While there are still challenges to be addressed, the potential impact of this technology on disease prevention and treatment is immense

## References

## References

- Weiguo, F., Wallace, L., Rich, S., Zhongju, Z.: "Tapping the Power of Text Mining", Communication of the ACM. 49(9), 77-82, 2006.
- Wu, R., Peters, W., Morgan, M.W.: "The Next Generation Clinical Decision Support: Linking Evidence to Best Practice", Journal Healthcare Information Management. 16(4), 50-55, 2002
- Chapman, P., Clinton, J., Kerber, R. Khabeza, T., Reinartz, T., Shearer, C., Wirth, R.: "CRISP-DM 1.0: Step by step data mining guide", SPSS, 1-78, 2000.
- Fayyad, U: "Data Mining and Knowledge Discovery in Databases: Implications from scientific databases", Proc. of the 9th Int. Conf. on Scientific and Statistical Database Management, Olympia, Washington, USA, 2-11, 1997
- Weiguo, F., Wallace, L., Rich, S., Zhongju, Z.: "Tapping the Power of Text Mining", Communication of the ACM. 49(9), 77-82, 2006.

Thank You!!