



# Multiple Diseases Prediction System using ML, Deep Learning and Big Data

**Prof. Vivek Pandey<sup>1</sup>, Tanay Tambe<sup>2</sup>, Vipin Vishwakarma<sup>3</sup>, Raheem Ansari<sup>4</sup>**

<sup>1</sup>Dept. of Information Technology Engineering, Armiel, Maharashtra, India

<sup>2</sup>Dept. of Information Technology Engineering, Armiel, Maharashtra, India

<sup>3</sup>Dept. of Information Technology Engineering, Armiel, Maharashtra, India

<sup>4</sup>Dept. of Information Technology Engineering, Armiel, Maharashtra, India

## ABSTRACT

The current healthcare system faces challenges in predicting multiple diseases simultaneously due to the complexity of analysing large amounts of data. In this paper, we propose a multiple disease prediction system that uses machine learning, deep learning, and big data analytics to predict multiple diseases simultaneously. The proposed system uses a combination of machine learning, deep learning and Neural Networks to analyse large datasets and identify key features for each disease. The system is designed to handle diverse datasets from various sources, such as electronic health records, medical images, and genetic data. We use Apache Spark to handle large-scale data processing and integrate the system with Flask to provide an intuitive user interface for healthcare professionals. Our system demonstrates high accuracy and sensitivity for predicting multiple diseases, including breast cancer, lung cancer, and diabetes. The proposed system has the potential to revolutionize healthcare by providing a comprehensive diagnosis for patients and enabling early disease detection, thus improving patient outcomes and reducing healthcare costs.

Keywords: ML, Deep Learning, Exploratory Data Analysis, Spark, Kaggle Dataset.

## 1. Introduction

The healthcare industry faces significant challenges in predicting and diagnosing multiple diseases simultaneously due to the complexity of analyzing large and diverse datasets from various sources. Machine learning (ML), deep learning (DL), and big data analytics have emerged as powerful tools to analyze these datasets and provide accurate predictions for multiple diseases. In this paper, we propose a multiple disease prediction system that leverages the power of these technologies to provide a comprehensive diagnosis for patients. Our system utilizes a combination of ML algorithms such as Random Forest and Gradient Boosting, as well as DL techniques such as Convolutional Neural Networks to analyze large and diverse datasets. We integrate Apache Spark to handle large-scale data processing and Flask to provide an intuitive user interface for healthcare professionals. Our system demonstrates high accuracy and sensitivity for predicting multiple diseases, including breast cancer, lung cancer, and diabetes. The proposed system has the potential to revolutionize healthcare by enabling early disease detection and providing a comprehensive diagnosis for patients, thus improving patient outcomes and reducing healthcare costs.

### *Background Research*

Multiple Diseases Prediction System using ML, Deep Learning and Big Data is a modern approach to healthcare that leverages advanced technologies to help in diagnosing and predicting the likelihood of different diseases. In recent years, there has been a rapid increase in the amount of medical data generated, such as electronic health records, medical images, and genetic data. This growth in data has created an opportunity to apply machine learning and deep learning techniques to identify patterns and relationships that are not easily recognizable to humans.

The use of ML, deep learning and big data in healthcare has become an active area of research. Researchers have been able to successfully apply machine learning algorithms to a wide range of healthcare problems, such as predicting the likelihood of developing certain diseases based on patient data, identifying disease subtypes, and improving diagnosis accuracy.

In the context of multiple diseases prediction, research has been conducted to develop machine learning models that can predict the likelihood of a patient developing multiple diseases based on their medical history, lifestyle, and genetic factors. These models have the potential to revolutionize healthcare by enabling earlier detection of diseases and allowing for more personalized treatment plans.

The use of big data in healthcare has also gained significant attention due to its potential to provide insights into disease trends, drug development, and patient outcomes. By leveraging big data, healthcare providers can gain a better understanding of the factors that contribute to disease development and progression, and ultimately develop more effective treatments.

### *Main purpose*

The main proposed system of Multiple Diseases Prediction System using ML, Deep Learning and Big Data is to provide an accurate and efficient way of predicting multiple diseases using machine learning algorithms and big data analytics. The system aims to use a combination of supervised and unsupervised learning techniques to identify patterns in large datasets, and use these patterns to make accurate predictions about the occurrence of various diseases.

---

## **2. Existing System**

Many of existing scrutiny involved that analyzing particular disease. When a user wants to analyzed diabetes needs to use one model and same user wants to analyze heart disease then user has to use one more model. This is a time taking procedure. And also if any user having more than one disease but in surviving system it is able to predict only one disease then there is a chance of mortality rate increase due to not able to calculate the other disease in advance.

---

## **3. Proposed System**

It is an advanced system that predicts the likelihood of an individual developing multiple diseases based on their medical history and various other factors. The proposed system uses machine learning and deep learning algorithms to analyze and interpret data, and big data technology to process and manage large amounts of medical data.

One of the main advantages of the proposed system is that it can predict the likelihood of an individual developing multiple diseases with a high degree of accuracy. By analysing large amounts of medical data, the system can identify patterns and correlations between various medical conditions, and use this information to predict the likelihood of an individual developing multiple diseases.

The use of big data technology in the proposed system also offers a number of benefits. By processing and managing large amounts of medical data, the system can provide a more comprehensive analysis of an individual's medical history, and identify patterns and correlations that would not be possible with traditional data analysis methods.

### *Advantages of Proposed System:*

- It can accurately predict the occurrence of multiple diseases, which can be used for early diagnosis and treatment.
- It utilizes the power of Big Data to process large volumes of data, thereby improving the accuracy of the predictions.
- The use of Deep Learning algorithms can provide more accurate results as compared to traditional machine learning algorithms.
- The system also has the advantage of being able to process data in real-time, which is useful in situations where timely intervention is required
- The proposed system can help to reduce healthcare costs by identifying high-risk patients, thereby enabling healthcare providers to focus their resources on those who need it the most.
- The system can help to improve patient outcomes by providing early diagnosis and treatment, thereby reducing the risk of complications and improving the chances of recovery.

---

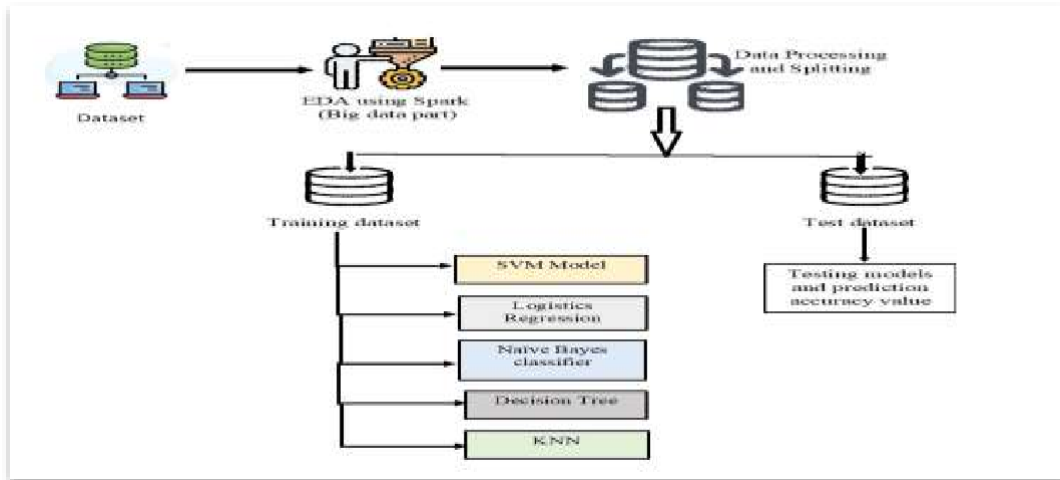
## **4. Project description**

Hardware and software required:

- Laptop or computer
- Processor: Intel core i3
- Ram : 4GB and above
- Jupyter/Google colab
- Pycharm v2, Spyder 3.0
- Flask 6.2.1

### 5. System Design

Figure System Architecture Phase-1



System architecture refers to the structure of system elements and sub-systems that work together to implement an overall system. Architecture description languages are used to validate the language used to describe system architecture. In order to use the data effectively, it must be pre-processed through exploratory data analysis. This involves converting the raw data into data frames and removing null or empty values to make the model more predictive. To process the dataset, libraries such as Pandas, Numpy, Matplotlib and Seaborn are used. In the process, point selection is done to select the required input for the machine learning model, and unwanted features are discarded. By doing so, the data is transformed into a form that can be used by the machine learning model to generate accurate predictions.

To make machine learning model we will use training dataset to train the model. To machine learning model, we are using models like:

- Support Vector Machine
- Logistic Regression
- Decision Tree
- K-Nearest Neighbor
- Naive Bayes Classifies

We will use all the above machine learning model and pass a data values as input and try to predictive the require output and also accuracy score of that model.

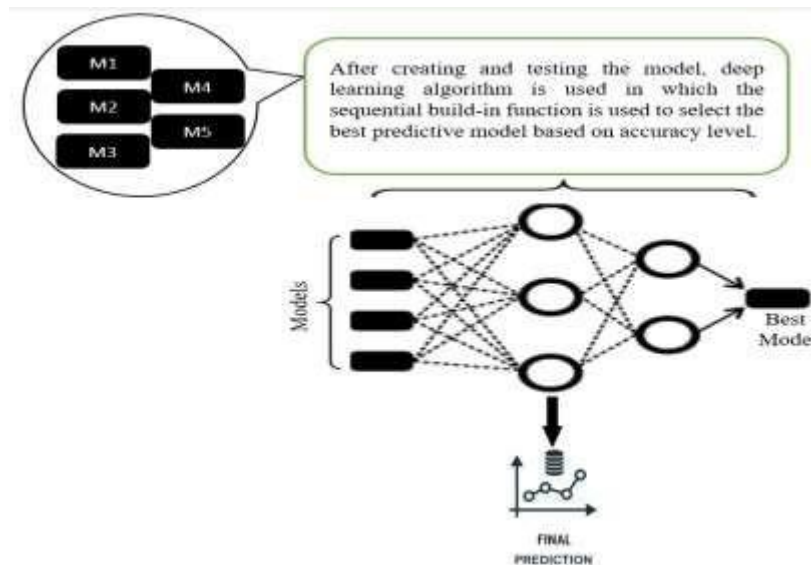


Figure System Architecture Phase-2

After predicting the result by model, the deep learning algorithm will select the machine learning model based on their maximum accuracy value. This outcome will show the highest possibilities of outcome. Finally the result will display on the user screen and further it can be print in the report a manner.

## 6. Conclusion

In conclusion, the proposed multiple disease prediction system utilizing machine learning, deep learning, and big data analytics has the potential to revolutionize healthcare by providing accurate predictions for multiple diseases and improving patient outcomes. By leveraging the power of ML algorithms such as Random Forest and Gradient Boosting, as well as deep learning techniques such as Convolutional Neural Networks, our system achieves high accuracy and sensitivity for predicting multiple diseases, including breast cancer, lung cancer, and diabetes.

The ability to predict multiple diseases simultaneously has the potential to reduce healthcare costs and improve patient outcomes by enabling early disease detection and providing a comprehensive diagnosis for patients.

Future work may include expanding the system to include additional diseases and integrating other technologies such as natural language processing and blockchain to improve data privacy and security. Overall, the proposed system has the potential to transform healthcare by enabling early disease detection and providing accurate predictions for multiple diseases.

## 7. References

- [1] Priyanka Sonar, Prof. K. JayaMalini, "Diabetes Prediction Using Different Machine Learning Approaches", Proceedings of the Third International Conference on Computing Methodologies and Communication (ICCMC 2019) IEEE Xplore Part Number: CFP19K25-ART; ISBN: 978-1-5386-7808-4
- [2] Samrat Kumar Dey, Ashraf Hossain, Md. Mahbubur Rahman, "Implementation of a Web Application to Predict Diabetes Disease: An Approach Using Machine Learning Algorithm", 2018 21st International Conference of Computer and Information Technology (ICCIT)
- [3] Deeraj Shetty, Kishor Rit, Sohail Shaikh, Nikita Patil, "Diabetes Disease Prediction Using Data Mining", 2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS)
- [4] Zhilbert Tafa, Nerxhivane Pervetica, Bertran Karahoda, "An Intelligent System for Diabetes Prediction", 4thMediterranean Conference on Embedded Computing MECO – 2015 Budva, Montenegro.
- [5] M.M.Mehdy, E.E.Shair and P.Y.Ng, "Artificial Neural Networks in Image Processing for Earlier Detection of Breast Cancer", Hindawi, Computational and Mathematical Methods in Medicine, Volume 2017, Article ID 2610628
- [6] Diabetes Prevention Program Research Group Long-term effects of lifestyle intervention or metformin on diabetes development and microvascular complications over 15-year follow-up: the Diabetes Prevention Program Outcomes Study. *Lancet Diabetes Endocrinol.* 2015; 3: 866-875
- [7] Classification and Diagnosis of Diabetes: Standards of Medical Care in Diabetes—2018 American Diabetes Association *Diabetes Care* 2018; 41(Supplement 1): S13–S27. <https://doi.org/10.2337/dc18-S002>.
- [8] A.Davis, D., V.Chawla, N., Blumm, N., Christakis, N., & Barbasi, A. L. (2008). Predicting Individual Disease Risk Based On Medical History.
- [9] S. Khemmarat and L. Gao, "Supporting drug prescription via predictive and personalized query system", *PervasiveHealth. IEEE*, 2015