

# Disease Prediction using Machine Learning



S.Sai Kumar, B.V.Subba Rao, J.Sirisha, B. Anusha, A. Sai Akhil, K.Venkata Suhesh, K.Bhavana

**Abstract:** As global population is increasing, life expectancy rises, and standards of living increases, and the causes of death across the world are changing. Now-a-days most of the deaths are occurring due to the lack of awareness among people about which disease may have the chance to occur for a particular symptom. So by taking the previous datasets of patients and analyzing it, we can predict which disease may have the scope to occur for a particular symptom. It creates an awareness among people about the disease and appropriate medication can be received. By this the death rate can be minimized. So, in this system, we are going to analyse the patient datasets by using machine-learning(linear regression) algorithms and python code to predict the diseases which have the chance to occur and so that we can reduce the death rate.

**Keywords:** Disease Prediction, Machine Learning, Linear Regression.

## I. INTRODUCTION

From the past decagon a lot of research has been going on “disease prediction” using machine learning techniques. Many researchers implemented it using data mining techniques to predict the disease from existing symptoms. These approaches tried to predict the reoccurrence of the disease. But in recent days, due to the success of deep learning in areas of machine learning has driven a shift towards machine learning models to predict the diseases. By using this approach we can execute the system with little preprocessing of raw data and the data can be represented hierarchically. So that we can get the most accurate results.

Machine can predict diseases but cannot predict the sub types of the diseases caused by occurrence of one disease. It fails to predict all possible conditions of the people. Existing system handles only structured data. The prediction system are broad and ambiguous. In current past, countless disease estimate classifications have been advanced and in procedure. The standing organizations arrange a blend of machine learning algorithms which are judiciously exact in envisaging diseases. However the restraint with the prevailing systems is speckled. First, the prevailing systems are dearer only rich people could pay for to such calculation systems. And also, when it comes to folks, it becomes even higher. Second, the guess systems are non-specific and indefinite so far. So that, a machine can envisage a positive disease but cannot expect the sub types of the diseases and diseases caused by the existence of one bug. For occurrence, if a group of people are foreseen with Diabetes, doubtless some of them might have complex risk for Heart viruses due to the actuality of Diabetes. The remaining schemes fail to foretell all possible surroundings of the tolerant[1][2].

### Disadvantages:

- High cost.
- Results are not accurate.

## II. PROPOSED WORK

Our application will be at affordable cost. Linear Regression Machine Learning Algorithm predicts Diseases as well as all sub diseases. Accuracy is improved using Machine Learning algorithm. The proposed system begin with the thought that was not executed by the ancestors. In this we go for subdividing the data such that a request would be scrutinized only in the explicit partition, which will increase effective proficiency but cut query rescue time. In tally to that, it provide definite rations for specific clients to pattern his/her condition. Thus making our presentation broadly open for all at cheap cost[3].

### Advantages:

- Low cost.
- Accuracy is high

## III. APPROACH

### Step 1: Collection of Data and Preparation of Dataset

This step includes collection of patient data from various hospitals, their medication slips etc. Once the data is collected, redundant data is removed by preprocessing. Finally, the dataset is implemented in an organized manner.

Manuscript received on March 15, 2020.

Revised Manuscript received on March 24, 2020.

Manuscript published on March 30, 2020.

\* Correspondence Author

**Dr.S.Sai Kumar\***, Assistant Professors, Department of Information Technology, PVPSIT, Vijayawada, AP, India.

**Dr.B.V.Subba Rao**, <sup>2</sup>Professor & Head, Department of Infromation Technology, PVPSIT, Vijayawada, AP, India.

**J.Sirisha**, Assistant Professors, Department of Information Technology, PVPSIT, Vijayawada, AP, India.

**B.Anusha**, B.Tech Students, Department of Information Technology, PVPSIT, Vijayawada, AP, India.

**Sai Akhil**, B.Tech Students, Department of Information Technology, PVPSIT, Vijayawada, AP, India.

**K.Venkata Suhesh**, B.Tech Students, Department of Information Technology, PVPSIT, Vijayawada, AP, India.

**K.Bhavana**, B.Tech Students, Department of Information Technology, PVPSIT, Vijayawada, AP, India.

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](https://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

## Disease Prediction using Machine Learning

### Step 2: Developing a linear regression model in machine learning

Here, in this step, machine learning approach based on linear regression modeling is developed.

And decision trees are generated which is used to deal with large amount of data.

### Step 3: Training and Testing on datasets

The disease predicting model will be trained on the dataset and tested to predict the disease more accurately based on symptoms.

### Step 4: Deployment and analysis

The trained and tested disease prediction model will be deployed in a real-life scenario made by the human experts, and the model can be further improved in the methodology. The model will follow the above architecture[4][5].

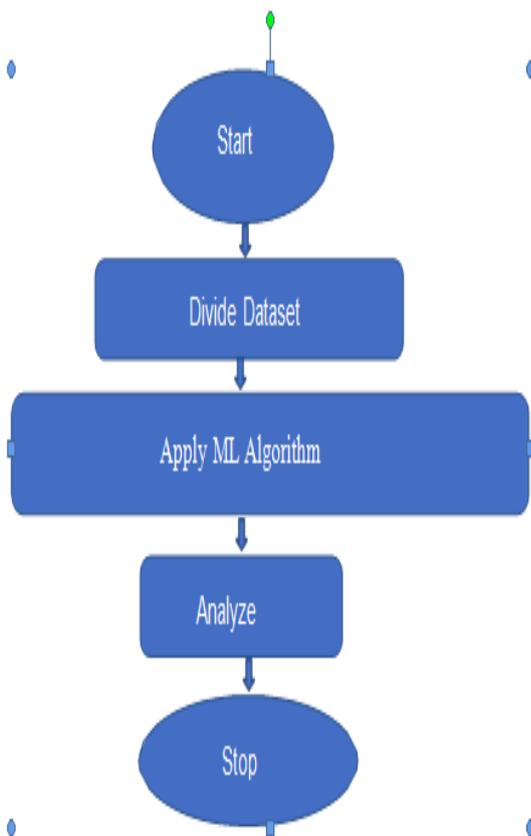


Fig.No.1: Work flow of disease prediction

## IV. MACHINE LEARNING

Artificial Intelligence can make the computer to think and take decisions appropriately. Computers are made more intelligent and they are even able to think like humans using artificial intelligence.

There are different types of machine learning techniques which are used to classify the datasets. They are:

- Supervised Learning
- Unsupervised Learning
- Semi-Supervised Learning

- Reinforcement Learning
- Evolutionary Learning
- Deep Learning

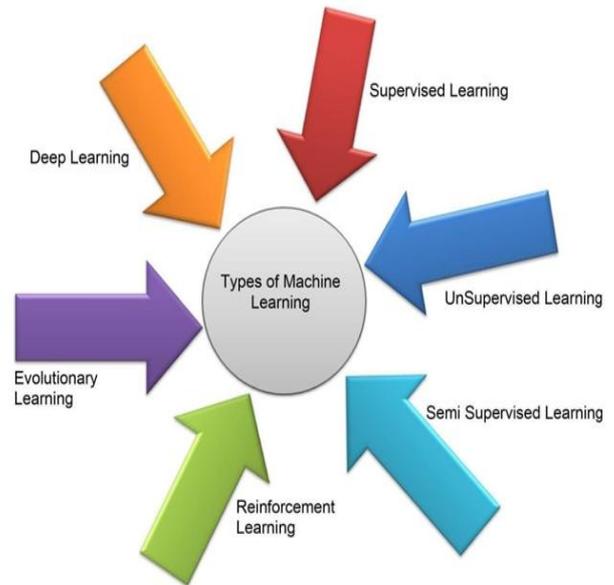


Fig No.:2 Types of Machine Learning

### Supervised Learning:

In this, we provide a training dataset for prediction. Based on this training dataset, algorithms are implemented to display the outputs for the given inputs. Supervised Learning is also known as “Learning from Exemplars”.

There are 2 types of supervised learning. They are

- Classification
- Regression

**Classification:** It will give the prediction value as “Yes/No”.

Example: “Is this a banana?”, “Are curries healthy?”

**Regression:** It gives the answer of “How much” and “How many”.

Example: “How many apples are there?”

### Unsupervised Learning:

In this we do not provide any training set for prediction. In unsupervised learning technique the system try to find out the common characteristics between the given input data and based on these similarities, unsupervised learning technique classifies the data.

Unsupervised Learning is also known as “Density Estimation”.

There are 2 types of unsupervised learning

- Clustering
- Association

**Clustering:** It makes clusters based on the similarities.

**Association:** We discover rules that describe large portions of your data.

### Semi-Supervised Learning:

Semi-Supervised learning is a category of Supervised learning techniques. This technique uses both labeled and unlabeled data for training purpose. It is a combination of both supervised-learning (labeled data) and unsupervised learning (unlabeled data).

**Reinforcement Learning:**

This algorithm receives input based on experience. Algorithm is designed to inform whenever the answer is wrong, but it does not provide any information of about how to correct it. It has to analyse and check for various possibilities until it finds the correct answer and it does not support any further improvements.

Reinforcement Learning is also named as “Learning with a Critic”.

Reinforcement learning is different from supervised learning, in the way that accurate input & output sets are not provided, nor sub-optimal actions clearly defined.

**Evolutionary Learning:**

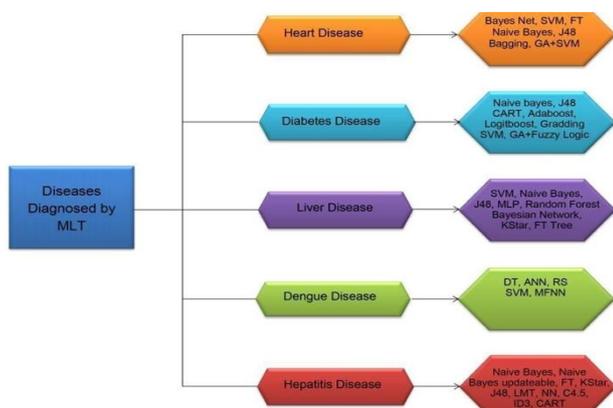
Evolutionary learning can be termed as biological learning. In general, biological organisms are clearly adjusted to make progress in their life durability rates and the possibilities of having new generations. By using this idea of fitness, we can use this model in a computer to check the accuracy of the solution.

**Deep learning:**

In deep learning, it consists of a set of algorithms. By using these algorithms abstraction can be achieved. It consists of deep layered graph with different types of processing layers which were made up of 2 different types of transformation like linear and non linear transformations. Pattern recognition process and data classification has its own importance from the past decagon. Data set is sub-divided into two or more classes. Such classifiers and classification methods are used for medical dataset analysis and disease prediction.

**Disease Prediction by using various machine learning algorithms:**

From the past decagon many researchers are working on various machine learning algorithms for predicting a disease and believed that these machine learning algorithms are the best to work with various disease predictions. The following figure shows the approach of predicting a disease by using various Machine Learning algorithms [6].



**Fig.No.3: Different algorithms used to predict disease in machine learning**

**Linear Regression**

A statistical approach which is used for modeling a relationship between a dependent variable and an independent variable is known as “Linear Regression”. In which, dependent variables are termed as “response”, whereas,

independent variables are termed as “features” for simple understanding. Simple Linear Regression is the basic version of linear regression and it also provides the most basic understanding of the algorithm [7].

**Simple Linear Regression**

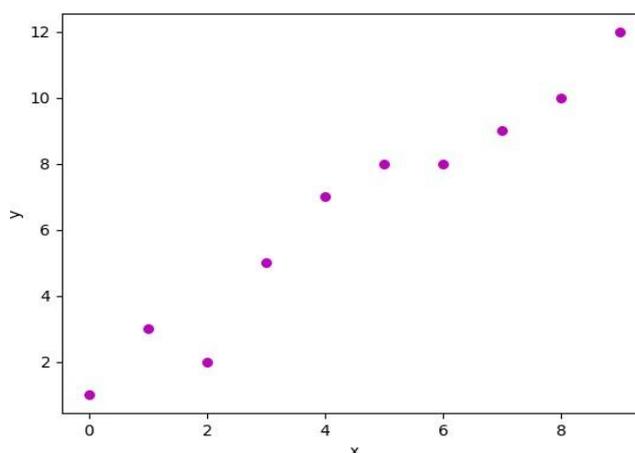
As an approach to predict a “response” using a “single feature” we use “Simple Linear Regression” algorithm. In this technique, we assume that the relation between two variables is linear. Therefore, we always look forward for finding a linear function which is used in predicting the response value(y) with more accuracy as a function of the feature or independent variable(x). Let us consider a dataset where the value of response is termed as y for every feature of x:

x	0	1	2	3	4	5	6	7	8	9
y	1	3	2	5	7	8	8	9	10	12

For understanding, we define:

A feature vector as x , where x = [x1, x2, ..... , xn], and a response vector as y, where y = [y1, y2, ..... , yn] where n is the number of observations (from the above example, we consider n=10).

The following shows the scatter plot for the above dataset:



Now, we have to find a best line which fits for the above scatter plot, so that the prediction of the response for any new feature values will be effective(i.e a value of x does not exist in the dataset). This line is termed or known as “Regression Line”[8].

Regression Line can be represented in the form of equation as follows:

$$h(x_i) = \beta_0 + \beta_1 x_i$$

Here, in the above equation,

- h(x<sub>i</sub>) is used to represent the response value which is predicted for each ith observation.
- β<sub>0</sub> and β<sub>1</sub> are termed as the regression coefficients and in which, β<sub>0</sub> is represented as y-intercept and β<sub>1</sub> is represented as slope of regression line.

We must estimate the values of regression coefficients β<sub>0</sub> and β<sub>1</sub>,

in order to create our model, after we estimate these coefficients, the model is further used to predict the responses.

Now consider:



```
C:\Users\DELL\OneDrive\Desktop\Akhil_mini>python inp.py
Enter 1 if you have disease else 0 : 0
Enter 1/0 if you have Shortness of breath : 0
Enter 1/0 if you have Chest pain : 0
Enter 1/0 if you have Increased sweating : 0
Enter 1/0 if you have Unresponsiveness : 0
Enter 1/0 if you have mental status changes : 0
Enter 1/0 if you have Feeling suicidal : 0
Enter 1/0 if you have Auditory Hallucinations : 1
Enter 1/0 if you have Visual Hallucinations : 1
Enter 1/0 if you have Sleeplessness : 1
Enter 1/0 if you have Irritable mood : 0
Enter 1/0 if you have Slowing of urinary stream : 0
Enter 1/0 if you have Fatigue : 0
Enter 1/0 if you have Hypotension : 0
Enter 1/0 if you have Extreme exhaustion : 0
Enter 1/0 if you have Prostatism : 0
Enter 1/0 if you have Wheezing : 0
Enter 1/0 if you have Cough : 0
Enter 1/0 if you have Chest tightness : 0
Enter 1/0 if you have Pleuritic pain : 0
Enter 1/0 if you have Distress respiratory : 0
Enter 1/0 if you have Abdominal pain : 0
Enter 1/0 if you have Vomiting : 0
Enter 1/0 if you have Nervousness : 0
Enter 1/0 if you have Hunger : 0
Enter 1/0 if you have Out of breath : 0
[0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]
Successful#####
```

**Description:** Taking the input in the form of symptoms from the user.

```
C:\Users\DELL\OneDrive\Desktop\Akhil_mini>python prediction.py
3
Diabetis
depression
kidneyfailure
```

**Description:** Predicting output in the form of disease name using prediction.csv.

```
C:\Users\DELL\OneDrive\Desktop\Akhil_mini>python inp.py
Enter 1 if you have disease else 0 : 0
Enter 1/0 if you have Shortness of breath : 0
Enter 1/0 if you have Chest pain : 0
Enter 1/0 if you have Increased sweating : 0
Enter 1/0 if you have Unresponsiveness : 0
Enter 1/0 if you have mental status changes : 0
Enter 1/0 if you have Feeling suicidal : 0
Enter 1/0 if you have Auditory Hallucinations : 0
Enter 1/0 if you have Visual Hallucinations : 0
Enter 1/0 if you have Sleeplessness : 0
Enter 1/0 if you have Irritable mood : 0
Enter 1/0 if you have Slowing of urinary stream : 1
Enter 1/0 if you have Fatigue : 1
Enter 1/0 if you have Hypotension : 1
Enter 1/0 if you have Extreme exhaustion : 0
Enter 1/0 if you have Prostatism : 1
Enter 1/0 if you have Wheezing : 0
Enter 1/0 if you have Cough : 0
Enter 1/0 if you have Chest tightness : 0
Enter 1/0 if you have Pleuritic pain : 0
Enter 1/0 if you have Distress respiratory : 0
Enter 1/0 if you have Abdominal pain : 0
Enter 1/0 if you have Vomiting : 0
Enter 1/0 if you have Nervousness : 0
Enter 1/0 if you have Hunger : 0
Enter 1/0 if you have Out of breath : 0
[0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]
successful#####
```

**Description:** Taking the input in the form of symptoms from the user.

```
C:\Users\DELL\OneDrive\Desktop\Akhil_mini>python prediction.py
3
Diabetis
depression
kidneyfailure
```

**Description:** Predicting output in the form of disease name using prediction.csv.

## VI. CONCLUSION

The Disease Prediction using Machine Learning is designed especially for the convenience of the users to save their valuable time and money. In this project we have used a Machine Learning algorithm like Linear Regression and implemented it using Python Programming Language, to predict the disease by taking symptoms as the input. The users of this project can easily enter the symptoms in the form of 0's & 1's i.e., '0' if symptom does not exist and '1' if the symptom exists. By using linear regression we predict the disease based on the provided symptoms and provide the output with disease name.

Based on the encouraging results from Disease Prediction using Machine Learning project and its benefits to the users in saving their time, we have created a user-friendly interface. We gathered lot of information regarding the flaws, difficulties that the users are facing, using the conventional methods. This project is useful to many people as it minimizes the unnecessary tests in hospitals, so that the user could save money and time. The user may check the scope of disease to occur through our project and undergo the tests accordingly. The user may be able to enter any number of symptoms and check for the disease which have the scope to occur.

## REFERENCES

- [https://www.researchgate.net/publication/330981991\\_Analysis\\_of\\_Heart\\_Disease\\_Prediction\\_using\\_VariousMachine\\_Learning\\_Techniques](https://www.researchgate.net/publication/330981991_Analysis_of_Heart_Disease_Prediction_using_VariousMachine_Learning_Techniques)
- <https://www.geeksforgeeks.org/linear-regression-python-implementation/> / <https://www.youtube.com/watch?v=NUXdtN1W1FE#action=share>
- <https://youtu.be/NUXdtN1W1FE>
- <https://youtu.be/E5RjzSK0fvY>
- <https://towardsdatascience.com/linear-regression-detailed-view-ea73175f6e86>
- <https://www.youtube.com/watch?v=iUriLVYtk&list=UUbAZq799T9mxxRyGz3XYkA&index=2>  
<https://www.kaggle.com/learn/machine-learning>
- <https://www.expertsystem.com/machine-learning-definition/>  
<https://www.analyticsvidhya.com/blog/2017/09/common-machine-learning-algorithms/>
- Grady Booch, James Rumbaugh, Ivar Jacobson : The Unified Modeling Language User Guide, Pearson
- Education.Rob Pandey, Pauline Wilcox
- Software Engineering, A Precise approach, Pankaj Jalote, Wiley. [13]
- Software Engineering, 6 th Edition, Roger S.Pressman , TMH. Applying UML Advanced Application, Elsevier.
- [filehorse.com/download-python-64/](http://filehorse.com/download-python-64/)

## AUTHORS PROFILE



**Dr. S. Sai Kumar**, is currently working as Assistant Professor in Department of Information Technology at Prasad V. Potluri Siddhartha Institute of Technology Permanently Affiliated to Jawaharlal Nehru Technological University, Kakinada. I completed my Ph.D in Computer Science & Engineering at Krishna University. He has a total of 10 years of rich experience comprising teaching and research. He has published 26 papers in International Journals. His research interests are in the areas like Mobile Computing, Data Mining, IoT, Artificial Intelligence, Machine Learning etc.



**Dr. B.V. Subba Rao**, presently working as Professor and Head of the Department of Information Technology branch in P.V.P Siddhartha Institute of Technology, Vijayawada, affiliated to Jawaharlal Nehru Technological University. He has a total of 18 years of rich experience comprising teaching, research and industry. He has published 35 papers in International Journals and 8 papers in national Journals and presented 12 papers in National /International Conference Proceedings. He published and obtained 2 patents. He has an Academic participation in 34 International / National Seminars / workshops and Conferences. He is an editorial Board member to various National and International journals like IJ-CA-ETS, IJ-ETA-ETS, International Journal of Knowledge and Research in Computer Engineering, International Journal of Computer Science and Information Technology and also reviewer of these journals. His current research interests are in the areas of Artificial Intelligence, Natural Language Processing and Information Storage and Retrieval systems.



**J. Sirisha**, presently working as Asst. Professor in the Department of Information Technology branch in P.V.P Siddhartha Institute of Technology, Vijayawada, affiliated to Jawaharlal Nehru Technological University. She has a total of 14 years of rich experience comprising teaching and research. She has published 14 papers in International Journals and 4 papers in national. Her current research interests are in the areas of Web Mining and Semantic web technologies.