

# ***DISEASE PREDICTION USING MACHINE LEARNING***



*A Project Report Submitted to*  
***SMARTINTERNZ***  
*For Fulfillment*  
*of*  
***EXTERNSHIP CERTIFICATION COURSE***  
*In*  
***APPLIED DATA SCIENCE***

***SUBMITTED BY***  
***Bandewar Sai Raghavendra (20BCE7077)***  
***Rishabh Sharma (20BCE7208)***  
***Dev Kapadia (20BCD7051)***  
***Miriyala Venkata Ramesh (20BCR7133)***

***VELLORE INSTITUTE OF TECHNOLOGY, ANDHRA PRADESH***  
***JULY 2023***

# **1. INTRODUCTION**

## **1.1 OVERVIEW**

*The project "Disease Prediction Using Machine Learning" aims to use advanced data analysis methods and machine learning algorithms to create models that can help predict diseases early on. The project team hopes to develop accurate and effective models that can identify patterns and signs that may indicate the onset of diseases by analyzing a variety of medical data, including patient demographics, clinical records, genetic information, and lifestyle factors.*

- The project aims to create predictive models that can help in early detection and diagnosis of diseases.*
- The project team hopes to develop accurate and effective models.*
- The project team will analyze a variety of medical data to identify patterns and signs that may indicate the onset of diseases.*

## **1.2 PURPOSE**

*This initiative aims to revolutionize healthcare by using the potential of machine learning algorithms to predict diseases. The traditional way of diagnosing diseases relies on symptoms, manual interpretation of medical data, and the knowledge of healthcare experts.*

*However, these methods are not always accurate or timely, which can lead to delayed diagnosis and poor patient outcomes. This initiative seeks to overcome these limitations by providing medical professionals with powerful tools for early diagnosis and intervention using machine learning techniques.*

- The initiative aims to use machine learning to predict diseases.*
- Traditional methods of diagnosing diseases are not always accurate or timely.*
- Machine learning can help to overcome these limitations.*
- The initiative aims to provide medical professionals with powerful tools for early diagnosis and intervention.*

# **2. LITERATURE SURVEY**

## **2.1 EXISTING PROBLEM**

*Traditional methods in healthcare have some limitations. They rely on a manual examination of patient data and professional interpretation, which can be inaccurate, time-consuming, and not scalable to large datasets. It can be difficult for healthcare professionals to identify small patterns and signs in large datasets, which can delay diagnosis and lead to subpar treatment outcomes.*

*Additionally, it is difficult to accurately determine a person's risk of developing a disease and predict future occurrences due to the complexity of disease development and the interaction of multiple risk factors.*

- Traditional methods of diagnosis and forecasting in healthcare are not always accurate.*
- They can be time-consuming and not scalable to large datasets.*
- It can be difficult to identify small patterns and signs in large datasets.*
- It is difficult to accurately determine a person's risk of developing a disease and predict future occurrences.*

## 2.2 PROPOSED SOLUTION

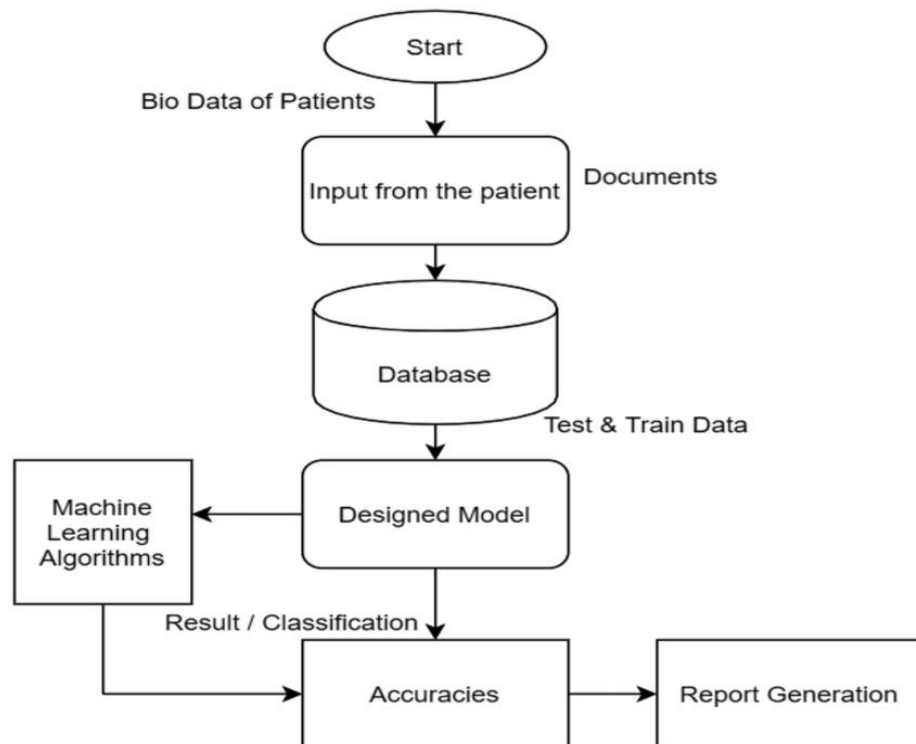
Machine learning is being used to solve the limitations of traditional methods for disease prediction. Machine learning can identify hidden patterns and correlations in large amounts of medical data, which can lead to more accurate and timely disease predictions. This is done by using advanced algorithms and data processing techniques.

The proposed approach involves creating prediction models that include multiple types of data, such as patient demographics, clinical records, genetic data, and lifestyle factors, to provide comprehensive risk assessment tools.

- Machine learning can be used to predict diseases more accurately and timely than traditional methods.
- Machine learning can identify hidden patterns and correlations in large amounts of data.
- Machine learning can be used to create comprehensive risk assessment tools.

## 3. THEORETICAL ANALYSIS

### 3.1 BLOCK DIAGRAM



### 3.2 HARDWARE / SOFTWARE DESIGNING

#### Hardware Design

- The hardware should have enough processing resources to handle the complexity of the machine learning algorithms and the amount of data to be processed.
- The hardware should have enough storage capacity to store and handle massive amounts of medical data.
- The hardware design should be scalable to accommodate the growing data processing needs of the project.
- The hardware should include strong security features to protect patient privacy and comply with data protection laws.

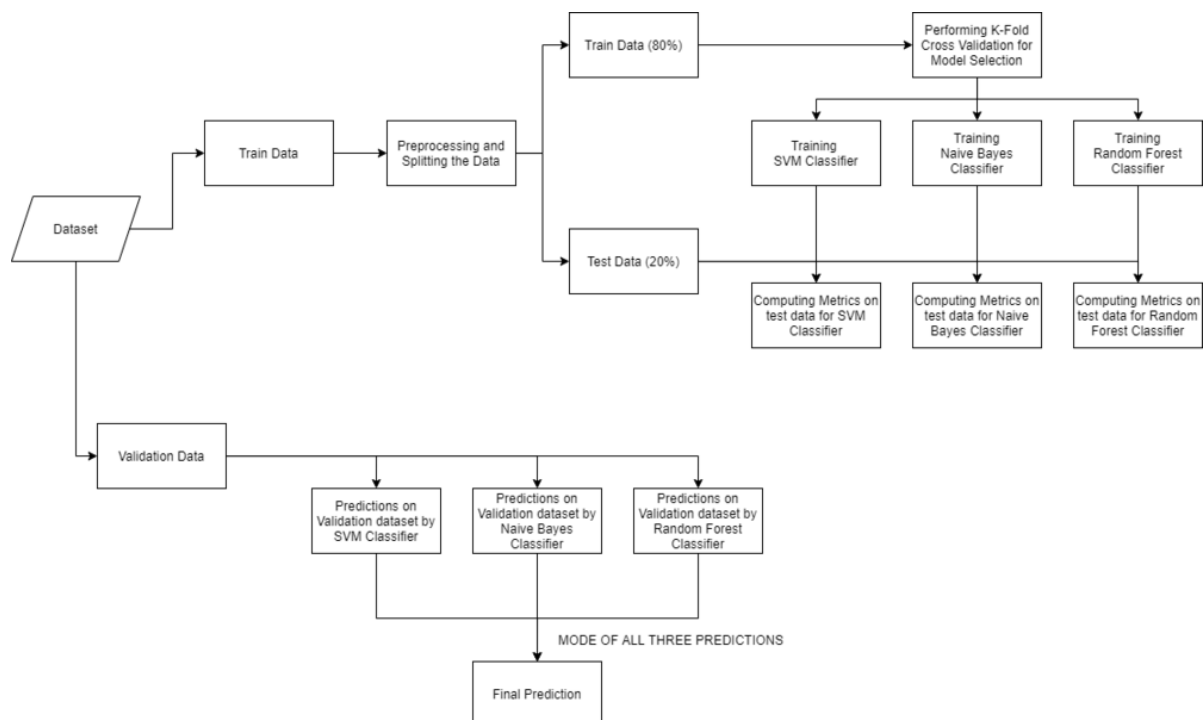
## Software Design

- The software should include modules for data preprocessing, machine learning algorithms, model training and evaluation, and model deployment.
- The data preprocessing modules should handle data cleansing, transformation, and feature extraction.
- The machine learning algorithms modules should support a variety of machine learning techniques, such as deep learning models, decision trees, and classification methods.
- The model training and evaluation modules should use cross-validation procedures and assessment criteria to evaluate the effectiveness of the models.
- The model deployment modules should be able to integrate the models into the healthcare system and make predictions in real-time.

## 4. EXPERIMENTAL INVESTIGATIONS

- The project aims to prove that machine learning can be used to predict diseases.
- The results of the trials support the proposed solution.
- The trials identify areas that need improvement.
- The trials guide future development and use of the disease prediction system.

## 5. FLOWCHART



## 6. RESULT

- We've tested the data with 4 different models namely Support Vector Machine, Random Forest, Logistic Regression, and Decision Tree with 50, 60, 70, 80, and 90 features.
- The best model "Random Forest classifier with 90 features" was able to achieve 99.25% accuracy on the given data set.

- A comparison of different machine learning algorithms gave us the most effective approaches for predicting disease.
- Although this model helps doctors diagnose, it is important to remember that this model is not a substitute for a doctor's expertise.
- The project has limitations such as data availability and can be further improved, such as incorporating additional data sources or improving the model architecture.
- Further research using larger and more diverse datasets is needed to develop machine learning models that can be effectively used in real-world clinical settings.

**Disease Predictor**

Refer to this table to get the predicted disease

Predicted Output	Disease
0	(vertigo) Paroymsal Positional Vertigo
1	AIDS
2	Acne
3	Alcoholic hepatitis
4	Allergy
5	Arthritis
6	Bronchial Asthma
7	Cervical spondylosis
8	Chicken pox
9	Chronic cholestasis
10	Common Cold
11	Dengue
12	Diabetes
13	Dimorphic hemmorhoid
14	Drug Reaction
15	Fungal infection
16	GERD
17	Gastroenteritis
18	Heart attack
19	Hepatitis B
20	Hepatitis C
21	Hepatitis D
22	Hepatitis E
23	Hypertension
24	Hyperthyroidism
25	Hypoglycemia
26	Hypothyroidism
27	Impetigo
28	Jaundice
29	Malaria
30	Migraine
31	Osteoarthritis
32	Paralysis (brain hemorrhage)
33	Peptic ulcer disease
34	Pneumonia
35	Psoriasis
36	Tuberculosis
37	Typhoid
38	Urinary tract infection
39	Varicose veins
40	hepatitis A

**Itching**

**Muscle\_weakness**

**Scurring**

**Stomach\_bleeding**

**Slurred\_speech**

**Knee\_pain**

**Loss\_of\_balance**

**Depression**

**Obesity**

**Foul\_smell\_of\_urine**

## 7. ADVANTAGES & DISADVANTAGES

### **Advantages:**

- Accuracy: Machine learning models can be very accurate in predicting diseases, especially when they are trained on large datasets.
- Speed: Machine learning models can be used to predict diseases much faster than traditional methods, such as manual diagnosis.
- Cost-effectiveness: Machine learning models can be cost-effective, especially when they are used to predict common diseases.
- Accessibility: Machine learning models can be made accessible to a wider range of people, such as in rural areas or developing countries.

### **Disadvantages:**

- Data requirements: Machine learning models require large datasets to train, which can be difficult and expensive to obtain.
- Interpretability: Machine learning models can be difficult to interpret, which can make it difficult to understand how they make predictions.
- Bias: Machine learning models can be biased, which can lead to inaccurate predictions.
- Security: Machine learning models can be vulnerable to security attacks, which can lead to the misuse of the model.

## 8. APPLICATIONS

- Early diagnosis: Machine learning models can be used to predict diseases early before they cause symptoms. This can help to improve patient outcomes by allowing doctors to start treatment earlier.
- Personalized medicine: Machine learning models can be used to personalize treatment for patients. This means that doctors can tailor treatment to the individual patient's needs, which can improve patient outcomes.
- Research: Machine learning models can be used to conduct research into diseases. This can help to improve our understanding of diseases and to develop new treatments.
- Public health: Machine learning models can be used to improve public health. This can help to track the spread of diseases and to develop interventions to prevent outbreaks.
- Healthcare education: Machine learning models can be used to educate healthcare professionals. This can help them to learn about diseases and to improve their diagnostic skills.
- Patient engagement: Machine learning models can be used to engage patients in their healthcare. This can help patients to understand their risk of developing diseases and to take steps to prevent them.

## 9. CONCLUSION

*Machine learning is a powerful tool that can be used to predict diseases. Machine learning models can be trained on large datasets of patient data to learn the relationships between different features and diseases. This information can then be used to predict the risk of a patient developing a disease or to personalize treatment for patients.*

*Machine learning has the potential to revolutionize disease prediction and improve patient outcomes.*

*However, it is important to be aware of the limitations of machine learning before using it in clinical practice. Machine learning models can be biased and difficult to interpret, and they require large datasets to train.*

*Despite these limitations, machine learning is a promising tool for disease prediction. As machine learning technology continues to develop, machine learning will likely play an increasingly important role in the healthcare industry.*

*Here are some of the key points to remember about disease prediction using machine learning:*

- Machine learning models can be used to predict diseases with a high degree of accuracy.*
- Machine learning models can be used to personalize treatment for patients.*
- Machine learning models can be used to improve public health.*
- Machine learning models have some limitations, such as bias and interpretability.*
- Machine learning is a promising tool for disease prediction and is likely to play an increasingly important role in the healthcare industry.*

## 10. FUTURE SCOPE

- Improved accuracy: Machine learning models are constantly being improved, and they will likely become more accurate in the future. This will allow doctors to make better diagnoses and provide more personalized treatment.*
- New applications: Machine learning is still a relatively new field, and many new applications are being developed. For example, machine learning could be used to predict the risk of a patient developing a disease based on their genetic makeup.*
- Lower costs: As machine learning technology becomes more mature, it is likely to become more affordable. This will make it possible for more healthcare providers to use machine learning, which could improve patient outcomes for a wider range of people.*
- Better understanding of diseases: Machine learning can be used to analyze large datasets of patient data to identify patterns and trends. This information can help researchers to better understand diseases and to develop new treatments.*
- Prevention: Machine learning can be used to identify people who are at risk of developing a disease. This information can be used to target prevention interventions, such as public health campaigns or screening programs.*
- Personalized medicine: Machine learning models can be used to personalize treatment for patients. This means that doctors can tailor treatment to the individual patient's needs, which can improve patient outcomes.*

*In addition to these, here are some other future scopes of disease prediction using machine learning:*

- Real-time prediction: Machine learning models could be used to predict diseases in real-time, which would allow doctors to intervene early and prevent serious complications.
- Mobile applications: Machine learning models could be used to create mobile applications that allow people to track their health and predict their risk of developing diseases.
- Wearable devices: Machine learning models could be used to analyze data from wearable devices, such as smartwatches and fitness trackers, to predict diseases.

## **11. BIBLIOGRAPHY**

The sites we took reference from:-

<https://www.hindawi.com/journals/bmri/2023/6864343/>

<https://www.mdpi.com/2227-9032/11/7/1031>

<https://www.geeksforgeeks.org/disease-prediction-using-machine-learning/>

<https://ieeexplore.ieee.org/document/9791739>

[https://www.youtube.com/watch?v=8Q\\_QQVQ1HZA&list=PLfFghEzKVmjuhQwKhYXvdU94GSU-6Jcjr](https://www.youtube.com/watch?v=8Q_QQVQ1HZA&list=PLfFghEzKVmjuhQwKhYXvdU94GSU-6Jcjr)

<https://www.youtube.com/watch?v=kk762SkWv4U&pp=ygUqZGlzZWZzZSBwcmVkaWN0aW9uIHVzaW5nIG1hY2hpbmUqbGVhcm5pbmcmcg>

## **APPENDIX (CODES)**

*SmartBridge\_Project.ipynb*:-

<https://drive.google.com/file/d/19wf-9hat1IjTzMslzVv0vrNLalaIRmES/view?usp=sharing>

*Model.py*:-

<https://drive.google.com/file/d/1IJ1LO7yJgbaIa0U4hz8ybF9B8GlnHwTv/view?usp=sharing>

*Home.html* :-

<https://drive.google.com/file/d/1PyQOBrjqAcOEPcvYlXfuGuYOragxJBw/view?usp=sharing>

*Diseaseprediction.py* :-

<https://drive.google.com/file/d/1p5GGMbSP3uZZhO2pCBdFCUeHaPMdZsa5/view?usp=sharing>

*App.py* :-

[https://drive.google.com/file/d/1CzsPHps\\_krh184N4Ab5efOeOvvfu0O5c/view?usp=sharing](https://drive.google.com/file/d/1CzsPHps_krh184N4Ab5efOeOvvfu0O5c/view?usp=sharing)