# INTRODUCTION

* 1. Overview

Disease Prediction system is based on predictive modelling and predicts the disease of the user on the basis of the symptoms that user provides as an input to the system. The system analyses the symptoms provided by the user as an input and gives the probability of the disease as an output. Nowadays health industry plays a major role in curing the diseases of the patients so this project will also be kind of a help for the industry to tell the user and also it can be useful for the user in case he/she doesn’t want to go to the hospital. This system is completely done with the help of Machine Learning and Python with Tkinter interface for it and also the dataset that is available previously by the hospitals using that we can predict the disease. Sometimes doctors may fail to take accurate decisions while diagnosing the disease of a patient and therefore this system can assist in such cases to get accurate results.

1.2 Purpose

This system is used to predict disease according to symptoms using Machine Learning Technology. For predicting diseases, we mainly use 3 algorithms ,the first algorithm is a [**Decision Tree**](https://codespeedy.com/decision-tree-regression-in-python-using-scikit-learn/), second is a [**Random Forest**](https://codespeedy.com/introduction-to-random-forest-algorithm/) and the last one is [**Naive Bayes**](https://codespeedy.com/naive-bayes-algorithm-in-python/). This system is for those people who are always fretting about their health, for this reason, we provide some features which acknowledge them and enhance their mood too. So, there is a system for the awareness of health 'Disease Predictor', which recognize disease according to symptoms.

# LITERATURE SURVEY

# The use of different ML algorithms enabled the early detection of many maladies such as heart, kidney, breast and brain diseases. The classifiers used for the existing works were – common diseases(CNN, KNN), kidney(KNN, NB, RF and SVM), heart(DT, KNN, LR, NB, ANN, SVM, Classification Tree), breast diseases(RF, Bayesian Networks, SVM) and Parkinson’s Disease(SVM, FKNN, NB). Throughout the literature, SVM, RF and LR algorithms were the most widely used at prediction, while accuracy was the most used performance metric. The CNN model proved to be the most adequate at predicting common diseases. Furthermore, SVM model showed superiority in accuracy at most times for kidney diseases and Parkinson’s Disease because of its reliability in handling high-dimensional, semi-structured and unstructured data. For Breast cancer prediction, RF showed more superiority in the probability of correct classification of the diseases because of its ability to scale well for large datasets and its susceptibility to avoid overfitting. Finally, the LR algorithm proved to be the most reliable in predicting heart diseases.

# 

* 1. Existing problem

1. MYCIN: MYCIN is a medical diagnosis expert system that was developed to capture the expertise of a human expert on blood diseases. Physicians used it to diagnose and treat patients with infectious blood diseases caused by bacteria in the blood and meningitis.
2. ONCOCIN: ONCOCIN is a medical expert system that was developed to capture the expertise of a human expert on cancer. It was designed to assist physicians in the treatment of cancer patients receiving chemotherapy.
3. DAVIAL: DAVIAL is a medical expert system that was developed to diagnose heart diseases through echocardiography and other cardiac anomalies.

2.2 Limitations

Do not have human-like decision-making power.

Cannot possess human capabilities.

Cannot produce accurate disease

Only for specific diseases .

Does not predict wide variety of diseases

* 1. Proposed solution

The system takes the symptoms as input from the user.

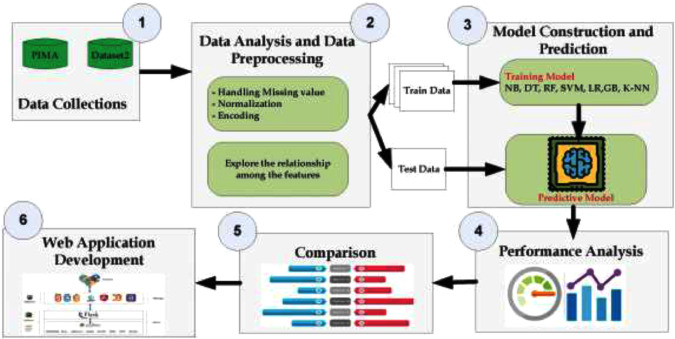
These symptoms are compared with the system’s dataset(i.e training and testing data set).

The dataset and symptoms go to the prediction model of the system where the data is pre-processed and then the classification of those data is done with the help of various algorithms and techniques such as Decision Tree, Naïve Bayes, Random Forest and etc.

If the prediction by any of the two algorithms is the same the user might have that disease.

# THEORITICAL ANALYSIS

* 1. Block diagram



* 1. Hardware / Software designing

Hardware requirements:

SYSTEM : intel core i3.i5.i7 and 2G GHz minimum

RAM : 512 Mb or above

HARD DISK : 10 GB or above

INPUT DEVICE : Keyboard and mouse

OUTPUT DEVICE : Monitor or PC

Software requirements:

OPERATING SYSTEM : Window 7,10 or higher versions

PLATFORM : JUPYTER NOTEBOOK

FRONT END : Python Tkinter

BACK END : Python and files

PROGRAMMING LANGUAGE : Python

# EXPERIMENTAL INVESTIGATIONS

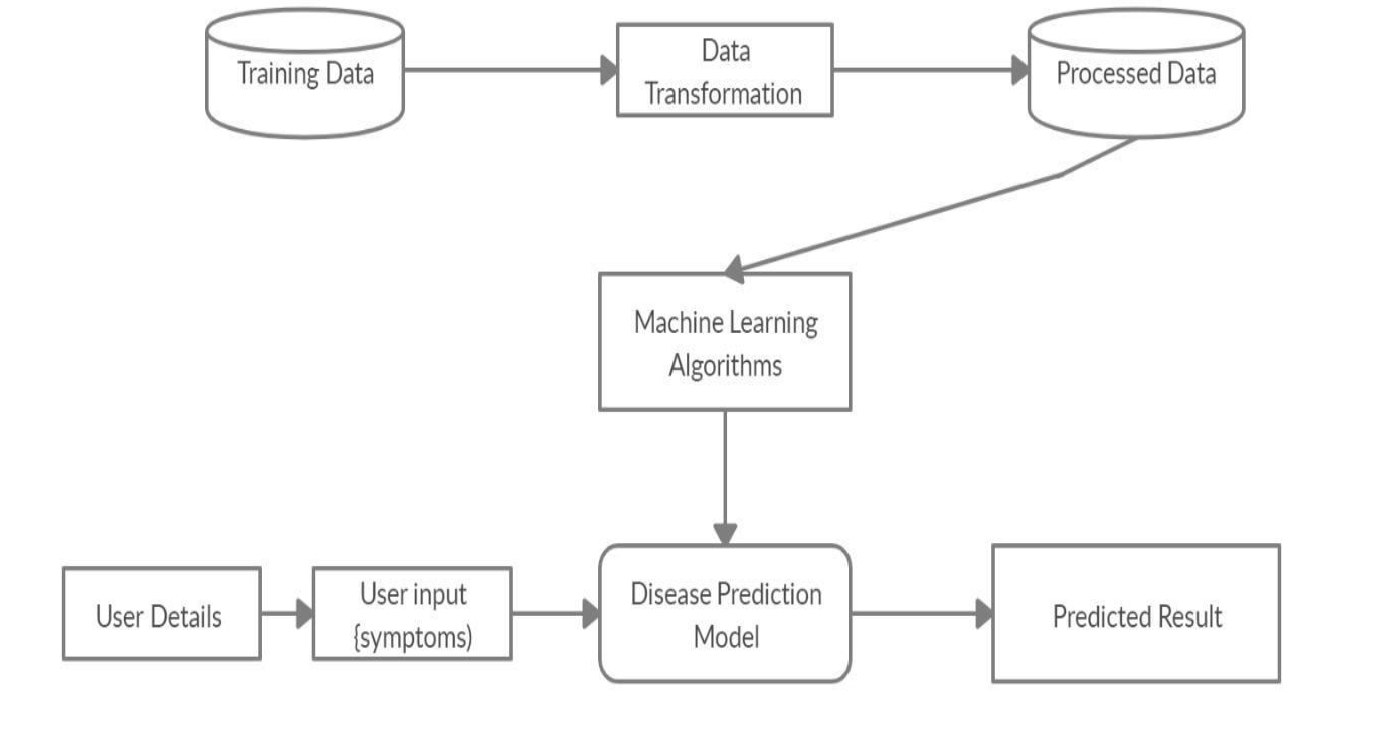
Data preparation is the primary step for any machine learning problem. We used a dataset from Kaggle for this problem. This dataset consists of two CSV files one for training and one for testing. There is a total of 133 columns in the dataset out of which 132 columns represent the symptoms and the last column is the prognosis.

**2**Cleaning is the most important step in a machine learning project. The quality of our data determines the quality of our machine-learning model. So it is always necessary to clean the data before feeding it to the model for training. In our dataset all the columns are numerical, the target column i.e. prognosis is a string type and is encoded to numerical form using a lab encoder

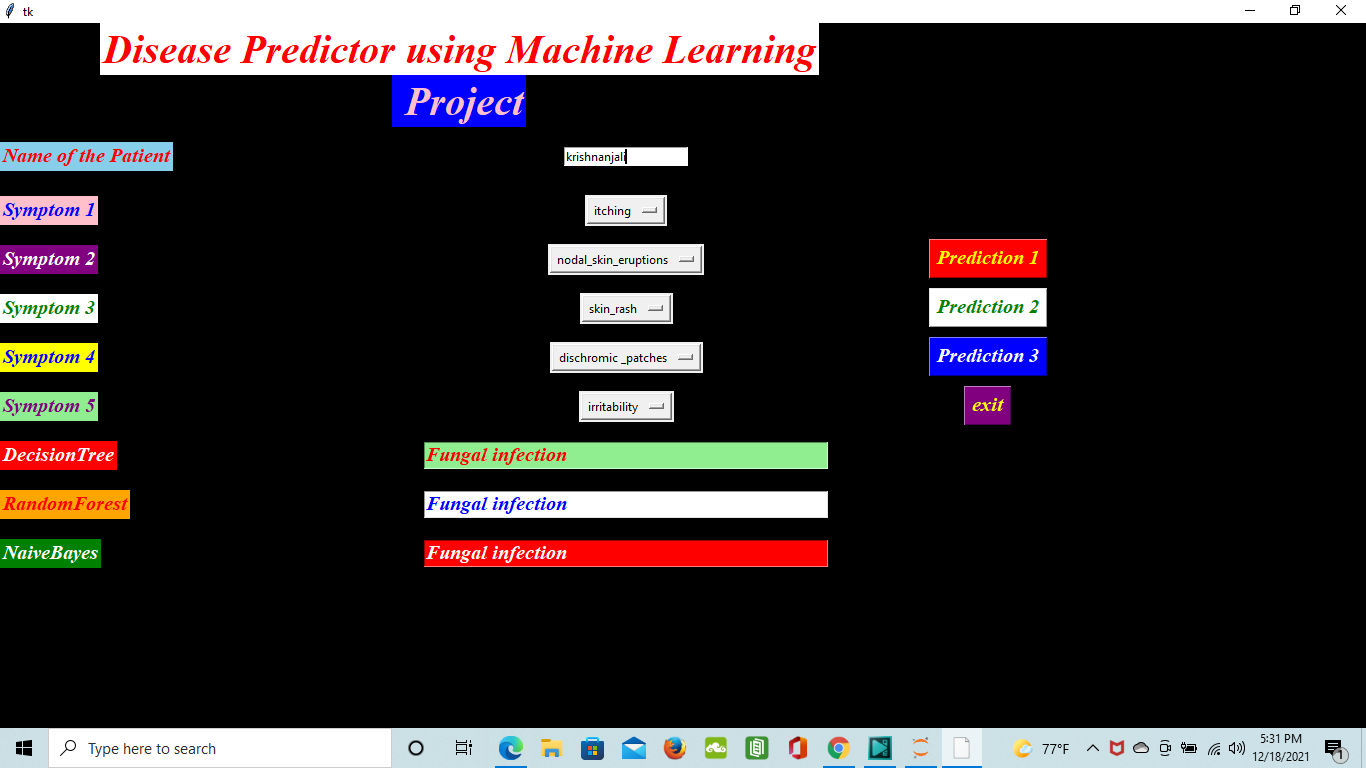
After gathering and cleaning the data, the data is ready and can be used to train a machine learning model. We will be using this cleaned data to train the Decision tree, Naive Bayes Classifier, and Random Forest Classifier. We will be using a [confusion matrix](https://www.geeksforgeeks.org/confusion-matrix-machine-learning/) to determine the quality of the models

After training the three models we will be predicting the disease for the input symptoms by combining the predictions of all three models. This makes our overall prediction more robust and accurate.

# FLOWCHART



# 7 RESULT



This section represents the proposed system results which can predict the disease faster, more accurate and with high reliability than the existing system. The results are obtained by implementing various Machine Learning algorithms. The Machine Learning classification techniques namely decision Tree, Random forest, Naïve Bayes are implemented using Python programming.

THE DISEASE PREDICTED BY

DECISION TREE:

Disease:Fungal Infection

Accuracy of test data =1.0 for 41 sample of test data

RANDOMFOREST:

Disease:Fungal Infection

Accuracy of test data =1.0 for 41 sample of test data

NAIVEBAYES:

Disease:Fungal Infection

Accuracy of test data =1.0 for 41 sample of test data

**ADVANTAGES:**

1. user feel comfortable and more secure
2. Children, adult, pet animal , old age people are used
3. spending more time for family, freedom for safety guards near the Swimming pool
4. Swimmers, resort are gain in the financal
5. drowning should be monitored

**DISADVANTAGE:**

1.For uneducated people will suffer from this technology

2.Electricity will be required

3.Software and hardware requirement will need

# APPLICATIONS

1. **Real-time Drowning Detection:** The proposed model can be deployed in a real-time surveillance system to monitor diseases ..
2. **Early Warning System**: By continuously analyzing the model can act as an early warning system

**3 Assistive Technology for Doctors**: Doctors can benefit from this model as an assistive tool. By analyzing the results the model can help healthcare professionals focus their attention on critical areas or individuals who require immediate assistance. This enhances their situational awareness and response capabilities.

**4 Training and Education:** The YOLOv5 model can be used as a training tool for healthcare professionals. By analyzing historical data, the model can identify common patterns and behaviors associated with diseases.

CONCLUSION

# The main aim of this project is to predict the disease in accordance with symptoms put down by the patients with proper implementation of Machine Learning algorithm. In this paper we have used three Machine Learning algorithm for prediction and achieved the mean accuracy around 100% which shows remarkable rectification and high accuracy than previous work and also makes this system more reliable than the existing one for this job and hence provides better satisfaction to the user in comparison with the other one.We have also created a GUI for better interaction with the system by users which is very easy to operate .This project shows that Machine Learning algorithm can be used to predict the disease easily with different parameters and models.

# FUTURE SCOPE

A. A prime account option available for the patients.

B. Video calling feature.

C. The website's account linking feature allows users to connect their account with other online services like Gmail and social media.

D. Addition of a map feature to the website, like adding an API for it.

E. Partner with a pharmacy and provide discounts on the medicine for the patients.

# BIBILOGRAPHY

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

https://www.irjet.net/archives/V7/i4/IRJET-V7I4993.pdf

https://www.kaggle.com/datasets/kaushil268/disease-prediction-using-machine-learning

# APPENDIX

https://github.com/anjna111ramdas/Disease-prediction