Human Disease Detection

A PROJECT REPORT

Submitted by
Yudhishter
(22BAI70012)
Harshit
(22BAI70085)
Shubham
(22BAI70002)
Aakansha
(22BAI70092)

in partial fulfillment for the award of the degree of

BACHELOR OF ENGINEERING

IN

ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

Chandigarh University

DECEMBER 2022

BONAFIDE CERTIFICATE

Certified that this project report "**Human disease detection"** is the bonafide work of "**Yudhishter, Harshit, Shubham, Aakansha"** who carried out the project work under my/oursupervision.

SIGNATURE
Aman kaushik
HEAD OF THE DEPARTMENT

SIGNATURE
Kirti
SUPERVISOR

Submitted for the project viva-voce examination held on _

INTERNAL EXAMINER

EXTERNAL EXAMINER

ACKNOWLEDGEMENT

We have taken efforts in this project. However, it would not have been possible without the kind support and help of our supervisor and organization. I would like to extend my sincere thanks to all of them. We are highly indebted to Prof. Ajay Pal for their guidance and constant supervision as well as for providing necessary information regarding the project & also for their support in completing the project. We would like to express my gratitude towards our family and department for their kind co-operation and encouragement which help us in completion of this project.

THANKS AGAIN TO ALL WHO HELPED

TABLE OF CONTENTS

ABSTRACT

With the rise in number of patient and disease every year medical system is overloaded and with time have become overpriced in many countries. Most of the disease involves a consultation with doctors to get treated. With sufficient data prediction of disease by an algorithm can be very easy and cheap. Prediction of disease by looking at the symptoms is an integral part of treatment. In our project we have tried accurately predict a disease by looking at the symptoms of the patient. We have used 4 different algorithms for this purpose and gained an accuracy of 92-95%. Such a system can have a very large potential in medical treatment of the future. We have also designed an interactive interface to facilitate interaction with the system. We have also attempted to show and visualized the result of our study and this project.

Introduction

Human Healthcare is one of the most significant subjects of society.The identification of the nature of the illness or other problems by examination of the symptoms i.e. diagnosis always stands first in overall curing procedure of disease. Thus, we can say that diagnosis and prediction of disease are most crucial aspects to be considered before thinking about exact procedure of curing the disease. This diagnosis procedure requires a lot of time as well as a lot of money. As a result, people belonging to poor financial background are not able to get accurate diagnose of disease which at the end may create life or death situation. So WeCare is basically a Human Disease Detection System which is supposed to input some data related like some sort of symptoms etc. and in response to that it gives type of disease that person is having as an output. So, this System will be salutary to people who are not able to pay huge amounts to get diagnosed as well as the ones who requires immediate diagnosis.

With the rise in number of patient and disease every year medical system is overloaded and with time have become overpriced in many countries. Most of the disease involves a consultation with doctors to get treated. With sufficient data prediction of disease by an algorithm can be very easy and cheap. Prediction of disease by looking at the symptoms is an integral part of treatment. In our project we have tried accurately predict a disease by looking at the symptoms of the patient. We have used 4 different algorithms for this purpose and gained an accuracy of 92-95%. Such a system can have a very large potential in medical treatment of the future. We have also designed an interactive interface to facilitate interaction with the system. We have also attempted to show and visualized the result of our study and this project.

Database Collection

Dataset for this project was collected from a study of university of Columbia performed at New York Presbyterian Hospital during 2004.

Library Used

In this project standard libraries for database analysis and model creation are used. The following are the libraries used in this project.

tkinter:

It's a standard GUI library of python. Python when combined with tkinter provides fast and easy way to create GUI. It provides powerful object-oriented tool for creating GUI. It provides various widgets to create GUI some of the prominent ones being:

- Button
- Canvas
- Label
- Entry
- Check Button
- List box
- Message
- Text
- Messagebox

Some of these were used in this project to create our GUI namely messagebox, button, label, Option Menu, text and title. Using tkinter we were able to create an interactive GUI for our model.

Creating a GUI application using Tkinter is an easy task. All you need to do is perform the following steps −

- Import the *Tkinter* module.
- Create the GUI application main window.
- Add one or more of the above-mentioned widgets to the GUI application.

- Enter the main event loop to take action against each event triggered by the user.

Numpy:

Numpy is core library of scientific computing in python. It provides powerful tools to deal with various multi-dimensional arrays in python. It is a general purpose array processing package. Numpy's main purpose is to deal with multidimensional homogeneous array. It has tools ranging from array creation to its handling.

It also has functions for working in domain of linear algebra, fourier transform, and matrices.

NumPy was created in 2005 by Travis Oliphant. It is an open source project and you can use it freely.

NumPy stands for Numerical Python.

NumPy aims to provide an array object that is up to 50x faster than traditional Python lists.

The array object in NumPy is called ndarray, it provides a lot of supporting functions that make working with ndarray very easy.

Arrays are very frequently used in data science, where speed and resources are very important.

NumPy arrays are stored at one continuous place in memory unlike lists, so processes can access and manipulate them very efficiently.

This behavior is called locality of reference in computer science.

This is the main reason why NumPy is faster than lists. Also it is optimized to work with latest CPU architectures.

NumPy is a Python library and is written partially in Python, but most of the parts that require fast computation are written in C or C++.

It makes it easier to create a n dimensional array just by using np.zeros() or handle its contents using various other methods such as replace, arrange, random, save, load it also helps I array processing using methods like sum, mean, std, max, min, all, etc.

Array created with numpy also behave differently then arrays created normally when they are operated upon using operators such as +,-,*,/.

All the above qualities and services offered by numpy array makes it highly suitable for our purpose of handling data.

pandas:

Pandas is a Python library used for working with data sets.

It has functions for analyzing, cleaning, exploring, and manipulating data.

The name "Pandas" has a reference to both "Panel Data", and "Python Data Analysis" and was created by Wes McKinney in 2008.

Pandas allows us to analyze big data and make conclusions based on statistical theories.

Pandas can clean messy data sets, and make them readable and relevant.

Relevant data is very important in data science.

Pandas gives you answers about the data. Like:

- Is there a correlation between two or more columns?
- What is average value?
- Max value?
- Min value?

Pandas are also able to delete rows that are not relevant, or contains wrong values, like empty or NULL values. This is called *cleaning* the data.

The source code for Pandas is located at this github repository https://github.com/pandas-dev/pandas

It is the most popular python library used for data analysis. It provides highly optimized performance with back-end source code purely written in C or python.

Data in python can be analysed with 2 ways :

- Series
- Dataframes

Series is one dimensional array defined in pandas used to store any data type. Dataframes are two-dimensional data structure used in python to store data consisting of rows and columns.

Pandas dataframe is used extensively in this project to use datasets required for training and testing the algorithms. Dataframes makes it easier to work with attributes and results. Several of its inbuilt functions such as replace were used in our project for data manipulation and preprocessing.

sklearn:

Scikit-learn (Sklearn) is the most useful and robust library for machine learning in Python. It provides a selection of efficient tools for machine learning and statistical modeling including classification, regression, clustering and dimensionality reduction via a consistence interface in Python. This library, which is largely written in Python, is built upon NumPy, SciPy and Matplotlib.

It was originally called ***scikits.learn*** and was initially developed by David Cournapeau as a Google summer of code project in 2007. Later, in 2010, Fabian Pedregosa, Gael Varoquaux, Alexandre Gramfort, and Vincent Michel, from FIRCA (French Institute for Research in Computer Science and Automation), took this project at another level and made the first public release (v0.1 beta) on 1st Feb. 2010.

Let's have a look at its version history −

- May 2019: scikit-learn 0.21.0
- March 2019: scikit-learn 0.20.3
- December 2018: scikit-learn 0.20.2
- November 2018: scikit-learn 0.20.1
- September 2018: scikit-learn 0.20.0
- July 2018: scikit-learn 0.19.2
- July 2017: scikit-learn 0.19.0
- September 2016. scikit-learn 0.18.0
- November 2015. scikit-learn 0.17.0
- March 2015. scikit-learn 0.16.0
- July 2014. scikit-learn 0.15.0
- August 2013. scikit-learn 0.14

Sklearn is an open source python library with implements a huge range of machinelearning, pre-processing, cross-validation and visualization algorithms. It features various simple and efficient tools for data mining and data processing. It features various classification, regression and clustering algorithm such as support vector machine, random forest classifier, decision tree, gaussian naïve-Bayes, KNN to name a few.

In this project we have used sklearn to get advantage of inbuilt classification algorithms like decision tree, random forest classifier, KNN and naïve Bayes. We have also used inbuilt cross validation and visualization features such as classification report, confusion matrix and accuracy score.

DOMAIN OF THE PROJECT

This project is ML based which is basically a subfield of artificial intelligence in which we use data and algorithmic approach to imitate a human behavior.
Broadly Machine Learning is divided into two parts:

1) **Supervised ML**

2) **Unsupervised ML**

Supervised ML :

Supervised learning is typically done in the context of classification, when we want to map input to output labels, or regression, when we want to map input to a continuous output. Common algorithms in supervised learning include logistic regression, naive bayes, support vector machines, artificial neural networks, and random forests. In both regression and classification, the goal is to find specific relationships or structure in the input data that allow us to effectively produce correct output data. Note that "correct" output is determined entirely from the training data, so while we do have a ground truth that our model will assume is true, it is not to say that data labels are always correct in real-world situations. Noisy, or incorrect, data labels will clearly reduce the effectiveness of your model.
When conducting supervised learning, the main considerations are model complexity, and the bias-variance tradeoff. Note that both of these are interrelated.

Unsupervised ML :

The most common tasks within unsupervised learning are clustering, representation learning, and density estimation. In all of these cases, we wish to learn the inherent structure of our data without using explicitly-provided labels. Some common algorithms include k-means clustering, principal component analysis, and autoencoders. Since no labels are provided, there is no specific way to compare model performance in most unsupervised learning methods.

Hardware Specification

Processor - Minimum 1 GHz; Recommended 2GHz or more

Windows 7 or newer

MAC: OS X v10.7 or higher or

Linux: Ubuntu

Software Specification
Jupyter Notebook

Python Libraries - TensorFlow , Pandas, NumPy, Matplotlib, Scikit-learn, Keras

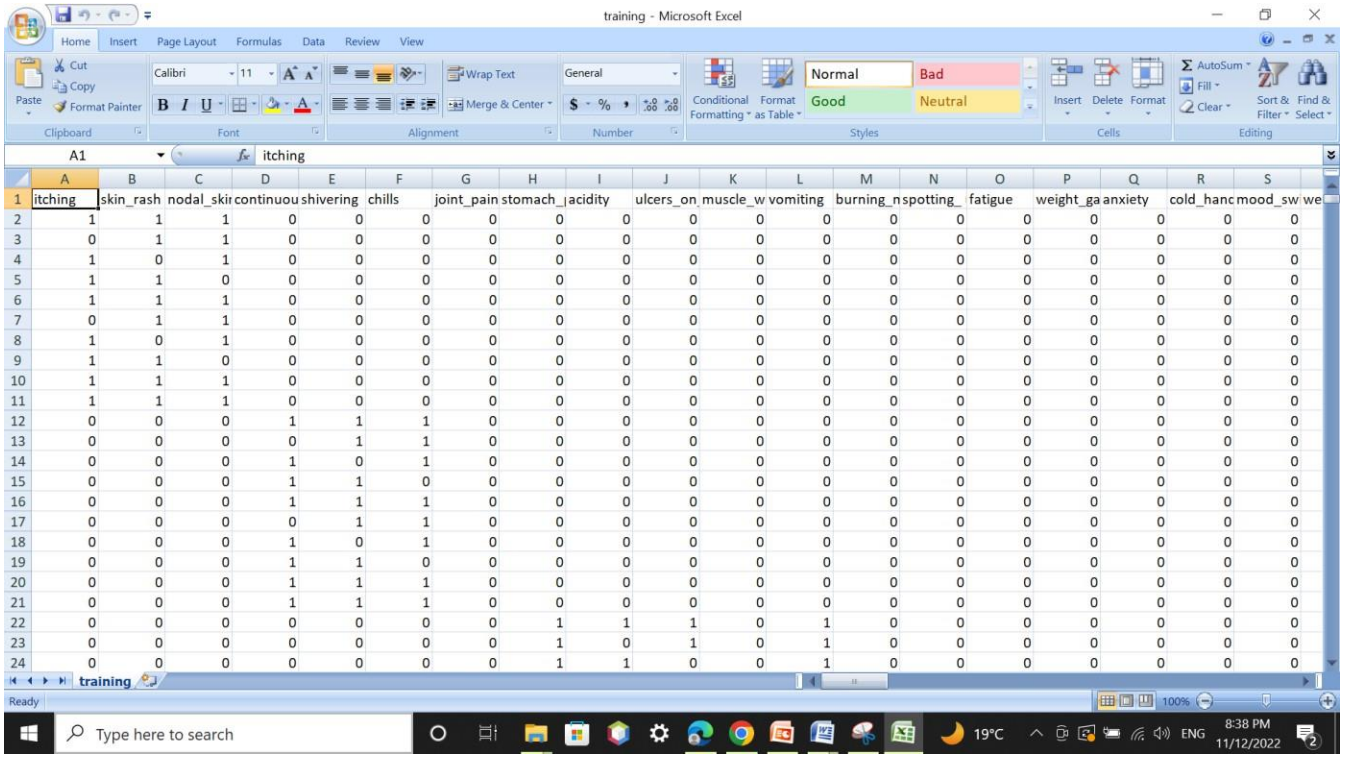
PROBLEM IDENTIFICATION

Now-a-days, people face various diseases due to the environmental condition and their living habits. Accurate and on-time analysis of any health-related problem is important for the prevention and treatment of the illness. So, the prediction of disease at earlier stage becomes important task. The traditional way of diagnosis may not be sufficient in the case of a serious ailment. 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 machine learning plays an important role to predict the disease. Developing a medical diagnosis system based on machine learning (ML) algorithms for prediction of any disease can help in a more accurate diagnosis than the conventional method. Medical science has large amount of data growth per year. 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, machine learning finds hidden pattern information in the huge amount of medical data. The exact examination of clinical dataset advantages in early illness expectation, patient consideration and local area administrations. The methodology of Machine Learning (ML) has been effectively utilized in grouped technologies including disease forecast. The objective of generating classifier framework utilizing Machine Learning (ML) models is to massively assist with addressing the well-being related issues by helping the doctors to foresee and analyse illnesses at a beginning stage. beginning phase. However, supervised machine learning (ML) algorithms have showcased significant potential in surpassing standard systems for disease diagnosis and aiding medical experts in the early detection of high-risk diseases. Here, the aim is to recognize trends across various types of supervised ML models in disease detection through the examination of performance metrics. We proposed general disease prediction based on symptoms of the patient. For the disease prediction, we use K-Nearest Neighbour (KNN), Naive Bayes, Decision Tree Classifier and Random Forest Classifier machine learning algorithms for accurate prediction of disease.

Dataset Discussion

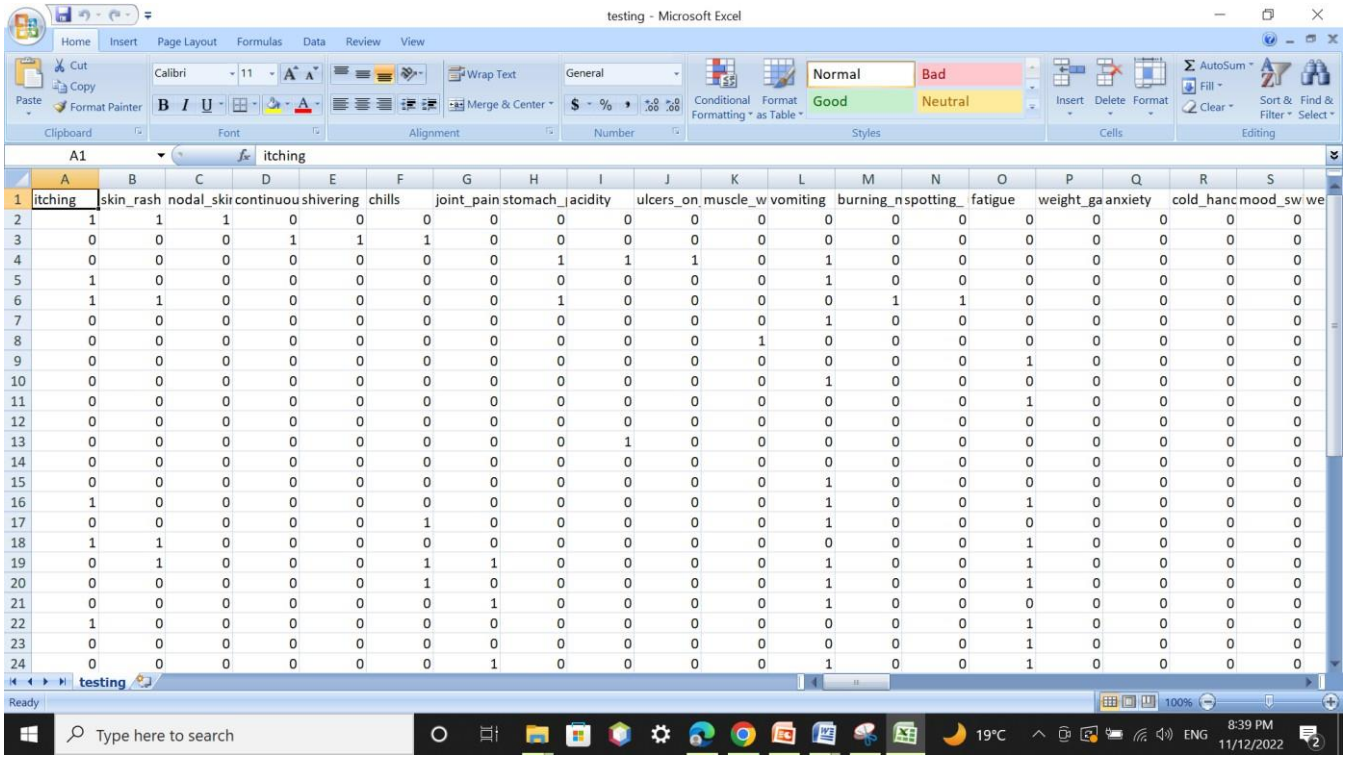
The dataset we have considered comprises of 132 indications, the blend or stages of which leads to 41 illnesses. In light of the 4920 documents of various patient samples, mainly to point foster a forecast algorithm that considers in the side effects of various client and forecasts the sickness that the person is bound to be affected. There are columns containing diseases, their symptoms, precautions to be taken, and their weights.

Training Data :



itching	skin_rash	nodal_skin	continuous	shivering	chills	joint_pain	stomach_p	acidity	ulcers_on	muscle_w	vomiting	burning_n	spotting_	fatigue	weight_ga	anxiety	cold_hanc	mood_sw	we
1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
1	1	0	0	0	0	0	0	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	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
1	1	0	0	0	0	0	0	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	
1	1	1	0	0	0	0	0	0	0	0	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	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	
0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	
0	0	0	1	1	0	0	0	0	0	0	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	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	
0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	
0	0	0	1	1	0	0	0	0	0	0	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	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	
0	0	0	0	0	0	0	1	1	1	0	1	0	0	0	0	0	0	0	
0	0	0	0	0	0	0	1	0	1	0	1	0	0	0	0	0	0	0	
0	0	0	0	0	0	0	1	1	0	0	1	0	0	0	0	0	0	0	

Testing Data:



itching	skin_rash	nodal_skin	continuous	shivering	chills	joint_pain	stomach_p	acidity	ulcers_on	muscle_w	vomiting	burning_n	spotting_	fatigue	weight_ga	anxiety	cold_hanc	mood_sw	we
1	1	1	0	0	0	0	0	0	0	0	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	0	0	0	1	1	1	0	1	0	0	0	0	0	0	0	
1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	
1	1	0	0	0	0	0	1	0	0	0	0	1	1	0	0	0	0	0	
0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	
0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	
0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	
0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	
0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
1	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	
0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	
1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	
0	1	0	0	0	1	1	0	0	0	0	1	0	0	1	0	0	0	0	
0	0	0	0	0	1	0	0	0	0	0	1	0	0	1	0	0	0	0	
0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	
1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	
0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	
0	0	0	0	0	0	1	0	0	0	0	1	0	0	1	0	0	0	0	

LITERATURE SURVEY

The identification of the nature of the illness or other problems by examination of the symptoms i.e. diagnosis always stands first in overall curing procedure of disease. Thus , we can say that diagnosis andprediction of disease are most crucial aspects to be considered before thinking about exact procedure of curing the disease. This diagnosis procedure requires a lot of time as well as a lot of money. As a result , people belonging to poor financial background are not able to get accurate diagnose of disease which at the end may create life or death situation.Here are some of the work which are already done by people inorder to solve this problem :

The (Senturk & Kara, 2014) contributed to early breast cancer diagnosis in this study. An analysis of the diagnosis of breast cancer for patients is provided. Seven different algorithms are used to realize the predictions of the other patients and give them precision. Patient data from UCI ML during the prediction process, the data mining tool RapidMiner 5.0, is used to apply data mining with the desired algorithms during the prediction process.

In (Ramana et al., 2011), they found that the AP datasets were better than the UCLA datasets for all the various chosen algorithms. The writers used two separate datasets of inputs. The AP datasets were calculated to be better than the UCLA

dataset. Based on the usefulness of their KNN classification, backward propagation and SVM give better outcomes. For the entire chosen algorithm, the AP data set is better than UCLA. Besides, 95.07, 96.27, 96.93, 97.47, & 97.07 % accuracy have C4.5, Backward propagation, Naïve Bayes, SVM, and KNN. (Kousarrizi et al., 2012) this analysis is focused on two databases on thyroid disease. The first dataset is taken from the UCI machine learning repository. The second is the actual data gathered from the Imam Khomeini hospital by the Intelligent Device Laboratory of the K.N.Toosi University of Technology. They obtained a classification accuracy of 98.62 % using SVM for the first dataset, which is the highestaccuracy achieved so far. (Chitra et al., 2018) in the paper, the SVM with a Radial base function kernel is used for classification. The output parameters are high, such as the classification accuracy, sensitivity, and specificity of the SVM and RBF, making it the right choice for the classificationprocess. (Fan et al., 2013) Twelve morphological features from the ST segment were extracted. Using the SVM classifier, they obtained 95.20% sensitivity, 93.29% specificity, and 93.63% accuracy. (Hariharan et al., 2014) to diagnose Parkinson's disease, in this approach, the neural networks and the SVM algorithm are fused. The experimental findings show that for Parkinson's dataset, the combination of feature preprocessing, feature reduction/selection methods, and classification give a maximum classification precision of 100 %.

The efficiency analysis of the ML techniques on diabetes disease detection is performed in this paper. The work uses various ML techniques (DT), LR, DA, SVM, k-NN and ensemble learners. Software from MATLAB is taken into account. The findings are analyzed based on the 10-fold cross- validation criterion, and the performance analysis uses average classification accuracy. The averageaccuracy scores obtained are in the 65.5 % and 77.9 % range. The LR method provides the best accuracy score of 77.9 %, and the worst one is provided by the Coarse Gaussian SVM technique of 65.5 % in (Al-Zebari & Sengur, 2019). The patient data sets are analyzed by Durai (2019), based solely on a commonly diagnosed classification model for predicting the subject having a liver disorder. A necessary assessment process is carried out, depending on the studies, to maintain the integrity of a specific representation of the outcome. The J48 algorithm is a higher-performing algorithm with an accuracy rate of 95.04 per cent for feature selection.

Intracerebral hemorrhage sources for high mortality rate as a result, (Liu et al., 2019) it is based on multivariate analysis to anticipate the expansion of hematoma in spontaneous ICH with normally accessible SVM data and pointed out 83. A randomized 179 search approach was used in this study for parameter tuning, and recursive function 180 elimination was used for feature selection. Patient selection for thrombolytic procedures is another significant factor. Rustam et al. (2020) used three types of the forecast for each model: the number of cases freshly infected, the number of casualties, and the number of recoveries over the next ten days. The outcomes provided by the Study Analysis indicate that the use of these methods in the

current COVID-19 pandemic scenario is a promising mechanism. The results show that of all the models used, and the ES performs best, followed by LASSO & LR, which performs well in forecasting newly recorded incidents, death rate and recovery rate, Although SVM does not perform well in the prediction scenarios, the available dataset is given. Tanveer et al. (2020) analyzed 165 articles from 2005-2019 using different feature extraction techniques and machine learning techniques. Three key categories are studied in ML techniques: SVM, ANN and DL, and the ensemble methods.

Literature Review Summary

Table2.1: Literature review summary

Year and citation	**Article By**	**Purpose of the study**	**Tools/ Software used**	**Findings**	**Data set (ifuse**	**Evaluation parameters**
2019 2019 2020 2020 2020	Javed et al. Liu et al. Kumar et al.	Research Research R	RSA , RF SVM ES , LR , SVM , LASSO	93.33% 83.3 % ---------- 96.9 %	Cleveland heartfailure 1157 Patients Github LI	Accuracy Accuracy Accuracy Accuracy Accuracy

	Naqi etal. Rustam et al.	esearch Research	DL CNN	97.2 %	DC-IDRI SN-AM	

Design flow

With the rise in number of patient and disease every year medical system is overloaded and with time have become overpriced in many countries. Most of the disease involves a consultation with doctors to get treated. With sufficient data prediction of disease by an algorithm can be very easy and cheap. Prediction of disease by looking at the symptoms is an integral part of treatment. In our project we have tried accurately predict a disease by looking at the symptoms of the patient. We have used 4 different algorithms for this purpose and gained an accuracy of 92-95%. Such a system can have a very large potential in medical treatment of the future. We have also designed an interactive interface to facilitate interaction with the system. We have also attempted to show and visualized the result of our study and this project.

In medical domains, artificial intelligence (AI) primarily focuses on developing the algorithms and techniques to determine whether a system's behavior is correct in disease diagnosis. Medical diagnosis identifies the disease or conditions that explain a person's symptoms and signs. Typically, diagnostic information is gathered from the patient's history and physical examination . It is frequently difficult due to the fact that many indications and symptoms are ambiguous and can only be diagnosed by trained health experts. Therefore, countries that lack enough health professionals for their populations, such as developing countries like Bangladesh and India, face difficulty providing proper diagnostic procedures for their maximum population of patients. Moreover, diagnosis procedures

Why To Choose ML Domain for this Project ?

Machine learning (ML) is used practically everywhere, from cutting-edge technology (such as mobile phones, computers, and robotics) to health care (i.e., disease diagnosis, safety). ML is gaining popularity in various fields, including disease diagnosis in health care. Many researchers and practitioners illustrate the promise of machine- learning-based disease diagnosis (MLBDD), which is inexpensive and time-efficient . Traditional diagnosis processes are costly, time-consuming, and often require human intervention. While the individual's ability restricts traditional diagnosis techniques, ML-based systems have no such limitations, and machines do not get exhausted as humans do. As a result, a method to diagnose disease with outnumbered patients' unexpected presence in health care may be developed. To create MLBDD systems,

health care data such as images (i.e., X-ray, MRI) and tabular data (i.e., patients' conditions, age, and gender) are employed.

Machine learning (ML) is a subset of AI that uses data as an input resource. The use of predetermined mathematical functions yields a result (classification or regression) that is frequently difficult for humans to accomplish. For example, using ML, locating malignant cells in a microscopic image is frequently simpler, which is typically

challenging to conduct just by looking at the images. Furthermore, since advances in deep learning (a form of machine learning), the most current study shows MLBDD accuracy of above 90% . Alzheimer's disease, heart failure, breast cancer, and pneumonia are just a few of the diseases that may be identified with ML. The emergence of machine learning (ML) algorithms in disease diagnosis domains illustrates the technology's utility in medical fields.

Recent breakthroughs in ML difficulties, such as imbalanced data, ML interpretation, and ML ethics in medical domains, are only a few of the many challenging fields to handle in a nutshell. In this paper, we provide a review that highlights the novel uses of ML and DL in disease diagnosis and gives an overview of development in this field in order to shed some light on this current trend, approaches, and issues connected with ML in disease diagnosis. We begin by outlining several methods to machine learning and deep learning techniques and particular architecture for detecting and categorizing various forms of disease diagnosis.

Recent Achievements in AI Health Care Industry:

Over the past two decades, artificial intelligence has created a robust place for itself in human society and our daily lives. Most of us are unaware of the fact that everything starting from our social media feeds to the online shopping experience is powered by AI. Fortunately, healthcare is a major area that is becoming increasingly reliant on technology. Diagnostics, treatment management, drug discovery and production, surgeries, etc are revolutionized by the influence of artificial intelligence. AI sub-technologies like machine learning, natural language processing (NLP), and data science are also helping in identifying healthcare needs and solutions faster with more accuracy. On the other hand, we also fear that these capabilities could surpass human ability and make us worthless. Although this could be eminent in the far future, technology is under human toes for now.

AI-based Cognitive Behavioural Therapy: Mental health issues are becoming very common in the modern world. However, addressing them is time-consuming and often very expensive. Unique mental issues like the opioid abuses' crises are very hectic and need new methods to deliver proper care to patients. Fortunately, AI-based cognitive behavioural therapy is here to help. This emerging method is detecting depression, eating disorders, and substance abuse and helps provide apt treatments through customized medical services.

Automatic Stroke Detection: Every minute matters when a person is suffering from a stroke. Getting access to skilled stroke care takes a lot of time, which could reflect on the patients' health in the future. Besides this, identifying the exact spot of collapse is still a hectic task for medical professionals. Automation and AI tools can aid the medical

staff in stroke detection and treatment. With the help of automation stroke detection device, doctors can close the gaps in high-quality imaging studies that can identify the type of stroke and the location of the clot or bleeding.

Social Media Analysis on Suicidal Thoughts: According to a 2016 report, over 45,000 people died of suicide in the US. The major cause of such suicidal and self- harming thoughts lies with socioeconomic and mental health factors. However, social media analysis on profiles can help flag such suicidal thoughts beforehand. Artificial intelligence can analyse social media records and health data to alarm self-harming tendencies. Following this, the patients can use AI-driven apps to maintain therapy and support until they are better.

3D Printed Limbs: Besides artificial intelligence, robotics capabilities are also invading the medical space. Existing robotic arms and legs can already function like human body parts. As an advanced method, 3D printed hands are here to enhance the capabilities even further. Simple tasks like changing the Television volume can be done from a distance by just moving the 3D printed limbs. It is also fused with voice recognition technology that can command smart home tasks.

Autism Virtual Reality: Although we know autistic children suffer from various struggles even on a daily basis, we often fail to take a step forward to help them. Autism virtual reality promotes empathy and helps people around autistic patients support them in a better way. When people wear this VR headset, they get to experience what it feels like to be with autism. This could help change their perspective on autistic children.

AI Cancer Detection: Detecting cancer at the early stage is unique. However, technology is making it possible with its growth. Artificial intelligence can detect abnormalities in radiographs at potential stages. This helps find any kind of unclear or out-of-ordinary gestures in the body that could signal cancer symptoms. Besides, AI is generally used in cancer detection, screening, treatment, chemotherapy, etc.

AI-enhanced Microscopes: AI-enhanced microscopes are used to scan harmful bacterias in blood samples. They can be faster and more accurate than human lab technicians. Beth Israel Deaconess Medical Centre has come up with a diagnostic microscope that uses artificial intelligence capabilities to detect deadly diseases earlier than human detection. It helps find harmful bacteria in blood with 95% accuracy.

Digital Nurses: During the Covid-19 health disaster, we came to know the importance of medical facilities across the globe. Along with that, we also realized there are not enough medical staff including doctors and nurses to handle the overflowing patients. Fortunately, digital nurses are here to transform patient care. They can help monitor patient conditions between doctor visits. For example, Molly is a digital nurse created

by Sense.ly to keep an eye on patient conditions.

Personalized Medical Care: Giving similar treatment to 100 patients who suffer from the same disease is not a good move. Everybody's body type, immunity, medical aspects, etc might differ. Therefore, healthcare institutions are coming up with personalized treatment facilities that could leverage tailored patient care by carefully analysing their health data. This will reduce the cost of extensive healthcare and increase effectiveness. Healthcare institutions use machine learning technology to match patient data with the most possible effective medications.

Surgical Robots: With the combination of artificial intelligence capabilities and collaborative robotic functions, surgical robots are revolutionizing operation theatres. Although today's surgical robots are capable of autonomously conducting surgical procedures, doctors still stand alongside them to look over the operations. However, they can do simple tasks and repetitive movements without the doctor's assistance. By infusing AI capabilities in the surgical robots, they can improve control accuracy to sub-millimetre precision.

The Future of AI in Healthcare: With a plethora of issues like the ageing population, growing chronic diseases, and scarcity of medical professions, the need for artificial intelligence and other technologies is drastically growing in healthcare. Although AI-powered solutions have made small changes towards these key issues, we have a long way to go before the technology matures and addresses all the concerns. However, over the next few years, hybrid AI models are anticipated to rule the healthcare space. With supreme diagnostics, sophisticated treatment, and early identity process at the core of patient care, disruptive trends like machine learning, NLP, robotics, etc will play a big role in answering complex questions. Besides, a good number of AI applications are awaiting approval to hit the market. Once they reach commercial facilities, patient care and the potential to save lives will improve exponentially.

Social , Political and Ethical Issues in AI Based Health Case Systems :

We have identified five types of use case for artificial intelligence technologies in health and medical research.

Use Cases	Examples	
	Current	Future
1. Process optimisation *Using AI to optimise 'back-end' processes in healthcare, such as procurement, logistics, and staff scheduling*	• Rota/staff schedule management, e.g. Hong Kong Health Authority • Emergency services dispatch management e.g. Corti	• Data-driven optimisation of logistics, procurement • Automated analysis/ completion of medical notes and other documentation • Patient experience analysis e.g. Alder Hey
2. Preclinical research *Using AI in preclinical applications such as drug discovery and genomic science*	• Candidate molecule screening, e.g. BenevolentAI, AtomNet • Repurposing drugs, e.g. Teva Pharmaceuticals • Predicting potential side effects, e.g. Cloud Pharmaceuticals • Analysis of large -omics datasets to gain insights.[16]	• Determining targets for gene editing, e.g. CRISPR

3. Clinical pathways *Fitting AI into existing and new clinical workflows, such as in diagnostics and prognostication*	• Analysis of optical coherence tomography (OCT) images, e.g. DeepMind-Moorfields collaboration • Analysis of radiological imaging, e.g. Viz.ai • Analysis of clinical conversations e.g. Corti • Prognostication e.g. prediction of all-cause mortality [Stanford, KenSci], prediction of cardiovascular risk [University of Nottingham]	• Radiologists' assistants, e.g. suggesting best imaging modality in particular clinical situation, improved image acquisition processes leading to radiation dose reduction • Management decision-support for healthcare practitioners, suggesting best treatment for particular patient • Automated transcription of clinical interactions • Automated completion and submission of investigation requests/referrals
4. Patient-facing applications *Using AI to interact directly with patients and other service users, including in the delivery of therapies or the provision of information*	• Chatbots, e.g. Oli [Alder Hey], AVA [Arthritis Research UK], Lark Weight Loss Coach • Autonomous (closed-loop) insulin pumps • Personalised health advice and interventions, e.g. CareSkore, Viome, DayTwo	• Smart homes and wearables • Robot carers • Robot surgeons

5. Population-level applications *Using AI to gain insights into population health, such as identifying epidemics and monitoring disease spread.*	• Prediction of infectious disease outbreaks, e.g. Dengue fever app in Malaysia • Better targeting of public health spending and other interventions, e.g. University of Southern California tool • Better understanding of risk-factors for non-communicable disease, e.g. childhood obesity [Indiana University tool]	

Effective delivery of healthcare services relies on the strategic deployment of resources, both physical and human. Even as spending on healthcare continues to outpace broader economic growth, it remains difficult to completely meet a community's or a country's health needs, particularly as people are living longer worldwide, with an attendant increase in complex chronic conditions. Although many of our interviewees indicated that optimising use of limited resources could be a major use of AI technologies in healthcare systems worldwide, examples of this use case category currently in practice are few and far between. One area where AI is being applied in healthcare systems is rostering. The Hong Kong Health Authority, for example, is using an AI-based tool developed at the local City University of Hong Kong to produce monthly or weekly staff rosters that satisfy a set of constraints, such as staff availability, staff preferences, allowed working hours, ward operational requirements and hospital regulations. This tool has been deployed across 40 public hospitals, and is responsible for the rostering of over 40,000 staff. Since being introduced, the Hospital Authority reports increased productivity, improved staff morale, and improved quality of service, as the system is seen to be fair, frees up managers' time, and can provide management with insights into working patterns and resource utilisation. This feedback contrasts with reports about the use of scheduling software in other sectors. Companies in industries such as hospitality and retail respond to real-time analysis of factors such as sales and weather and modify their staffing accordingly. However, this increased efficiency when it comes to the use of staff resources could lead to significant disruption to the lives of low-wage employees, who may receive as little as a day's notice of their changed timetable. More broadly, the trend towards increased use of 'people analytics' - the comprehensive collection of data about employees' behaviour, which is then used to inform managerial decisions - has been criticised as having a dehumanising effect on work, and may not even be effective in increasing productivity or optimising working practices. Another 'back- end' application of AI in a healthcare system is found in Denmark. In 2016, the

Copenhagen-based start-up Corti partnered with the city's emergency medical service (EMS) to provide an AI assistant to augment human operators receiving emergency calls on the 112 emergency number. Besides helping with triage (see 'Clinical Pathways' section, below), Corti's technology is also being used to oversee and optimise the whole dispatch process, for example by identifying and alerting human operators to errors in the address any emergency response is sent to. An innovative approach to using AI will be to apply it to supporting quality and service improvement. This is currently in the final rounds of a research grant application for investigation by the team at Alder Hey Children's NHS

POTENTIAL BENEFITS OF USING AI IN HEATHCARE SYSTEMS :

Pushing boundaries of human performance. The flashiest use of medical AI is to do things that human providers—even excellent ones—cannot yet do. For instance, Google Health has developed a program that can predict the onset of acute kidney injury up to two days before the injury occurs; compare that to current medical practice, where the injury often isn't noticed until *after* it happens Su.ch algorithms can improve care beyond the current boundaries of human performance.

Democratizing medical knowledge and excellence. AI can also share the expertise and performance of specialists to supplement providers who might otherwise lack that expertise. Ophthalmology and radiology are popular targets, especially because AI image-analysis techniques have long been a focus of development. Several programs use images of the human eye to give diagnoses that otherwise would require an ophthalmologist. Using these programs, general practitioner, technician, or even a patient can reach that conclusion. Such democratization matters because specialists, especially highly skilled experts, are relatively rare compared to need in many areas.

Automating drudgery in medical practice. AI can automate some of the computer tasks that take up much of medical practice today. Providers spend a tremendous amount of time dealing with electronic medical records, reading screens, and typing on keyboards, even in the exam room. If AI systems can queue up the most relevant information in patient records and then distill recordings of appointments and conversations down into structured data, they could save substantial time for providers and might increase the amount of facetime between providers and patients and the quality of the medical encounter for both.

Managing patients and medical resources. Finally, and least visibly to the public, AI can be used to allocate resources and shape business. For instance, AI systems might predict which departments are likely to need additional short-term staffing, suggest which of two patients might benefit most from scarce medical resources, or, more controversially, identify revenue-maximizing practices.

RISKS AND CHALLENGES OF USING AI IN HEALTHCARE SYSTEMS:

While AI offers a number of possible benefits, there also are several risks:

Data availability. Training AI systems requires large amounts of data from sources such as electronic health records, pharmacy records, insurance claims records, or consumer-generated information like fitness trackers or purchasing history. But health data are often problematic. Data are typically gmented across many different systems. Even aside from the variety just mentioned, patients typically see different providers and switch insurance companies, leading to data split in multiple systems and multiple formats. This fragmentation increases the risk of error, decreases the comprehensiveness of datasets, and increases the expense of gathering data—which also limits the types of entities that can develop effective health-care AI.

RISKS AND CHALLENGES OF USING AI IN HEALTHCARE SYSTEMS:

Privacy concerns. Another set of risks arise around privacy.The requirement of large datasets creates incentives for developers to collect such data from many patients. Some patients may be concerned that this collection may violate their privacy, and lawsuits have been filed based on data-sharing between large health systems and AI developers.AI could implicate privacy in another way: AI can predict private information about patients even though the algorithm never received that information. (Indeed, this is often the goal of health-care AI.) For instance, an AI system might be able to identify that a person has Parkinson's disease based on the trembling of a computer mouse, even if the person had never revealed that information to anyone else (or did not know). Patients might consider this a violation of their privacy, especially if the AI system's inference were available to third parties, such as banks or life insurance companies.

Bias and inequality. There are risks involving bias and inequality in health-care AI. AI systems learn from the data on which they are trained, and they can incorporate biases from those data. For instance, if the data available for AI are principally gathered in academic medical centers, the resulting AI systems will know less about— and therefore will treat less effectively—patients from populations that do not typically frequent academic medical centers. Similarly, if speech-recognition AI systems are used to transcribe encounter notes, such AI may perform worse when the provider is of a race or gender underrepresented in training data. Even if AI systems learn from accurate, representative data, there can still be problems if that information reflects underlying biases and inequalities in the health system. For example, African- American patients receive, on average, less treatment for pain than white patients; an AI system learning from health-system records might learn to suggest lower doses of painkillers to African-American patients even though that decision reflects systemic bias, not biological reality. Resource-allocation AI systems could also exacerbate inequality by assigning fewer resources to patients considered less desirable or less profitable by health systems for a variety of problematic reasons.

Professional realignment. Longer-term risks involve shifts in the medical profession. Some medical specialties, such as radiology, are likely to shift substantially as much of their work becomes automatable. Some scholars are concerned that the widespread use of AI will result in decreased human knowledge and capacity over time, such that providers lose the ability to catch and correct AI errors and further to develop medical knowledge.

The nirvana fallacy. One final risk bears mention. AI has the potential for tremendous good in health care. The nirvana fallacy posits that problems arise when policymakers

and others compare a new option to perfection, rather than the status quo. Health-care AI faces risks and challenges. But the current system is also rife with problems. Doing nothing because AI is imperfect creates the risk of perpetuating a problematic status quo.

POSSIBLE SOLUTIONS :

There are several ways we can deal with possible risks of health-care AI:

Data generation and availability. Several risks arise from the difficulty of assembling high-quality data in a manner consistent with protecting patient privacy. One set of potential solutions turns on government provision of infrastructural resources for data, ranging from setting standards for electronic health records to directly providing technical support for high-quality data-gathering efforts in health systems that otherwise lack those resources. A parallel option is direct investment in the creation of high-quality datasets. Reflecting this direction, both the United States' All of Us initiative and the U.K.'s BioBank aim to collect comprehensive health-care data on huge numbers of individuals. Ensuring effective privacy safeguards for these large-scale datasets will likely be essential to ensuring patient trust and participation.

Quality oversight. Oversight of AI-system quality will help address the risk of patient injury. The Food and Drug Administration (FDA) oversees some health-care AI products that are commercially marketed. The agency has already cleared several products for market entry, and it is thinking creatively about how best to oversee AI systems in health. However, many AI systems in health care will not fall under FDA's purview, either because they do not perform medical functions (in the case of back-end business or resource-allocation AI) or because they are developed and deployed in- house at health systems themselves—a category of products FDA typically does not oversee. These health-care AI systems fall into something of an oversight gap. Increased oversight efforts by health systems and hospitals, professional organizations like the American College of Radiology and the American Medical Association, or

insurers may be necessary to ensure quality of systems that fall outside the FDA's exercise of regulatory authority.

Provider engagement and education. The integration of AI into the health system will undoubtedly change the role of health-care providers. A hopeful vision is that providers will be enabled to provide more-personalized and better care, freed to spend more time interacting with patients as humans.A less hopeful vision would see providers struggling to weather a monsoon of uninterpretable predictions and recommendations from competing algorithms. In either case—or in any option in- between—medical education will need to prepare providers to evaluate and interpret the AI systems they will encounter in the evolving health-care environment.

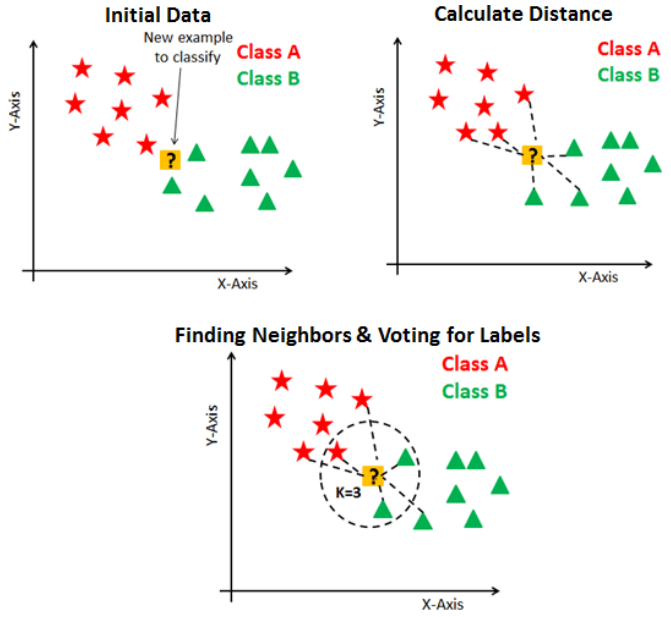
Process

MODELS USED:

The following methodology will be followed to achieve the objectives defined for proposedresearch work:

1. ***K-nearest neighbors (KNN)***
   The K-nearest neighbors (KNN) algorithm used is a type of supervised machine learning algorithm.

It simply calculated the distance of a new data point to all other training data points. The distance can be of Euclidean or Manhattan type. After this, it selects the Knearest data points, where K can be any integer. Lastly, it assigns the data point to the class to which the majority of K data points belong. The below figure shows how to model using KNN:



2. ***Naïve Bayes***

It is a machine learning algorithm for classification problems and is based on Bayes' probability theorem. The primary use of this is to do text classification which involves high dimensional training data sets. We used the Bayes theorem that can be defined as:

$$P(h|d) = P(d|h) \cdot P(h) / P(d)$$

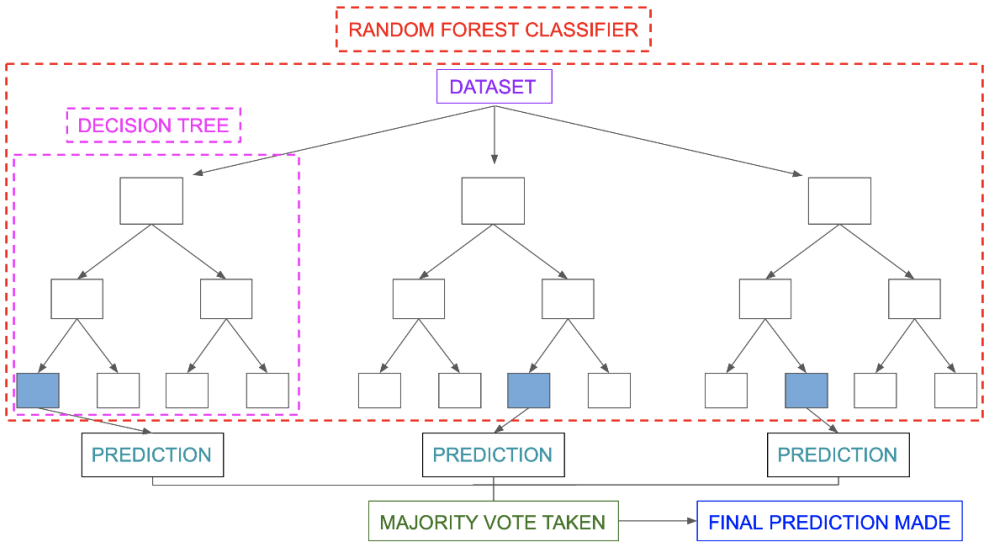
where P(h|d) is the probability of hypothesis h given the data d. This is called the posterior probability. P(d|h) is the probability of data d given that the hypothesis h was true. P(h) is the probability of hypothesis h being true (regardless of the data). This is called the prior probability of h. P(d) is the probability of the data (regardless of the hypothesis).

3. *Decision Trees*

Decision trees algorithm belongs to the family of supervised learning algorithms. It is used for regression and classification. In the decision tree, for prediction, it uses the method of tree diagram at the top. It contains a root node after which it gets split in the dominant input feature and then it again gets split. These processes continue till all input is placed and at the node, the extreme last node contains the weights on the bases of these weights it classifies the input. In a coarse tree, the maximum number of splits from each node is 4. Whereas in a Medium tree, the maximum number of splits from each node is 20. In a fine tree, the maximum number of splits from each node is 100.

4. *Random Forest Classfier*

Random Forest algorithm developed from trees algorithm and bagging algorithm is modelled. The developed the algorithm found that it can potentially improve classification accuracy. It is also work well with a data set with large number of input variables. The algorithm is started by creating a combination of trees which each will vote for a class as shown in Fig. The figure presents how to model the Random Forest. Suppose that there are N data and M input variables in a data set where the real data used in this paper compose of data and input variables. Let k be the number of sampling groups, ni and mi be number of data and variables in group i where i is equal to 1, 2, ... and k.



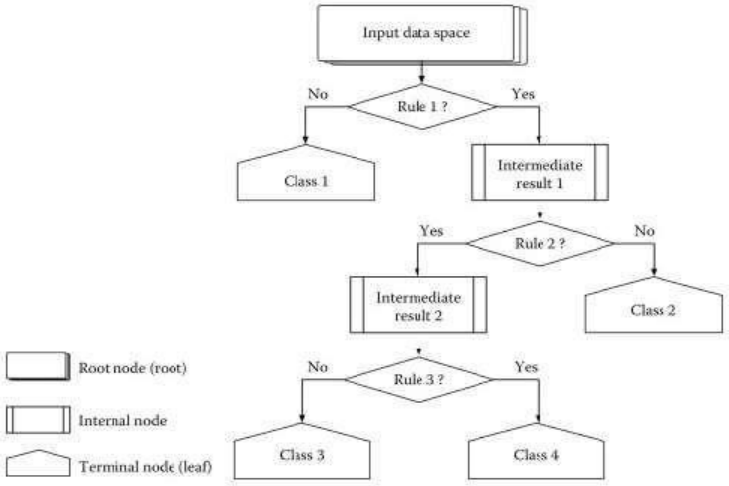
Models Working In Our Project :

There are four different kind of models present in our project to predict the disease these are

- Decision tree
- Random forest tree
- Gaussian Naïve Bayes
- KNN

Decision tree is classified as a very effective and versatile classification technique. It is used in pattern recognition and classification for image. It is used for classification in very complex problems dew to its high adaptability. It is also capable of engaging problems of higher dimensionality. It mainly consists of three parts root, nodes and leaf.

Roots consists of attribute which has most effect on the outcome, leaf tests for value of certain attribute and leaf gives out the output of tree.



Decision tree is the first prediction method we have used in our project. It gives us an accuracyof ~95%.

Random Forest Algorithm is a supervised learning algorithm used for both classification and regression. This algorithm works on 4 basic steps –

1. It chooses random data samples from dataset.
2. It constructs decision trees for every sample dataset chosen.
3. At this step every predicted result will be compiled and voted on.
4. At last most voted prediction will be selected and be presented as result of classification.

In this project we have used random forest classifier with 100 random samples and the result given is ~95% accuracy.

K Nearest Neighbour is a supervised learning algorithm. It is a basic yet essential algorithm. It finds extensive use in pattern finding and data mining.

It works by finding a pattern in data which links data to results and it improves upon the patter recognition with every iteration.

We have used K Nearest Neighbour to classify our dataset and achieved ~95% accuracy.

Naïve Bayes algorithm is a family of algorithms based on naïve bayes theorem. They share a common principle that is every pair of prediction is independent of each other. It also makes an assumption that features make an independent and equal contribution to the prediction.

In our project we have used naïve bayes algorithm to gain a ~95% accurate prediction.

Project Workflow :

DataFrame -> Preprocessing of the Data -> Visualization -> Building Models

Step 1 - DataFrame

```
In [8]: DF.head()
Out[8]:
```

prognosis	itching	skin_rash	nodal_skin_eruptions	continuous_sneezing	shivering	chills	joint_pain	stomach_pain	acidity	ulcers_on_tongue	...	pus_fil
Fungal infection	1	1	1	0	0	0	0	0	0	0	...	
Fungal infection	0	1	1	0	0	0	0	0	0	0	...	
Fungal infection	1	0	1	0	0	0	0	0	0	0	...	
Fungal infection	1	1	0	0	0	0	0	0	0	0	...	
Fungal infection	1	1	1	0	0	0	0	0	0	0	...	

5 rows × 132 columns

Step 2 – Preprocessing

```
In [6]: l2=[]
        for i in range(0,len(l1)):
            l2.append(0)
        print(l2)

[0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]
```

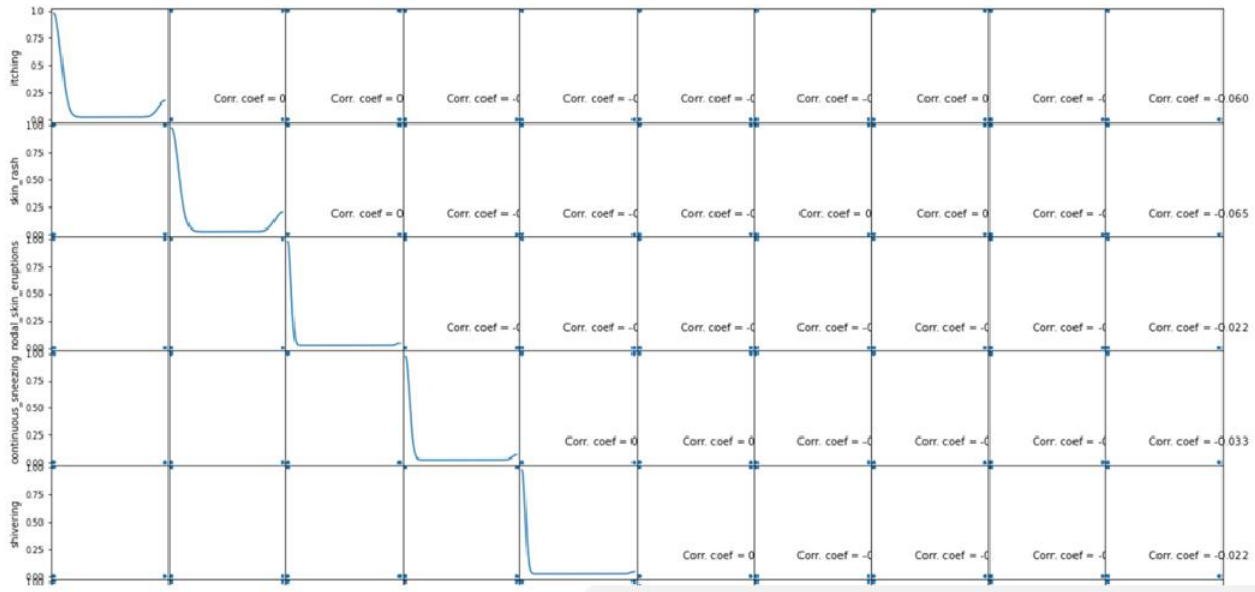
```
In [7]: #Reading the training .csv file
        df=pd.read_csv("D:/Areesha/training.csv")
        DF= pd.read_csv('D:/Areesha/training.csv', index_col='prognosis')

        #Label encoding the disease column
        df.replace({'prognosis':{'Fungal infection':0,'Allergy':1,'GERD':2,'Chronic cholestasis':3,'Drug Reaction':4,
            'Peptic ulcer diseae':5,'AIDS':6,'Diabetes ':7,'Gastroenteritis':8,'Bronchial Asthma':9,'Hypertension ':10,
            'Migraine':11,'Cervical spondylosis':12,
            'Paralysis (brain hemorrhage)':13,'Jaundice':14,'Malaria':15,'Chicken pox':16,'Dengue':17,'Typhoid':18,'hepatitis A':19,
            'Hepatitis B':20,'Hepatitis C':21,'Hepatitis D':22,'Hepatitis E':23,'Alcoholic hepatitis':24,'Tuberculosis':25,
            'Common Cold':26,'Pneumonia':27,'Dimorphic hemmorhoids(piles)':28,'Heart attack':29,'Varicose veins':30,'Hypothyroidism':
            'Hyperthyroidism':32,'Hypoglycemia':33,'Osteoarthristis':34,'Arthritis':35,
            '(vertigo) Paroymsal  Positional Vertigo':36,'Acne':37,'Urinary tract infection':38,'Psoriasis':39,
            'Impetigo':40}},inplace=True)
```

Step 3 – Visualization

DATA VISUALIZATION

In [10]:
```
# Scatter and density plots
def plotScatterMatrix(df1, plotSize, textSize):
    df1 = df1.select_dtypes(include =[np.number]) # keep only numerical columns
    # Remove rows and columns that would lead to df being singular
    df1 = df1.dropna('columns')
    df1 = df1[[col for col in df if df[col].nunique() > 1]] # keep columns where there are more than 1 unique values
    columnNames = list(df)
    if len(columnNames) > 10: # reduce the number of columns for matrix inversion of kernel density plots
        columnNames = columnNames[:10]
    df1 = df1[columnNames]
    ax = pd.plotting.scatter_matrix(df1, alpha=0.75, figsize=[plotSize, plotSize], diagonal='kde')
    corrs = df1.corr().values
    for i, j in zip(*plt.np.triu_indices_from(ax, k = 1)):
        ax[i, j].annotate('Corr. coef = %.3f' % corrs[i, j], (0.8, 0.2), xycoords='axes fraction', ha='center', va='center',
    plt.suptitle('Scatter and Density Plot')
    plt.show()
```

In [11]:
```
plotScatterMatrix(df, 20, 10)
```



Scatter and Density Plot

Step 4 – Building Models

Decision Trees

```
from sklearn import tree

clf3 = tree.DecisionTreeClassifier()
clf3 = clf3.fit(X,y)

from sklearn.metrics import classification_report,confusion_matrix,accuracy_score
y_pred=clf3.predict(X_test)
print("Decision Tree")
print("Accuracy")
print(accuracy_score(y_test, y_pred))
print(accuracy_score(y_test, y_pred,normalize=False))
print("Confusion matrix")
conf_matrix=confusion_matrix(y_test,y_pred)
print(conf_matrix)

psymptoms = [Symptom1.get(),Symptom2.get(),Symptom3.get(),Symptom4.get(),Symptom5.get()]

for k in range(0,len(l1)):
    for z in psymptoms:
        if(z==l1[k]):
            l2[k]=1

inputtest = [l2]
predict = clf3.predict(inputtest)
predicted=predict[0]
```

Random Forest

```
from sklearn.ensemble import RandomForestClassifier
clf4 = RandomForestClassifier(n_estimators=100)
clf4 = clf4.fit(X,np.ravel(y))

# calculating accuracy
from sklearn.metrics import classification_report,confusion_matrix,accuracy_score
y_pred=clf4.predict(X_test)
print("Random Forest")
print("Accuracy")
print(accuracy_score(y_test, y_pred))
print(accuracy_score(y_test, y_pred,normalize=False))
print("Confusion matrix")
conf_matrix=confusion_matrix(y_test,y_pred)
print(conf_matrix)

psymptoms = [Symptom1.get(),Symptom2.get(),Symptom3.get(),Symptom4.get(),Symptom5.get()]

for k in range(0,len(l1)):
    for z in psymptoms:
        if(z==l1[k]):
            l2[k]=1

inputtest = [l2]
predict = clf4.predict(inputtest)
predicted=predict[0]
```

Naïve Bayes

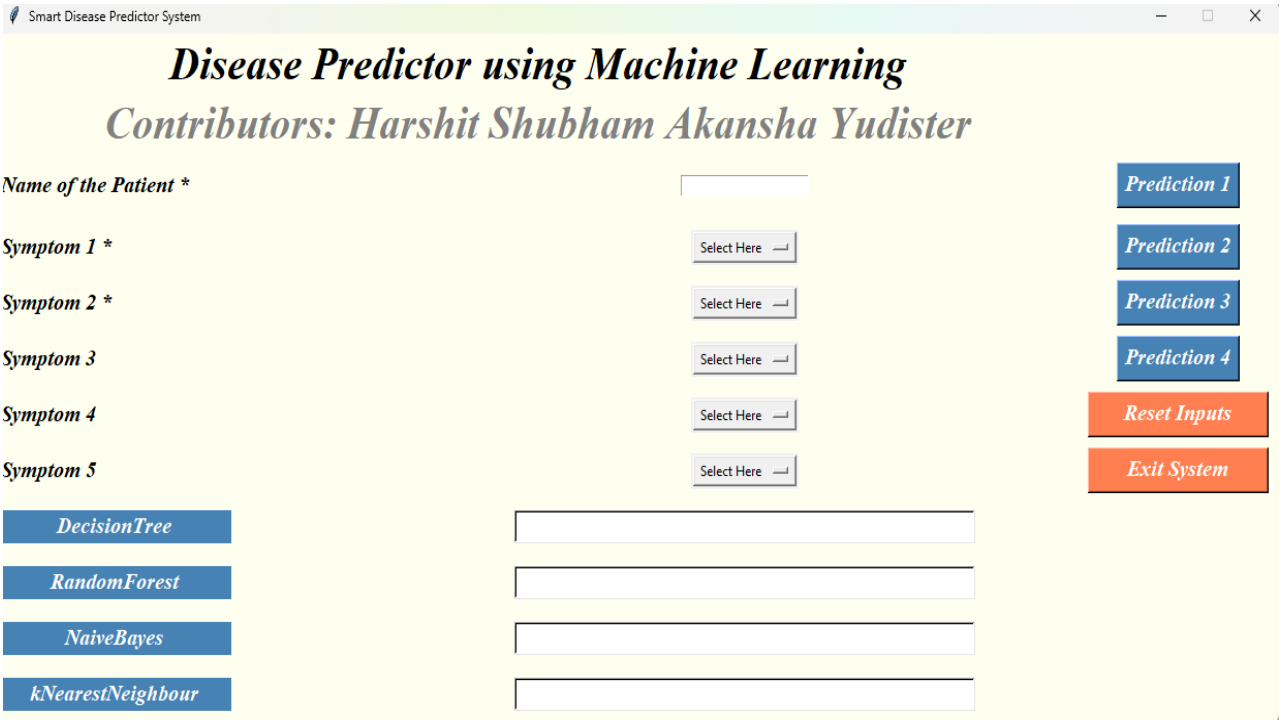
```
from sklearn.naive_bayes import GaussianNB
gnb = GaussianNB()
gnb=gnb.fit(X,np.ravel(y))

from sklearn.metrics import classification_report,confusion_matrix,accuracy_score
y_pred=gnb.predict(X_test)
print("Naive Bayes")
print("Accuracy")
print(accuracy_score(y_test, y_pred))
print(accuracy_score(y_test, y_pred,normalize=False))
print("Confusion matrix")
conf_matrix=confusion_matrix(y_test,y_pred)
print(conf_matrix)
psymptoms = [Symptom1.get(),Symptom2.get(),Symptom3.get(),Symptom4.get(),Symptom5.get()]
for k in range(0,len(l1)):
    for z in psymptoms:
        if(z==l1[k]):
            l2[k]=1

inputtest = [l2]
predict = gnb.predict(inputtest)
predicted=predict[0]
```

K-Nearest Neighbour

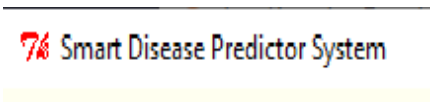
```
from sklearn.neighbors import KNeighborsClassifier
knn=KNeighborsClassifier(n_neighbors=5,metric='minkowski',p=2)
knn=knn.fit(X,np.ravel(y))

from sklearn.metrics import classification_report,confusion_matrix,accuracy_score
y_pred=knn.predict(X_test)
print("kNearest Neighbour")
print("Accuracy")
print(accuracy_score(y_test, y_pred))
print(accuracy_score(y_test, y_pred,normalize=False))
print("Confusion matrix")
conf_matrix=confusion_matrix(y_test,y_pred)
print(conf_matrix)

psymptoms = [Symptom1.get(),Symptom2.get(),Symptom3.get(),Symptom4.get(),Symptom5.get()]

for k in range(0,len(l1)):
    for z in psymptoms:
        if(z==l1[k]):
            l2[k]=1

inputtest = [l2]
predict = knn.predict(inputtest)
predicted=predict[0]
```

Graphical User Interface:

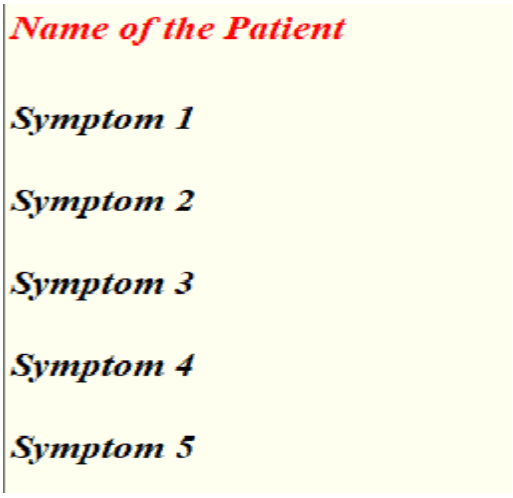
GUI made for this project is a simple tkinter GUI consisting of labels, messagebox, button, text, title and option menu



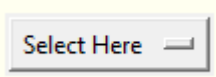
Root.title() is used to set the the title as Smart Disease Predictor System



Labels are further used for different sections



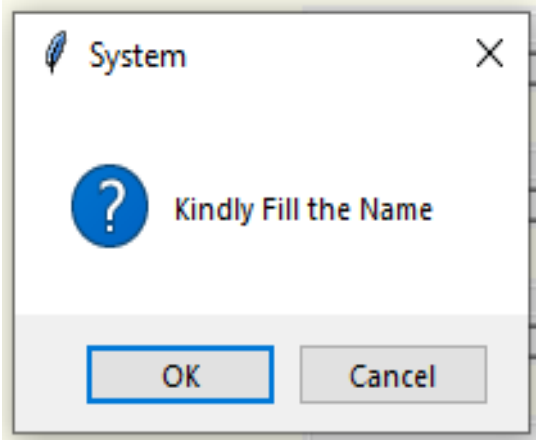
OptionMenu is used to create drop down menu



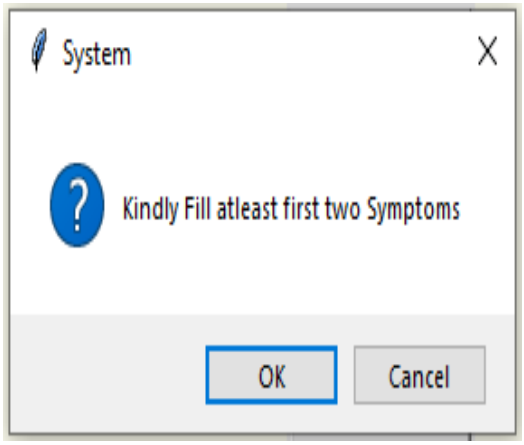
Buttons are used to give functionalities and predict the out come of models also two utilitybuttons namely exit and rest are also created.

Text is used to show output of the prediction using blank space.

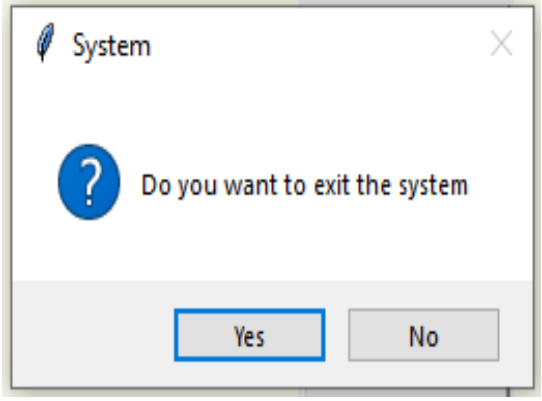
Messagebox are used at three different places, one- to restrain the to enter name



two- to ask for at least two symptoms,

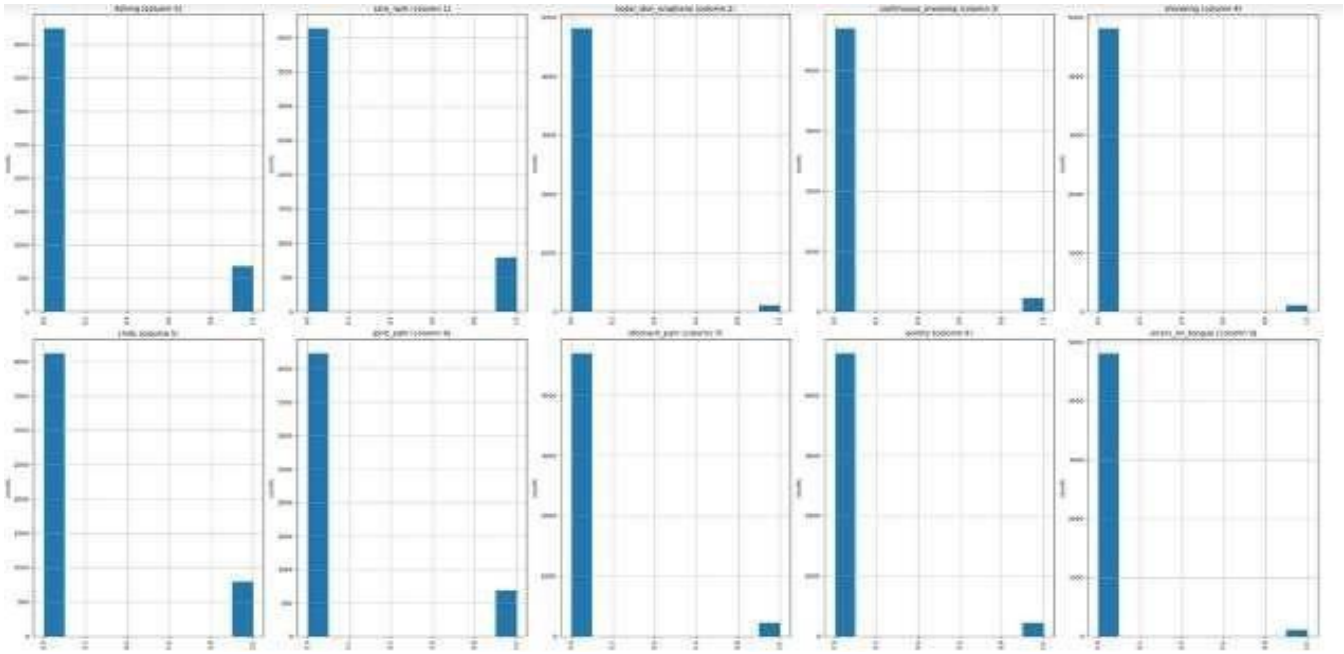


three- to confirm to exit system.



Modules

```
# Distribution graphs (histogram/bar graph) of column data
def plotPerColumnDistribution(df1, nGraphShown, nGraphPerRow):
    nunique = df1.nunique()
    df1 = df1[[col for col in df if nunique[col] > 1 and nunique[col] < 50]] # For displaying purposes, pick columns that have
    nRow, nCol = df1.shape
    columnNames = list(df1)
    nGraphRow = (nCol + nGraphPerRow - 1) / nGraphPerRow
    plt.figure(num = None, figsize = (6 * nGraphPerRow, 8 * nGraphRow), dpi = 80, facecolor = 'w', edgecolor = 'k')
    for i in range(min(nCol, nGraphShown)):
        plt.subplot(nGraphRow, nGraphPerRow, i + 1)
        columnDf = df.iloc[:, i]
        if (not np.issubdtype(type(columnDf.iloc[0]), np.number)):
            valueCounts = columnDf.value_counts()
            valueCounts.plot.bar()
        else:
            columnDf.hist()
        plt.ylabel('counts')
        plt.xticks(rotation = 90)
        plt.title(f'{columnNames[i]} (column {i})')
    plt.tight_layout(pad = 1.0, w_pad = 1.0, h_pad = 1.0)
    plt.show()
    print(nunique)
```



Functions like plotpercolumbdistribution() plotScatterMatrix() is used to viaualize the data.

```
#list1 = DF['prognosis'].unique()
def scatterplt(disea):
    x = ((DF.loc[disea]).sum())
    x.drop(x[x==0].index,inplace=True)
    print(x.values)
    y = x.keys()
    print(len(x))
    print(len(y))
    plt.title(disea)
    plt.scatter(y,x.values)
    plt.show
```

```
def scatterinp(sym1,sym2,sym3,sym4,sym5):
    x = [sym1,sym2,sym3,sym4,sym5]
    y = [0,0,0,0,0]
    if(sym1!='Select Here'):
        y[0]=1
    if(sym2!='Select Here'):
        y[1]=1
    if(sym3!='Select Here'):
        y[2]=1
    if(sym4!='Select Here'):
        y[3]=1
    if(sym5!='Select Here'):
        y[4]=1
    plt.scatter(x,y)
    plt.show
```

Function like scatterplt and scatterinp are used to compare input to training data.

Decision Tree Algorithm

```
root = Tk()
pred1=StringVar()
def DecisionTree():
    if len(NameEn.get()) == 0:
        pred1.set(" ")
        comp=messagebox.askokcancel("System","Kindly Fill the Name")
        if comp:
            root.mainloop()
    elif((Symptom1.get()=="Select Here") or (Symptom2.get()=="Select Here")):
        pred1.set(" ")
        sym=messagebox.askokcancel("System","Kindly Fill atleast first two Symptoms")
        if sym:
            root.mainloop()
    else:
        from sklearn import tree

        clf3 = tree.DecisionTreeClassifier()
        clf3 = clf3.fit(X,y)

        from sklearn.metrics import classification_report,confusion_matrix,accuracy_score
        y_pred=clf3.predict(X_test)
        print("Decision Tree")
        print("Accuracy")
        print(accuracy_score(y_test, y_pred))
        print(accuracy_score(y_test, y_pred,normalize=False))
        print("Confusion matrix")
        conf_matrix=confusion_matrix(y_test,y_pred)
        print(conf_matrix)
```

```
    for k in range(0,len(l1)):
        for z in psymptoms:
            if(z==l1[k]):
                l2[k]=1

    inputtest = [l2]
    predict = clf3.predict(inputtest)
    predicted=predict[0]

    h='no'
    for a in range(0,len(disease)):
        if(predicted == a):
            h='yes'
            break


    if (h=='yes'):
        pred1.set(" ")
        pred1.set(disease[a])

    else:
        pred1.set(" ")
        pred1.set("Not Found")
    #Creating the database if not exists named as database.db and creating table if not exists named as DecisionTree usin
    import sqlite3
    conn = sqlite3.connect('database.db')
    c = conn.cursor()
    c.execute("CREATE TABLE IF NOT EXISTS DecisionTree(Name StringVar,Symtom1 StringVar,Symtom2 StringVar,Symtom3 StringV
    c.execute("INSERT INTO DecisionTree(Name,Symtom1,Symtom2,Symtom3,Symtom4,Symtom5,Disease) VALUES(?,?,?,?,?,?,?)",(Name
    conn.commit()
    c.close()
```

Algorithm of decision tree and database storage.

Random Forest Algorithm

```
pred2=StringVar()
def randomforest():
    if len(NameEn.get()) == 0:
        pred1.set(" ")
        comp=messagebox.askokcancel("System","Kindly Fill the Name")
        if comp:
            root.mainloop()
    elif((Symptom1.get()=="Select Here") or (Symptom2.get()=="Select Here")):
        pred1.set(" ")
        sym=messagebox.askokcancel("System","Kindly Fill atleast first two Symptoms")
        if sym:
            root.mainloop()
    else:
        from sklearn.ensemble import RandomForestClassifier
        clf4 = RandomForestClassifier(n_estimators=100)
        clf4 = clf4.fit(X,np.ravel(y))

        # calculating accuracy
        from sklearn.metrics import classification_report,confusion_matrix,accuracy_score
        y_pred=clf4.predict(X_test)
        print("Random Forest")
        print("Accuracy")
        print(accuracy_score(y_test, y_pred))
        print(accuracy_score(y_test, y_pred,normalize=False))
        print("Confusion matrix")
        conf_matrix=confusion_matrix(y_test,y_pred)
        print(conf_matrix)
```

Algorithm of random forest classifier.

KNearestNeighbour Algorithm

```
pred4=StringVar()
def KNN():
    if len(NameEn.get()) == 0:
        pred1.set(" ")
        comp=messagebox.askokcancel("System","Kindly Fill the Name")
        if comp:
            root.mainloop()
    elif((Symptom1.get()=="Select Here") or (Symptom2.get()=="Select Here")):
        pred1.set(" ")
        sym=messagebox.askokcancel("System","Kindly Fill atleast first two Symptoms")
        if sym:
            root.mainloop()
    else:
        from sklearn.neighbors import KNeighborsClassifier
        knn=KNeighborsClassifier(n_neighbors=5,metric='minkowski',p=2)
        knn=knn.fit(X,np.ravel(y))

        from sklearn.metrics import classification_report,confusion_matrix,accuracy_score
        y_pred=knn.predict(X_test)
        print("KNN")
        print("Accuracy")
        print(accuracy_score(y_test, y_pred))
        print(accuracy_score(y_test, y_pred,normalize=False))
        print("Confusion matrix")
        conf_matrix=confusion_matrix(y_test,y_pred)
        print(conf_matrix)

        psymptoms = [Symptom1.get(),Symptom2.get(),Symptom3.get(),Symptom4.get(),Symptom5.get()]
```

Algorithm of K nearest neighbour.

Naive Bayes Algorithm

```
pred3=StringVar()
def NaiveBayes():
    if len(NameEn.get()) == 0:
        pred1.set(" ")
        comp=messagebox.askokcancel("System","Kindly Fill the Name")
        if comp:
            root.mainloop()
    elif((Symptom1.get()=="Select Here") or (Symptom2.get()=="Select Here")):
        pred1.set(" ")
        sym=messagebox.askokcancel("System","Kindly Fill atleast first two Symptoms")
        if sym:
            root.mainloop()
    else:
        from sklearn.naive_bayes import GaussianNB
        gnb = GaussianNB()
        gnb=gnb.fit(X,np.ravel(y))

        from sklearn.metrics import classification_report,confusion_matrix,accuracy_score
        y_pred=gnb.predict(X_test)
        print("Naive Bayes")
        print("Accuracy")
        print(accuracy_score(y_test, y_pred))
        print(accuracy_score(y_test, y_pred,normalize=False))
        print("Confusion matrix")
        conf_matrix=confusion_matrix(y_test,y_pred)
        print(conf_matrix)

        psymptoms = [Symptom1.get(),Symptom2.get(),Symptom3.get(),Symptom4.get(),Symptom5.get()]
```

Algorithm of naïve bayes classifier

All these classifier is connected to database and GUI to function seamlessly.

```
Symptom1 = StringVar()
Symptom1.set("Select Here")

Symptom2 = StringVar()
Symptom2.set("Select Here")

Symptom3 = StringVar()
Symptom3.set("Select Here")

Symptom4 = StringVar()
Symptom4.set("Select Here")

Symptom5 = StringVar()
Symptom5.set("Select Here")
Name = StringVar()
```

```
prev_win=None
def Reset():
    global prev_win

    Symptom1.set("Select Here")
    Symptom2.set("Select Here")
    Symptom3.set("Select Here")
    Symptom4.set("Select Here")
    Symptom5.set("Select Here")
    NameEn.delete(first=0,last=100)
    pred1.set(" ")
    pred2.set(" ")
    pred3.set(" ")
    pred4.set(" ")
```

Code of GUI to set initial values of labels.

```
from tkinter import messagebox
def Exit():
    qExit=messagebox.askyesno("System","Do you want to exit the system")

    if qExit:
        root.destroy()
        exit()
```

Code of message box.

```
#Taking name as input from user
NameEn = Entry(root, textvariable=Name)
NameEn.grid(row=6, column=1)

#Taking Symptoms as input from the dropdown from the user
S1 = OptionMenu(root, Symptom1,*OPTIONS)
S1.grid(row=7, column=1)

S2 = OptionMenu(root, Symptom2,*OPTIONS)
S2.grid(row=8, column=1)

S3 = OptionMenu(root, Symptom3,*OPTIONS)
S3.grid(row=9, column=1)

S4 = OptionMenu(root, Symptom4,*OPTIONS)
S4.grid(row=10, column=1)

S5 = OptionMenu(root, Symptom5,*OPTIONS)
S5.grid(row=11, column=1)
```

Code of option menu

```
#Buttons for predicting the disease using different algorithms
dst = Button(root, text="Prediction 1", command=DecisionTree,bg="Red",fg="yellow")
dst.config(font=("Times",15,"bold italic"))
dst.grid(row=6, column=3,padx=10)

rnf = Button(root, text="Prediction 2", command=randomforest,bg="Light green",fg="red")
rnf.config(font=("Times",15,"bold italic"))
rnf.grid(row=7, column=3,padx=10)

lr = Button(root, text="Prediction 3", command=NaiveBayes,bg="Blue",fg="white")
lr.config(font=("Times",15,"bold italic"))
lr.grid(row=8, column=3,padx=10)

kn = Button(root, text="Prediction 4", command=KNN,bg="sky blue",fg="red")
kn.config(font=("Times",15,"bold italic"))
kn.grid(row=9, column=3,padx=10)

rs = Button(root,text="Reset Inputs", command=Reset,bg="yellow",fg="purple",width=15)
rs.config(font=("Times",15,"bold italic"))
rs.grid(row=10,column=3,padx=10)

ex = Button(root,text="Exit System", command=Exit,bg="yellow",fg="purple",width=15)
ex.config(font=("Times",15,"bold italic"))
ex.grid(row=11,column=3,padx=10)
```

Code of option menu

```
#Buttons for predicting the disease using different algorithms
dst = Button(root, text="Prediction 1", command=DecisionTree,bg="Red",fg="yellow")
dst.config(font=("Times",15,"bold italic"))
dst.grid(row=6, column=3,padx=10)

rnf = Button(root, text="Prediction 2", command=randomforest,bg="Light green",fg="red")
rnf.config(font=("Times",15,"bold italic"))
rnf.grid(row=7, column=3,padx=10)

lr = Button(root, text="Prediction 3", command=NaiveBayes,bg="Blue",fg="white")
lr.config(font=("Times",15,"bold italic"))
lr.grid(row=8, column=3,padx=10)

kn = Button(root, text="Prediction 4", command=KNN,bg="sky blue",fg="red")
kn.config(font=("Times",15,"bold italic"))
kn.grid(row=9, column=3,padx=10)

rs = Button(root,text="Reset Inputs", command=Reset,bg="yellow",fg="purple",width=15)
rs.config(font=("Times",15,"bold italic"))
rs.grid(row=10,column=3,padx=10)

ex = Button(root,text="Exit System", command=Exit,bg="yellow",fg="purple",width=15)
ex.config(font=("Times",15,"bold italic"))
ex.grid(row=11,column=3,padx=10)
```

Code of buttons.

```
#Showing the output of different aldorithms
t1=Label(root,font=("Times",15,"bold italic"),text="Decision Tree",height=1,bg="Light green"
        ,width=40,fg="red",textvariable=pred1,relief="sunken").grid(row=15, column=1, padx=10)

t2=Label(root,font=("Times",15,"bold italic"),text="Random Forest",height=1,bg="Purple"
        ,width=40,fg="white",textvariable=pred2,relief="sunken").grid(row=17, column=1, padx=10)

t3=Label(root,font=("Times",15,"bold italic"),text="Naive Bayes",height=1,bg="red"
        ,width=40,fg="orange",textvariable=pred3,relief="sunken").grid(row=19, column=1, padx=10)

t4=Label(root,font=("Times",15,"bold italic"),text="kNearest Neighbour",height=1,bg="Blue"
        ,width=40,fg="yellow",textvariable=pred4,relief="sunken").grid(row=21, column=1, padx=10)
```

Code of result display.

Results analysis and validation:

Decision Tree:

```
Decision Tree
Accuracy
0.9512195121951219
39
Confusion matrix
[[1 0 0 ... 0 0 0]
 [0 1 0 ... 0 0 0]
 [0 0 1 ... 0 0 0]
 ...
 [0 0 0 ... 1 0 0]
 [0 0 0 ... 0 1 0]
 [0 0 0 ... 0 0 1]]
['irritability', 'swelling_of_stomach', 'blister', 'dizziness', 'mild_fever']
[1, 1, 1, 1, 1]
```

Random Forest :

```
Random Forest
Accuracy
0.9512195121951219
39
Confusion matrix
[[1 0 0 ... 0 0 0]
 [0 1 0 ... 0 0 0]
 [0 0 0 ... 0 0 0]
 ...
 [0 0 0 ... 1 0 0]
 [0 0 0 ... 0 1 0]
 [0 0 0 ... 0 0 1]]
[114 102 114 114 114]
5
5
```

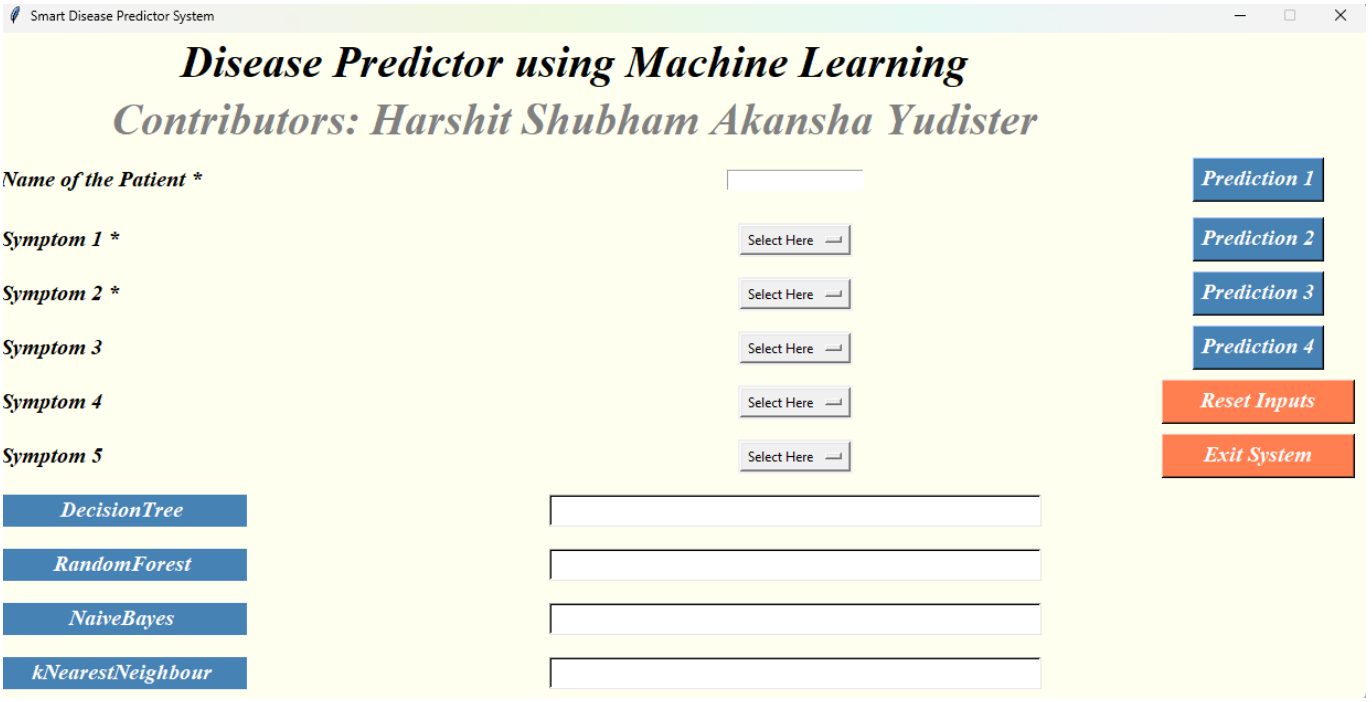
Naïve Bayes :

```
Naive Bayes
Accuracy
0.9512195121951219
39
Confusion matrix
[[1 0 0 ... 0 0 0]
 [0 1 0 ... 0 0 0]
 [0 0 1 ... 0 0 0]
 ...
 [0 0 0 ... 1 0 0]
 [0 0 0 ... 0 1 0]
 [0 0 0 ... 0 0 1]]
[114 102 114 114 114]
5
5
```

K Nearest Neighbour :

```
kNearest Neighbour
Accuracy
0.9512195121951219
39
Confusion matrix
[[1 0 0 ... 0 0 0]
 [0 1 0 ... 0 0 0]
 [0 0 0 ... 0 0 0]
 ...
 [0 0 0 ... 1 0 0]
 [0 0 0 ... 0 1 0]
 [0 0 0 ... 0 0 1]]
[108 108 108 108]
4
4
```

Final GUI :


Smart Disease Predictor System
Disease Predictor using Machine Learning
Contributors: Harshit Shubham Akansha Yudister
Name of the Patient *
Symptom 1 *
Symptom 2 *
Symptom 3
Symptom 4
Symptom 5
Select Here
Prediction 1
Prediction 2
Prediction 3
Prediction 4
Reset Inputs
Exit System
DecisionTree
RandomForest
NaiveBayes
kNearestNeighbour

Discussions

We have presented recently published research studies that employed AI-based Learning techniques for diagnosing the disease in the current review. This study highlights research on disease diagnosis prediction and predicting the post-operative life expectancy of diseased patients using AI-based learning techniques.

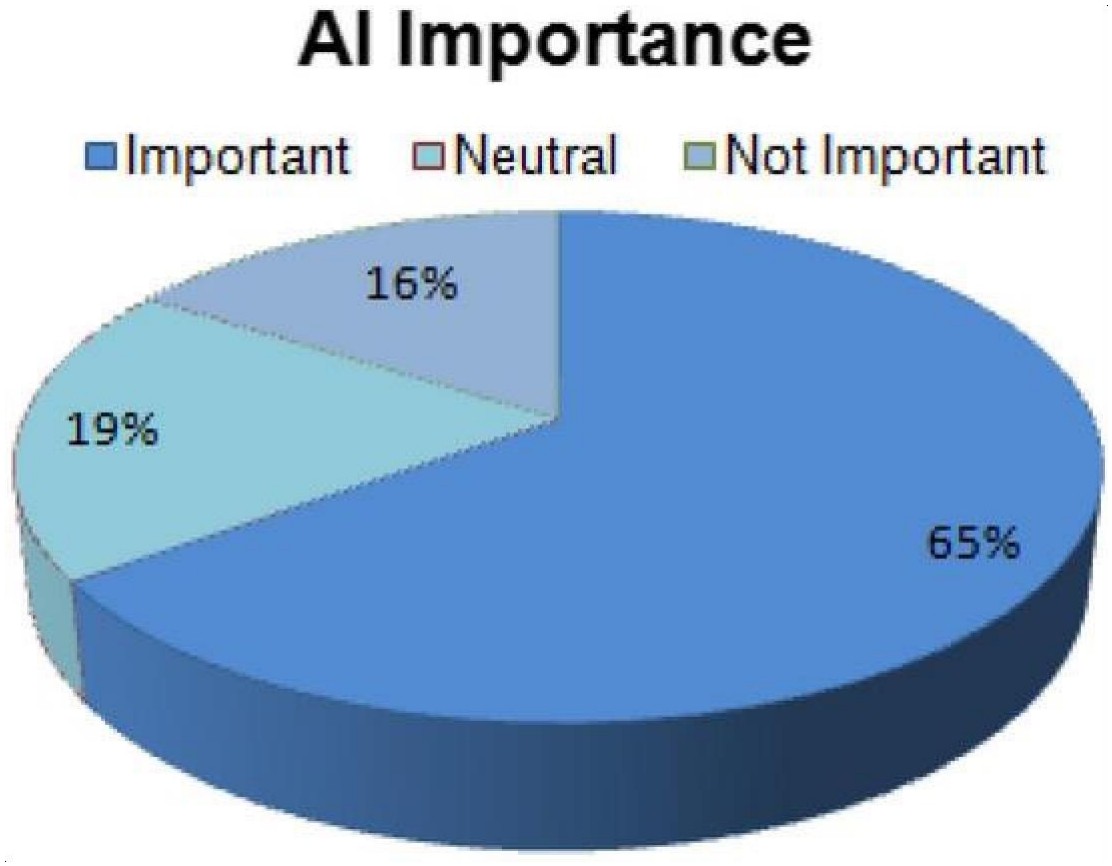
Investigation 1: Why do we need AI?

We know that AI is the simulation of human processes by machines (computer systems) and that this simulation includes learning, reasoning, and self-correction. We require AI since the amount of labour we must perform is rising daily. As a result, it's a good idea to automate regular tasks. It conserves the organization's staff and also boosts production (Vasal et al. 2020).

In terms of the healthcare industry, AI in health refers to a set of diverse technologies that enable robots

to detect, comprehend, act, and learn1 to execute administrative and clinical healthcare activities. AI has the potential to transform healthcare by addressing some of the industry's most pressing issues. For example, AI can result in improved patient outcomes and increased productivity and efficiency in care delivery (Gouda et al. 2020). It can also enhance healthcare practitioners' daily lives by spending more time caring for patients, therefore increasing staff morale and retention. In addition, it may potentially help bring life-saving medicines to market more quickly. Figure 8 shows the significance of AI in the medical field.

Fig. 8



Importance of artificial intelligence in healthcare

Investigation 2: Why is AI important, and how is it used to analyse the disease?

The emergence of new diseases remains a critical parameter in human health and society. Hence, the advances in AI allow for rapid processing and analysis of such massive and complex data. It recommends the correct decision for over ten different diseases (as mentioned in the literature) with at least 98% accuracy.

Doctors use technologies such as computed tomography scan or magnetic resonance imaging to produce a detailed 3D map of the area that needs to be diagnosed. Later, AI technology analyse s the system- generated image using machine and deep learning models to spot the diseased area's features in seconds. As shown in the framework

section, an artificial intelligence model using machine and deep learning algorithms is initially trained with the help of a particular disease dataset (Owasis et al. 2019). The dataset is then pre-processed using data cleaning and transformation techniques so that the disease symptoms in the form of feature vectors can be extracted and further diagnosed.

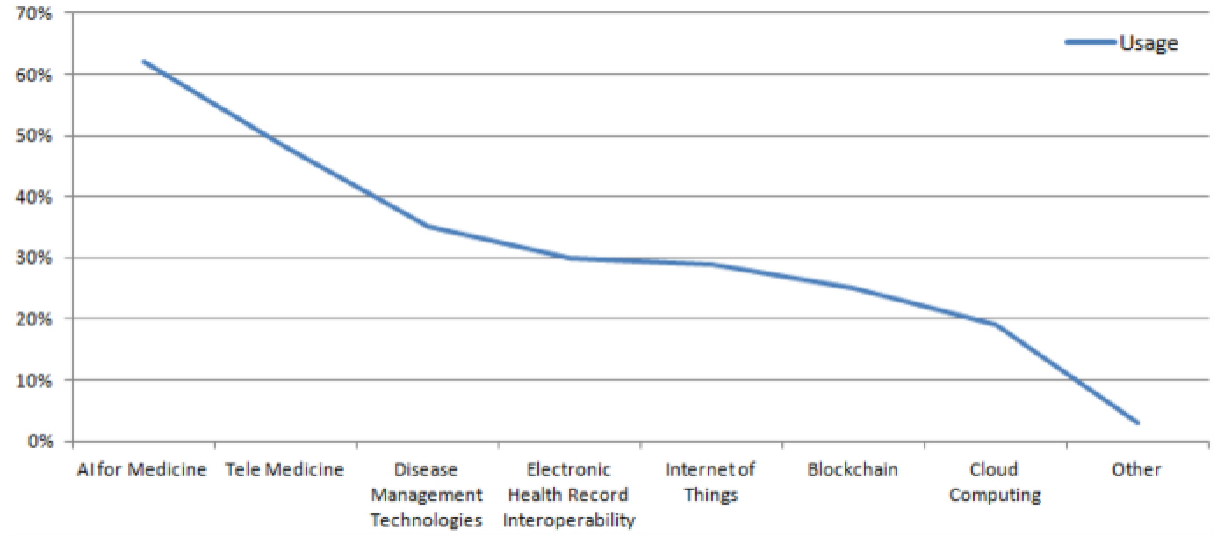
Suppose doctors do not use AI techniques. In that case, it will cause a delay in treating the patients as itis tough to interpret the scanned image manually, and it also takes a considerable amount of time. But,

on the other hand, it shows that an AI technique helps the patients and helps the doctors save thepatient's life by treating them as early as possible (Luo et al. 2019).

Investigation 3: What is the impact of AI in medical diagnosis?

Due to advancements in computer power, learning algorithms, and the availability of massive datasets (big data) derived from medical records and wearable health monitors. The best part of implementing AI in healthcare is that it helps to enhance various areas, including illness detection, disease classification, decision-making processes, giving optimal treatment choices, and ultimately, helping people live longer. In terms of disease diagnosis, AI has been used to enhance medical diagnosis (Chen et al. 2019a, b). For example, the technology, which is currently in use in China, may detect hazardous tumors and nodulesin patients with lung cancer, allowing physicians to provide an early diagnosis rather than sending tissue samples to a lab for testing, allowing for earlier treatment (Keenan et al. 2020). Figure 9 illustrates the influence of artificial intelligence and other approaches.

Fig. 9



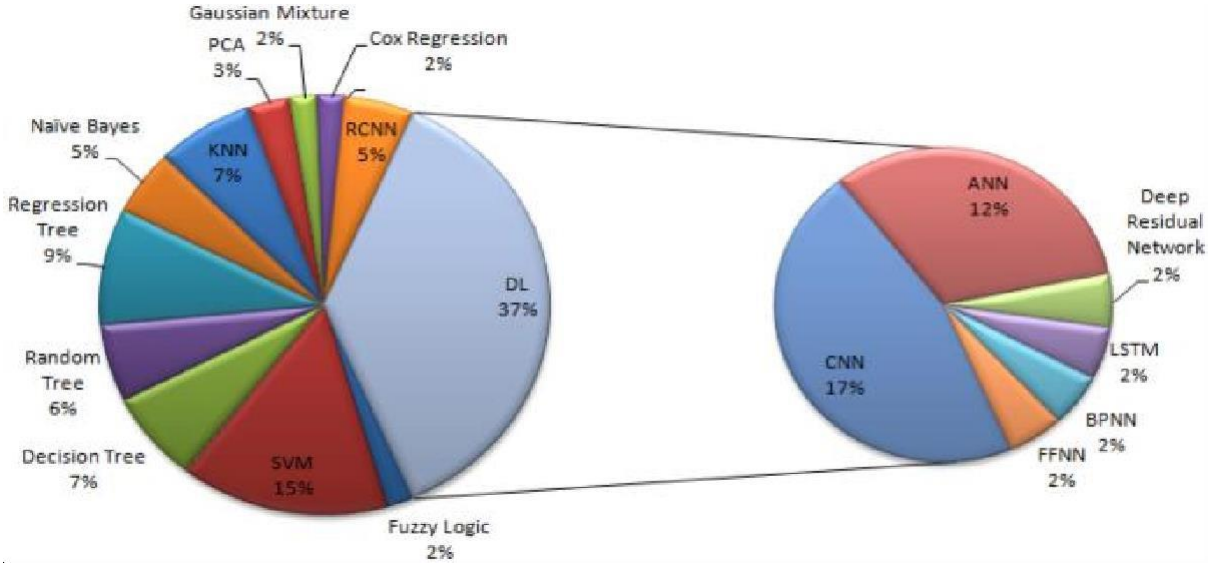
Comparison between AI and other techniques

Investigation 4: Which AI-based algorithm is used in disease diagnosis?

Disease detection algorithms driven by AI demonstrated to be an effective tool for identifying undiagnosed patients with under-diagnosed, uncoded, and rare diseases. Therefore, AI models fordisease detection have an ample opportunity to drive earlier diagnosis for patients in need and guide pharmaceutical companies with highly advanced, targeted diagnostics to help these patients get correctlydiagnosed and treated earlier in their disease journey (Keenan et al. 2020). The research work mentioned in the literature has covered both machine and deep learning models for diagnosing the diseases such as cancer, diabetes, chronic, heart disease, alzheimer, stroke and cerebrovascular, hypertension, skin, and liver disease. Machine learning models,

Random Forest Classifier, Logistic Regression, Fuzzy logics, Gradient Boosting Machines, Decision Tree, K nearest neighbour (KNN), and Support vector machines (SVM) are primarily used in literature. Among deep learning models, Convolutional Neural Networks (CNN) have been used most commonly for disease diagnosis. In addition, faster Recurrent Convolution Neural Network, Multilayer Perceptron, Long Short Term Memory (LSTM) have also been used extensively in the literature. Figure 10 displays the usage of AI-based prediction models in the literature.

Fig. 10



Artificial intelligence-based prediction models

Investigation 5: What are the challenges faced by the researchers while using AI models in several disease diagnosis?

Although AI-based techniques have marked their significance in disease diagnosis, there are still many challenges faced by the researchers that need to be addressed.

- *Limited Data size* The most common challenge faced by most of the studies was insufficient data to train the model. A small sample size implies a smaller training set which does not authenticate the efficiency of the proposed approaches. On the other hand, good sample size can train the model better than the limited one (Rajalakshmi et al. 2018).
- *High dimensionality* Another data-related issue faced in cancer research is high dimensionality. High dimensionality is referred to a vast number of features as compared to cases. However, multiple dimensionality reduction techniques are available to deal with this issue (Bibault et al. 2020).
- *Efficient feature selection technique* Many studies have achieved exceptional prediction outcomes. However, a computationally effective feature selection method is required to eradicate the data cleaning procedures while generating high disease prediction accuracy (Koshimizu et al.2020).
- *Model Generalizability* A shift in research towards improving the generalizability of the model is required. Most of the studies have proposed a prediction model that is validated on a single site. There is a need to validate the models on multiple sites that can help improve the model's generalizability (Fukuda et al. 2019).

- *Clinical Implementation* AI-based models have proved their dominance in medical research; still,the practical implementation of the models in the clinics is not incorporated. These models need to be validated in a clinical setting to assist the medical practitioner in affirming the diagnosis verdicts (Huang et al. 2020).

Investigation 6: How artificial intelligence-based techniques are helping doctors in diagnosing diseases?

AI improves the lives of patients, physicians, and hospital managers by doing activities usually performed by people but in a fraction of the time and the expense. For example, AI assists physicians in making suggestions by evaluating vast amounts of healthcare data such as electronic health records, symptom data, and physician reports to improve health outcomes and eventually save the patient's life (Kohlberger et al. 2019). Additionally, this data aids in the improvement and acceleration of decision- making while diagnosing and treating patients' illnesses using artificial intelligence-based approaches. Not only that, AI assists physicians in detecting diseases by utilizing complicated algorithms, hundreds of biomarkers, imaging findings from millions of patients, aggregated published clinical studies, and thousands of physicians' notes to improve the accuracy of diagnosis.

Management of the Project

Our Project Management includes five stages :

Project initiation

Initiation is the formal start of a project. It usually begins with the issue of a **project mandate** which briefly describes the purpose of the project and authorises budget spend.

At this stage, you should define the project at a broad level. This often begins with:

- a **business case** - justifying the need for the project and estimating potential benefits
- a **feasibility study** - evaluating the problem and determining if the project will solve it

If you decide to undertake the project, you should then create a **project initiation document (PID)**. This is the foundation of your project and a critical reference point for the next stages. Key components of your PID should be:

- your business case
- project goals, scope and size
- project organisation (defining the 'who, why, what, when and how' of the project)
- project constraints
- project risks

- stakeholders
- project controls and reporting framework
- the criteria for closing and assessing the project

Project definition and planning

Project planning is key to successful project management. This stage typically begins with **setting goals**. The two most common approaches include:

- the **SMART method** (specific, measurable, attainable, realistic and timely)
- the **CLEAR method** (collaborative, limited, emotional, appreciable, refinable)

At this stage, you will also define the **project scope**, and develop a **project plan** and **work breakdown schedule**. This involves identifying:

- time, cost and resources that are at your disposal
- roles and responsibilities for the project
- quality
- milestones
- baseline performance measures
- progress checkpoints
- risk and resources for resolving unforeseen issues

During this stage, you may also want to develop a communication plan (especially if you have external stakeholders), as well as a risk management plan.

Project launch and implementation

Implementation (also called **project execution**) simply means putting your project plan into action. It often begins with a project 'kick-off meeting'.

During this phase, you will carry out the tasks and activities from your project plan to produce the **project deliverables**. For example, if you are creating a promotional pack for a trade show, early deliverables might be to gather product information and prices, and complete all of your product photography and get it signed off by the customer.

Project managers may direct this work by:

- overseeing a team
- managing budget and resources
- communicating to stakeholders

Careful monitoring and control at this stage can help you keep the project plan on track. You can use a range of tools and processes to help you manage things like time, cost, quality and risks, or to communicate progress and manage customer acceptance.

Project monitoring and control

Monitoring and control often overlap with execution as they often occur at the same time. They require measuring **project progression and performance**, and dealing with any issues that arise from day-to-day work.

You can use **key performance indicators (KPIs)** to determine if your project is on track. Things you could measure include, for example:

- if your project is on schedule and budget
- if specific tasks are being completed
- if issues are adequately addressed

During this time, you may need to adjust schedules and resources to ensure that your project remains on track.

Project close

During this last phase, you will complete your work and dissolve the project. Closure doesn't necessarily mean success, but simply the final point of the project - eg closure can happen when you cancel projects that fail.

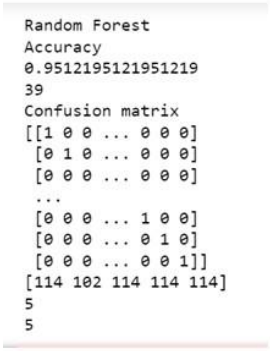
Project closure often involves things like:

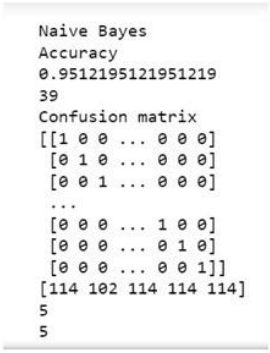
- handing over the deliverables
- releasing staff and resources
- archiving or handing over any relevant project documents
- cancelling supplier contracts
- completion of all activities across the project
- preparing the final project budget and report
- handover into business as usual if this applies

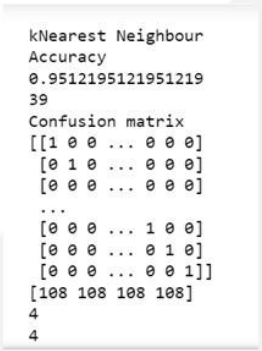
After closure, you can carry out a **post-implementation project review** (sometimes referred to as a 'post mortem' meeting). This is an opportunity to evaluate what went well and what didn't. Understanding failures, if there were any, can help you learn lessons and improve the way you carry out future projects.

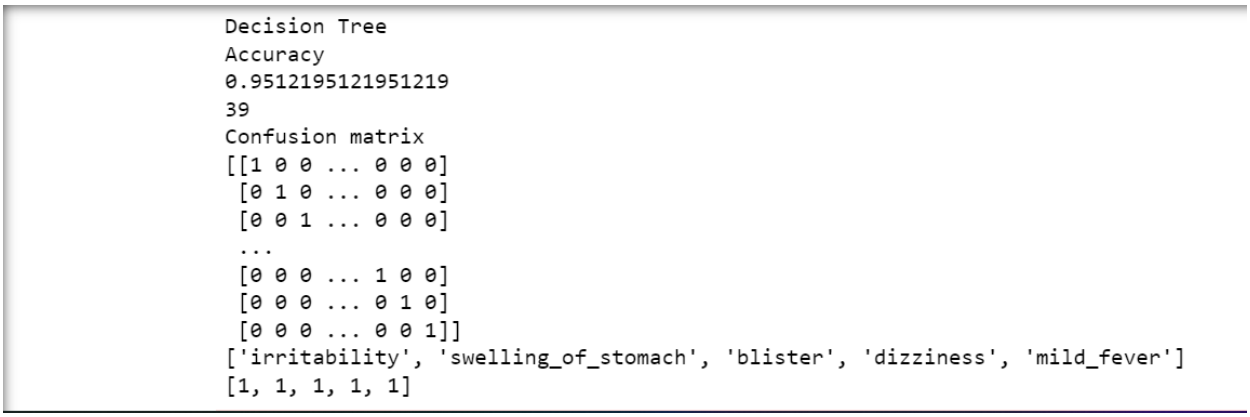
Testing Of our Project :

Testing of project yeilds the following Results:









Conclusion and future work

When it comes to disease diagnosis, accuracy is critical for planning, effective treatment, and ensuring the well-being of patients. AI is a vast and diverse realm of data, algorithms, analytics, deep learning, neural networks, and insights that is constantly expanding and adapting to the needs of the healthcare industry and its patients. According to the findings of this study, AI approaches in the healthcare system,particularly for illness detection, are essential. Aiming at illuminating how machine and deep learning techniques work in various disease diagnosis areas, the current study has been divided into several sections that cover the diagnosis of alzheimer's, cancer, diabetes, chronic diseases, heart disease, stroke and cerebrovascular disease, hypertension, skin disease, and liver disease. The introduction and contribution were covered in the first section, followed by an evaluation of the quality of the work and an examination of AI approaches and applications. Later, various illness symptoms and diagnostic difficulties, a paradigm for AI in disease detection models, and various AI applications in healthcare were discussed. The reported work on multiple diseases and the comparative analysis of different techniques with the used dataset as well as the results of an applied machine and deep learning methods in terms of multiple parameters such as accuracy, sensitivity, specificity, an area under the curve, and F- score has also been portrayed. Finally, the work that assisted researchers in determining the most effective method for detecting illnesses is finished, as in future scope. In a nutshell, medical experts better understand how AI may be used for illness diagnosis, leading to more appropriate proposals for the future development of AI based techniques.

Contrary to considerable advancements over the past several years, the area of accurate clinical diagnostics faces numerous obstacles that must be resolved and improved constantly to treat emerging illnesses and diseases effectively. Even healthcare professionals recognize the barriers that must be overcome before sickness may be detected in conjunction with artificial intelligence. Even doctors do not entirely rely on AI-based approaches at this time since they are unclear of their ability to anticipate illnesses and associated symptoms.Thus much work is required to train the AI-based systems so that there will be an increase in the accuracy to predict the methods for diagnosing diseases. Hence, in the future, AI-based research should be conducted by keeping the flaw mentioned earlier in

consideration to provide a mutually beneficial relationship between AI and clinicians. In addition to this, a decentralized federated learning model should also be applied to create a single training model for disease datasets at remote places for the early diagnosis of diseases.

The future of AI in health care could include tasks that range from simple to complex—everything from answering the phone to medical record review, population health trending and analytics, therapeutic drug and device design, reading radiology images, making clinical diagnoses and treatment plans, and even talking with patients.

The future of artificial intelligence in health care presents:

- A health care-oriented overview of artificial intelligence (AI), natural language processing (NLP), and machine learning (ML)
- Current and future applications in health care and the impact on patients, clinicians, and the pharmaceutical industry
- A look at how the future of AI in health care might unfold as these technologies impact the practice of medicine and health care over the next decade

From patient self-service to chat bots, computer-aided detection (CAD) systems for diagnosis, and image data analysis to identify candidate molecules in drug discovery, AI is already at work increasing convenience and efficiency, reducing costs and errors, and generally making it easier for more patients to receive the health care they need.

While NLP and ML are already being used in health care, they will become increasingly important for their potential to:

1. Improve provider and clinician productivity and quality of care
2. Enhance patient engagement in their own care and streamline patient access to care
3. Accelerate the speed and reduce the cost to develop new pharmaceutical treatments
4. Personalize medical treatments by leveraging analytics to mine significant, previously untapped stores of non-codified clinical data

While each AI technology can contribute significant value alone, the larger potential lies in the synergies generated by using them together across the entire patient journey, from diagnoses, to treatment, to ongoing health maintenance.

We set out to create a system which can predict disease on the basis of symptoms given to it. Such a system can decrease the rush at OPDs of hospitals and reduce the workload on medical staff. We were successful in creating such a system and use 4 different algorithm to do so. On an average we achieved accuracy of ~94%. Such a system can be largely reliable to do the job. Creating this system we also added a way to store the data entered by the user in the database which can be used in future to help in creating better version of such system. Our system also has an easy to use interface. It also has various visual representation of data collected and results achieved.

WHAT MORE CAN BE ADDED IN FUTURE?

1)HealthCare Chatbot :

Computer programs using textual conversational mediums are growing popular among healthcare institutions/organizations. Based on the market intelligence report published by BIS Research titled Global Chatbots in Healthcare Market – Analysis and Forecast, 2019-2029, the chatbots in the healthcare market generated a revenue of $36.5 million in 2018. These intelligent programs are able to detect symptoms, manage medications, and assist chronic health issues. They guide people rightly for serious illness and also assists them in scheduling appointments with professionals. With technological advancements in chatbots over the decade, there has been significant growth in the healthcare sector, along with other AI tools.

World renowned healthcare companies like Pfizer, the UK NHS, Mayo Clinic, and others are all using Healthcare Chatbots to easily meet the demands of their patients.

Here are some of the main benefits of chatbots in healthcare:

Availability around the clock
Medical emergencies could arise at any time. From identifying symptoms to scheduling surgeries, a patient might need anything at any time. All medical professionals might not be able to tend to everything at all times. Here's where chatbots come into play. They are available for everyone, at all times.

Provides critical information instantly
Time is of the essence in Healthcare. Chatbots provide helpful information instantly, especially in times where every second is important. For example, if a patient rushes in with an attack, the doctor can get the patient's information like previous records, other diseases, allergies, check-ups, etc., instantly over a bot.

Builds a rapport with patients and provides assistance
Imagine your patient is looking up a symptom on your website but doesn't know where exactly to learn about the solution and how to book an appointment with you? They're going to leave your website disappointed.

But a healthcare chatbot can help you turn that around. When a patient visits your website, your chatbot will give them a warm greeting and help them through the symptoms, predict possible diagnoses, and provide them with the option of booking an appointment with you directly.

They also take in user information by asking them questions, which gets stored for any form of reference and to personalize the patient's experience. This is how chatbots build a good rapport with your patients.

Empathy is a very important factor in healthcare. A website on its own might not be able to address all queries, an intuitive chatbot tends to solve more doubts and adds a personal touch. They also take in user information by asking them questions, which gets stored for any form of reference and to personalize the patient's experience. This is how chatbots build a good rapport with your patients.

Here are some of the main uses of chatbots in the healthcare industry:

Schedule appointments

Patients can book appointments directly from your chatbot. Not just that, the bots can even assign a doctor to your patient, send the doctor an email with patient information and create a slot in both the patient, as well as the doctor's calendar. This helps people to chedule online appointments with minimum effort and stress!

Checking for symptoms

Patients can easily check for symptoms on a healthcare chatbot and measure the severity of the situation. Since the chatbot remembers individual patient details, patients don't have to enter the same information every time they want to get an update.

Providing support and relevant information

When patients keep calling with the same set of basic questions, it could be helpful to have an intelligent chatbot answer them. Hence, healthcare chatbots are also given the responsibility of managing such additional queries and reducing repetitive calls.

Additionally, chatbots can even be useful to healthcare providers when they need to access patient information.

Coverage and claims

Whether patients want to check their existing coverage, file for claims or track the status of a claim, a healthcare chatbot will provide them with an easy gateway to find relevant information.

Similarly, since doctors will have easy access to patient information as well as queries, it is convenient for them to pre-authorize billing payments and other requests coming from patients or healthcare authorities.

Enhancing patient experience

Imagine how amazing it would be to have a doctor who can be there to support you at your beck-and-call. Unfortunately, that can't be the reality, but a bot can help. Healthcare chatbots typically have an uptime of over 99.9% so whenever your patients need an answer, a chatbot is there to give it to them.

Helping with therapy

Therapy chatbots can help provide mental health assistance. Chatbots for mental health are not intended to replace licensed mental health professionals, but rather to help them. Conversational chatbots can even practice cognitive behavioral therapy to some extent.

HOW THIS PROJECT IS BENEFICIAL ?

PROVIDES REAL-TIME DATA

A critical component of diagnosing and addressing medical issues is acquiring accurate information in a timely manner. With AI, doctors and other medical professionals can leverage immediate and precise data to expedite and optimize critical clinical decision-making. Generating more rapid and realistic results can lead to improved preventative steps, cost-savings and patient wait times.

STREAMLINES TASKS

Artificial intelligence in medicine has already changed healthcare practices everywhere. Innovations include appointment-scheduling, translating clinical details and tracking patient histories. AI is enabling healthcare facilities to streamline more tedious and meticulous tasks. For example, intelligent radiology technology is able to identify significant visual markers, saving hours of intense analysis. Other automated systems exist to automate appointment scheduling, patient tracking and care recommendations.

One specific task that is streamlined with AI **is reviewing insurance**. AI is used to minimize costs resulting from insurance claim denials. With AI, health providers can identify and address mistaken claims before insurance companies deny payment for them. Not only does this streamline the claims process, AI saves hospital staff the time to work through the denial and resubmit the claim.

Enabling faster payments and greater claims accuracy, hospitals can be more confident about reimbursement time frames, making them more willing to accept a larger number of insurance plans. AI essentially allows hospitals to accept a wide array of plans, benefiting potential and existing patients.

SAVES TIME AND RESOURCES

As more vital processes are automated, medical professionals have more time to assess patients and diagnose illness and ailment. AI is accelerating operations to save medical establishments precious productivity hours. In any sector, time equals money, so AI has the potential to save hefty costs.

It’s estimated around $200 billion is wasted in the healthcare industry annually. A good portion of these unnecessary costs are attributed to administrative strains, such as filing, reviewing and resolving accounts. Another area for improvement is

in medical necessity determination. Hours of reviewing patient history and information are traditionally needed to properly assess medical necessity. **New natural language processing (NLP) and deep learning (DL)** algorithms can assist physicians in reviewing hospital cases and avoiding denials.
By freeing vital productivity hours and resources, medical professionals are allotted more time to assist and interface with patients.

ASSISTS RESEARCH

AI enables researchers to amass large swaths of data from various sources. The ability to draw upon a rich and growing information body allows for more effective analysis of deadly diseases. Related to real-time data, research can benefit from the wide body of information available, as long as it's easily translated.
Medical research bodies like the **Childhood Cancer Data Lab** are developing useful software for medical practitioners to better navigate wide collections of data. AI has also been used to assess and detect symptoms earlier in an illness's progression. **Telehealth solutions** are being implemented to track patient progress, recover vital diagnosis data and contribute population information to shared networks.

MAY REDUCE PHYSICIAN STRESS

Some **latest research** reports over half of primary physicians feel stressed from deadline pressures and other workplace conditions. AI helps streamline procedures, automate functions, instantly share data and organize operations, all of which help relieve medical professionals of juggling too many tasks.

Yang explains, "The most significant contributor to physician burn out is patient load and the nature of the profession. However, as AI can assist with more time-intensive operations, explaining diagnoses for example, medical professionals may experience some stress alleviation."

COMING TRENDS IN AI :

The Ongoing Democratization of AI

AI will only achieve its full potential if it's available to everyone and every company and organization is able to benefit. Thankfully in 2023, this will be

easier than ever. An ever-growing number of apps put AI functionality at the fingers of anyone, regardless of their level of technical skill. This can be as simple as predictive text suggestions reducing the amount of typing needed to search or write emails to apps that enable us to create sophisticated visualizations and reports with a click of a mouse.

If there isn't an app that does what you need, then it's increasingly simple to create your own, even if you don't know how to code, thanks to the growing number of no-code and low-code platforms. These enable just about anyone to create, test and deploy AI-powered solutions using simple drag-and-drop or wizard-based interfaces. Examples include SwayAI, used to develop enterprise AI applications, and Akkio, which can create prediction and decision-making tools.

Ultimately, the democratization of AI will enable businesses and organizations to overcome the challenges posed by the AI skills gap created by the shortage of skilled and trained data scientists and AI software engineers. By empowering anybody to become "armchair" data scientists and engineers, the power and utility of AI will become within reach for us all.

Generative AI

If you ask most people what they think AI is useful for, they will probably tell you that it's mainly for automating routine, repetitive tasks. While this is often true, a growing branch of the science is dedicated to building AI tools and applications that can mimic one of the most uniquely human of all skill sets – creativity.

Generative AI algorithms take existing data – video, images or sounds, or even computer code – and uses it to create entirely new content that's never existed in the non-digital world.

One of the most well-known generative AI models is GPT-3, developed by OpenAI and capable of creating text and prose close to being indistinguishable from that created by humans. A variant of GPT-3 known as DALL-E is used to create images.

The technology has achieved mainstream exposure thanks to experiments such as the famous deepfaked Tom Cruise videos and the Metaphysic act, which took America's Got Talent by storm this year. But in 2023, we will see it used increasingly frequently to create synthetic data that can be used by businesses for all manner of purposes. Synthetic audio and video data can remove the need to

capture film and speech on video – simply type what you want the audience to see and hear into your generative tools, and the AI creates it for you!

Ethical and Explainable AI

The development of more ethical and explainable AI models is essential for a number of reasons. Most pressingly, though, it comes down to trust. AI requires data in order to learn, and often this means personal data. For many of the potentially most useful and powerful AI use cases, this might be very sensitive data like health or financial information. If we, the general public, don't trust AI or understand how it makes decisions, we simply won't feel safe handing over our information, and the whole thing falls apart.

In 2023 there will be efforts to overcome the "black box" problem of AI. Those responsible for putting AI systems in place will work harder to ensure that they are able to explain how decisions are made and what information was used to arrive at them. The role of AI ethics will become increasingly prominent, too, as organizations get to grips with eliminating bias and unfairness from their automated decision-making systems. Biased data has been shown to lead to prejudice in automated outcomes that can potentially lead to discrimination and unfair treatment – which simply won't be acceptable in a world where AI plays a part in decisions involving employment and access to justice or healthcare.

Augmented Working

In 2023, more of us will find ourselves working alongside robots and smart machines specifically designed to help us do our jobs better and more efficiently. This could take the form of smart handsets giving us instant access to data and analytics capabilities – as we have seen increasingly used in retail as well as industrial workplaces. It could mean augmented reality (AR)-enabled headsets that overlay digital information on the world around us. In a maintenance or manufacturing use case, this could give us real-time information that can help us identify hazards and risks to our own safety – such as pointing out when a wire is likely to be live or a component may be hot. Management and leadership teams will increasingly have access to real-time dashboards and reporting, giving an instant up-to-the-minute overview of operational effectiveness. AI-powered virtual assistants will also become more prevalent in the workplace, able to quickly answer questions as well as automatically suggest an alternative, more efficient methods of accomplishing objectives. Overall, developing the ability to work with and alongside intelligent, smart machines will become an increasingly indispensable work skill. I would even go as far as to say that for many of us, it

will go a long way towards mitigating the dangers of finding our roles becoming redundant!

Sustainable AI

In 2023 all companies will be under pressure to reduce their carbon footprint and minimize their impact on the environment. In this respect, the race to adopt and profit from AI can be both a blessing and a hindrance. AI algorithms – as well as all the infrastructure needed to support and deliver them, such as cloud networks and edge devices – require increasing amounts of power and resources. One study in 2019 found that training a single deep-learning model can result in the emission of 284,000 kilograms of CO2. At the same time, the technology has the potential to help companies understand how to build products, services, and infrastructure in a more energy-efficient way by identifying sources of waste and inefficiency. Ongoing efforts to implement more green and renewable energy-powered infrastructure are also a part of the drive toward delivering more sustainable AI.

AI can be a driver of sustainability in other industries and areas of operation, too – for example, computer vision is used in conjunction with satellite imagery to identify deforestation and illegal logging activity in the rainforests, as well as illegal fishing activity, which impacts biodiversity in the oceans. This year, I expect to see a continued drive towards deployment of AI initiatives aimed at tackling some of the most pressing problems facing our planet – rather than simply in pursuit of increased corporate profits.

Rapid Growth of Reinforcement Learning

Since AlphaGo developed by DeepMind defeated the Korean chess player Lee Sedol in Go in 2015, the proportion of reinforcement learning mentioned in artificial intelligence-related research papers has grown from 4.7% at that time to 20% after 2020. Now, reinforcement learning is also gradually creating huge value in various industries. Google's data centers use this technology to reduce energy consumption by more than 50%.

AI-Driven Business Decisions

Although the wisdom of AI is based on data, the so-called AI-driven and data-driven are actually very different. The former focuses on data, while the latter is the ability to process data. Now in 2020, AI is involved in more business decisions that would otherwise be the task of decision makers, ranging from operations,

marketing and sales, and even design. Artificial intelligence will gradually become the only link between data and business decisions.

RPA Penetration Increases

Process automation, also known as RPA (Robotic Process Automation), is the most frequent application of artificial intelligence. In a study of 152 AI use cases, it was found that nearly half of the cases in the industry are based on RPA. In recent years, due to the gradual maturity of the technology, the penetration of RPA will greatly increase in most industries, completing many of our existing tasks at a near-zero error and high-efficiency rate.

AI Will No Longer Be So Reliant on Big Data

In the past, training a deep learning model based on neural network often required a very large amount of data, but such data is not so easy to obtain in many fields such as medical treatment. This is why researchers often use certain data augmentation techniques, such as turning the same photo over, to increase the amount of existing data. However, with the increasing maturity of GAN technology, research in many fields can directly simulate new data, so that many meaningful models can be built in environments with only a small amount of data.

Ethical AI and AI Trustworthiness

Based on our many controversial developments in AI, such as simulating other people's voices and videos, or AI-driven surveillance systems, etc., as well as our fears about the potential of AI, how to humanely develop artificial intelligence technology is also gradually Gain momentum in academic research. Among them, developments such as explainable artificial intelligence and transparent AI decision-making are enhancing the trustworthiness of AI for users and consumers. At the same time, many policies and industrial norms are gradually echoing this trend.

More Relevant Interaction Models

The AI-driven interaction model Cognitive Engagement, often translated as cognitive investment, is driven by breakthroughs in NLP research and the maturity

of neural networks, and now has very complete applications in various fields. For example, a chatbot for 24-hour customer service, a product and service recommendation system that provides a personalized experience through communication, or an intelligent assistant that combines an expert system to work with professionals, AI will be used in many fields in the future. interact with users.

AI unlocks the metaverse

While the debate about the boundaries and regularities of the Metaverse continues, businessmen are already arranging and populating it. In fact, 2023 will be a defining period for many companies with regard to how they will create their world within the Metaverse.

The latest technology in artificial intelligence is one of the key elements of Metaverse. For example, AI allows us to overcome the limitations inherent in AR and VR, and create realistic 3D images and spaces. In addition, AI systems have a built-in potential for improvement due to self-controlled learning, which may not require human participation.

It can be said about any of the latest AI technologies that find their application in the Metaverse. For example, artificial intelligence avatars, or so-called digital humans, have emerged largely due to natural language processing (NLP) and computer vision. Virtual versions of people begin to understand not only language but also the movements and emotions of each other. In this way, users can transfer a part of their interaction with others to the avatar. At the same time, the digital copy can be given complete similarity to the original or other specific features.

The integration of AI systems and the IoT whose devices supply the necessary data, has made possible the creation of digital twins. Such virtual versions of environments or systems allow modeling their development under various scenarios. This makes it possible to predict their condition and behavior under the influence of different external factors. Therefore, the accumulation of data on digital duplicates of real objects and their aggregates, as well as the creation of management systems for such databases, is in demand. The challenge is to improve the quality of digitization, which, in turn, entails increasing the requirements for storing such huge volumes of data and for the productivity of their processing and analysis.

It is advisable for businesses to evaluate the following:

- 3D models of which objects and systems will be primarily useful for companies and their consumers.
- Processes involving clients and partners that can be painlessly transferred to the virtual world, and perhaps even made better there. Nowadays, no one is

surprised by virtual fitting or product testing, digital modeling of decor, and much more.

- Priorities, preferences, and expectations of your target audience in the Metaverse, which becomes a space for communication, sales, and other forms of interaction. It is also worth considering the evolution of Metaverse in the context of the decentralized concept of modern life, preparing to enable customers to independently store and share data, process important information on their AI devices, use decentralized applications (DApps), etc. At the same time, hardware and services should not lag behind the software.

  The current trend shows that appetites for AI-based products and services are only growing. According to the Worldwide Semiannual Artificial Intelligence Tracker provided by IDC (International Data Corporation), worldwide revenues for the artificial intelligence (AI) market in 2024 may be close to 500 billion US dollars. In turn, each of these AI-powered business applications can give impetus to developing other elements of the Metaverse.

AI ENHANCES SECURITY AND SURVEILLANCE

A new level of quality in security systems has also become possible thanks to the new artificial intelligence technologies. Video surveillance can now be combined with biometric authentication using face and voice recognition, and automated image analysis. AI-based security and monitoring systems lend themselves to more precise settings and more accurately identify objects that should be responded to when they appear. Video capture and analysis software helps secure large public and private spaces by detecting potential threats. Sensitive systems provide operators with timely information, not only about the identification of unwanted guests, but also about any behavioral anomalies or suspicious activity of visitors.

Identifying a person, including their age, gender, and emotional state, through voice recognition has also become an important feature of AI-powered applications. A built-in anti-spoofing feature that detects synthesized and recorded voice is just what is needed to keep such tools safe. Biometric facial recognition is also essential for maintaining security.

AI IN REAL-TIME VIDEO PROCESSING

For efficient processing of real-time video streams, it is critical to achieve data transmission accuracy and minimize video processing latency. AI solutions are involved in the root element – data pipeline processing.

The real-time video processing system using recent developments in artificial intelligence involves close integration of a pre-trained neural network model, user scenario implementation algorithms, and cloud infrastructure. It is thanks to the integration of these elements that real-time streaming speed is achieved. Acceleration of video processing is possible in two ways – by improving algorithms and by parallelizing processes. Parallelization of processes can be done through file splitting or by applying a pipeline approach.

The pipeline architecture allows the use of an AI algorithm for real-time video processing, avoiding additional complications and maintaining model accuracy. That is why the pipeline architecture is the optimal choice for fast and high- quality video processing. Moreover, it allows for the use of additional effects for face recognition and blurring. You can delve deeper into this topic with our article on AI in real-time video processing

It is hard to imagine processing streams in real-time without the ability to apply background removal and blurring. The rapid development of video conferencing has led to a growing interest in tools for these effects. And this trend will continue to gain momentum as the global video conferencing market is expected to grow from USD 9.2 billion in 2021 to USD 22.5 billion by 2026, according to GlobeNewswire forecasts.

Background removal and blurring in real-time video are based on creating a model that separates the person in the frame from the background. This task relies on a neural network. For its operation, you can choose one of the existing models, such as BodyPix, MediaPipe, or PixelLib. Next, you must integrate the chosen model with the relevant framework and organize the optimal execution process using WebAssembly, WebGL, or WebGPU.

USER MANUAL :

Step by Step Procedure to run the Project:

Importing Libraries

```
In [1]: #Importing Libraries
from mpl_toolkits.mplot3d import Axes3D
from sklearn.preprocessing import StandardScaler
import matplotlib.pyplot as plt
from tkinter import *
import numpy as np
import pandas as pd
import os
```

List of Symptoms

```
In [3]: #List of the symptoms is listed here in list l1.

l1=['back_pain','constipation','abdominal_pain','diarrhoea','mild_fever','yellow_urine',
    'yellowing_of_eyes','acute_liver_failure','fluid_overload','swelling_of_stomach',
    'swelled_lymph_nodes','malaise','blurred_and_distorted_vision','phlegm','throat_irritation',
    'redness_of_eyes','sinus_pressure','runny_nose','congestion','chest_pain','weakness_in_limbs',
    'fast_heart_rate','pain_during_bowel_movements','pain_in_anal_region','bloody_stool',
    'irritation_in_anus','neck_pain','dizziness','cramps','bruising','obesity','swollen_legs',
    'swollen_blood_vessels','puffy_face_and_eyes','enlarged_thyroid','brittle_nails',
    'swollen_extremeties','excessive_hunger','extra_marital_contacts','drying_and_tingling_lips',
    'slurred_speech','knee_pain','hip_joint_pain','muscle_weakness','stiff_neck','swelling_joints',
    'movement_stiffness','spinning_movements','loss_of_balance','unsteadiness',
    'weakness_of_one_body_side','loss_of_smell','bladder_discomfort','foul_smell_of urine',
    'continuous_feel_of_urine','passage_of_gases','internal_itching','toxic_look_(typhos)',
    'depression','irritability','muscle_pain','altered_sensorium','red_spots_over_body','belly_pain',
    'abnormal_menstruation','dischromic _patches','watering_from_eyes','increased_appetite','polyuria','family_hist
    'rusty_sputum','lack_of_concentration','visual_disturbances','receiving_blood_transfusion',
    'receiving_unsterile_injections','coma','stomach_bleeding','distention_of_abdomen',
    'history_of_alcohol_consumption','fluid_overload','blood_in_sputum','prominent_veins_on_calf',
    'palpitations','painful_walking','pus_filled_pimples','blackheads','scurring','skin_peeling',
    'silver_like_dusting','small_dents_in_nails','inflammatory_nails','blister','red_sore_around_nose',
    'yellow_crust_ooze']
len(l1)
```

```python
#List of Diseases is listed in list disease.

disease=['Fungal infection', 'Allergy', 'GERD', 'Chronic cholestasis',
       'Drug Reaction', 'Peptic ulcer diseae', 'AIDS', 'Diabetes ',
       'Gastroenteritis', 'Bronchial Asthma', 'Hypertension ', 'Migraine',
       'Cervical spondylosis', 'Paralysis (brain hemorrhage)', 'Jaundice',
       'Malaria', 'Chicken pox', 'Dengue', 'Typhoid', 'hepatitis A',
       'Hepatitis B', 'Hepatitis C', 'Hepatitis D', 'Hepatitis E',
       'Alcoholic hepatitis', 'Tuberculosis', 'Common Cold', 'Pneumonia',
       'Dimorphic hemmorhoids(piles)', 'Heart attack', 'Varicose veins',
       'Hypothyroidism', 'Hyperthyroidism', 'Hypoglycemia',
       'Osteoarthristis', 'Arthritis',
       '(vertigo) Paroymsal  Positional Vertigo', 'Acne',
       'Urinary tract infection', 'Psoriasis', 'Impetigo']
```

```python
l2=[]
for i in range(0,len(l1)):
    l2.append(0)
print(l2)
```

```
[0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]
```

Reading The .csv files

```python
#Reading the training .csv file
df=pd.read_csv("training.csv")
DF= pd.read_csv('training.csv', index_col='prognosis')
#Replace the values in the imported file by pandas by the inbuilt function replace in pandas.

df.replace({'prognosis':{'Fungal infection':0,'Allergy':1,'GERD':2,'Chronic cholestasis':3,'Drug Reaction':4,
    'Peptic ulcer diseae':5,'AIDS':6,'Diabetes ':7,'Gastroenteritis':8,'Bronchial Asthma':9,'Hypertension ':10,
    'Migraine':11,'Cervical spondylosis':12,
    'Paralysis (brain hemorrhage)':13,'Jaundice':14,'Malaria':15,'Chicken pox':16,'Dengue':17,'Typhoid':18,'hepatitis A':19,
    'Hepatitis B':20,'Hepatitis C':21,'Hepatitis D':22,'Hepatitis E':23,'Alcoholic hepatitis':24,'Tuberculosis':25,
    'Common Cold':26,'Pneumonia':27,'Dimorphic hemmorhoids(piles)':28,'Heart attack':29,'Varicose veins':30,'Hypothyroidism':31,
    'Hyperthyroidism':32,'Hypoglycemia':33,'Osteoarthristis':34,'Arthritis':35,
    '(vertigo) Paroymsal  Positional Vertigo':36,'Acne':37,'Urinary tract infection':38,'Psoriasis':39,
    'Impetigo':40}},inplace=True)
#df.head()
DF.head(10)
```

prognosis	itching	skin_rash	nodal_skin_eruptions	continuous_sneezing	shivering	chills	joint_pain	stomach_pain	acidity	ulcers_on_tongue	...	pus_fill
Fungal infection	1	1	1	0	0	0	0	0	0	0	...	
Fungal infection	0	1	1	0	0	0	0	0	0	0	...	
Fungal infection	1	0	1	0	0	0	0	0	0	0	...	
Fungal infection	1	1	0	0	0	0	0	0	0	0	...	
Fungal infection	1	1	1	0	0	0	0	0	0	0	...	
Fungal infection	0	1	1	0	0	0	0	0	0	0	...	
Fungal infection	1	0	1	0	0	0	0	0	0	0	...	
Fungal infection	1	1	0	0	0	0	0	0	0	0	...	
Fungal infection	1	1	1	0	0	0	0	0	0	0	...	
Fungal infection	1	1	1	0	0	0	0	0	0	0	...	

10 rows × 132 columns

Visualizations

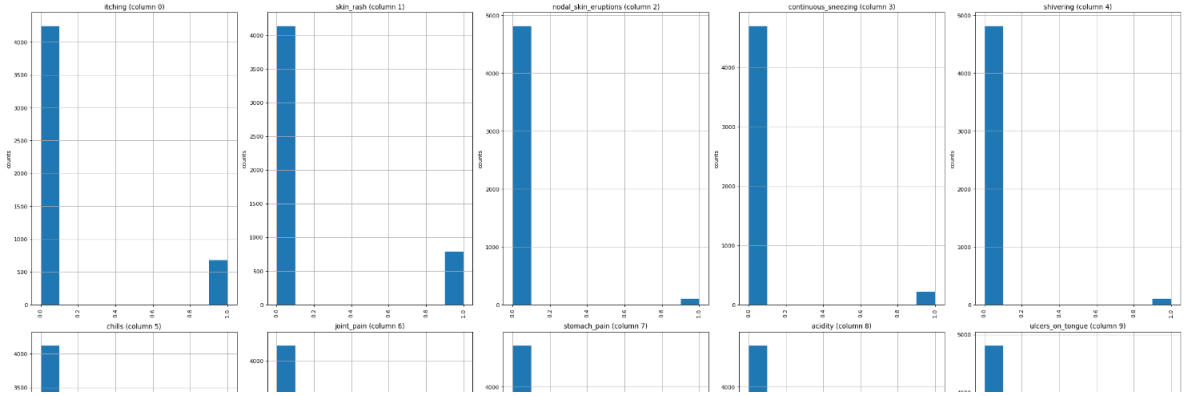
In [8]:
```
# Distribution graphs (histogram/bar graph) of column data
def plotPerColumnDistribution(df1, nGraphShown, nGraphPerRow):
    nunique = df1.nunique()
    df1 = df1[[col for col in df if nunique[col] > 1 and nunique[col] < 50]] # For displaying purposes, pick columns that have
    nRow, nCol = df1.shape
    columnNames = list(df1)
    nGraphRow = (nCol + nGraphPerRow - 1) / nGraphPerRow
    plt.figure(num = None, figsize = (6 * nGraphPerRow, 8 * nGraphRow), dpi = 80, facecolor = 'w', edgecolor = 'k')
    for i in range(min(nCol, nGraphShown)):
        plt.subplot(nGraphRow, nGraphPerRow, i + 1)
        columnDf = df.iloc[:, i]
        if (not np.issubdtype(type(columnDf.iloc[0]), np.number)):
            valueCounts = columnDf.value_counts()
            valueCounts.plot.bar()
        else:
            columnDf.hist()
        plt.ylabel('counts')
        plt.xticks(rotation = 90)
        plt.title(f'{columnNames[i]} (column {i})')
    plt.tight_layout(pad = 1.0, w_pad = 1.0, h_pad = 1.0)
    plt.show()
```

Scatter Plots to check Correlation among various symptoms

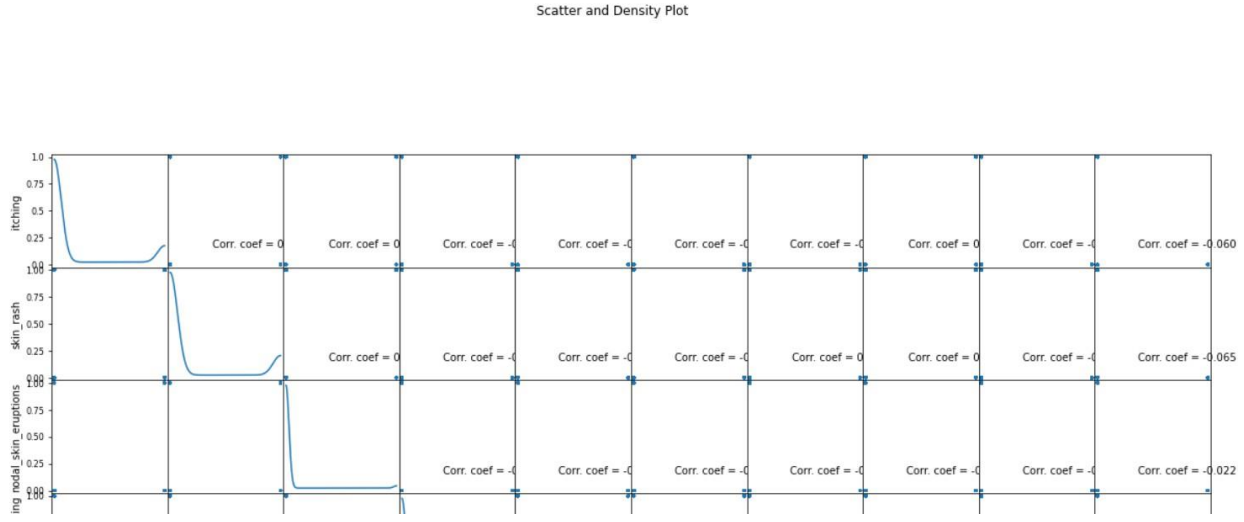
In [9]:
```
# Scatter and density plots
def plotScatterMatrix(df1, plotSize, textSize):
    df1 = df1.select_dtypes(include =[np.number]) # keep only numerical columns
    # Remove rows and columns that would lead to df being singular
    df1 = df1.dropna('columns')
    df1 = df1[[col for col in df if df[col].nunique() > 1]] # keep columns where there are more than 1 unique values
    columnNames = list(df)
    if len(columnNames) > 10: # reduce the number of columns for matrix inversion of kernel density plots
        columnNames = columnNames[:10]
    df1 = df1[columnNames]
    ax = pd.plotting.scatter_matrix(df1, alpha=0.75, figsize=[plotSize, plotSize], diagonal='kde')
    corrs = df1.corr().values
    for i, j in zip(*plt.np.triu_indices_from(ax, k = 1)):
        ax[i, j].annotate('Corr. coef = %.3f' % corrs[i, j], (0.8, 0.2), xycoords='axes fraction', ha='center', va='center', si
    plt.suptitle('Scatter and Density Plot')
    plt.show()
```

In [10]:
```
plotPerColumnDistribution(df, 10, 5)
```

```
<ipython-input-8-3de5d0542a9c>:10: MatplotlibDeprecationWarning: Passing non-integers as three-element position specification i
s deprecated since 3.3 and will be removed two minor releases later.
  plt.subplot(nGraphRow, nGraphPerRow, i + 1)
```

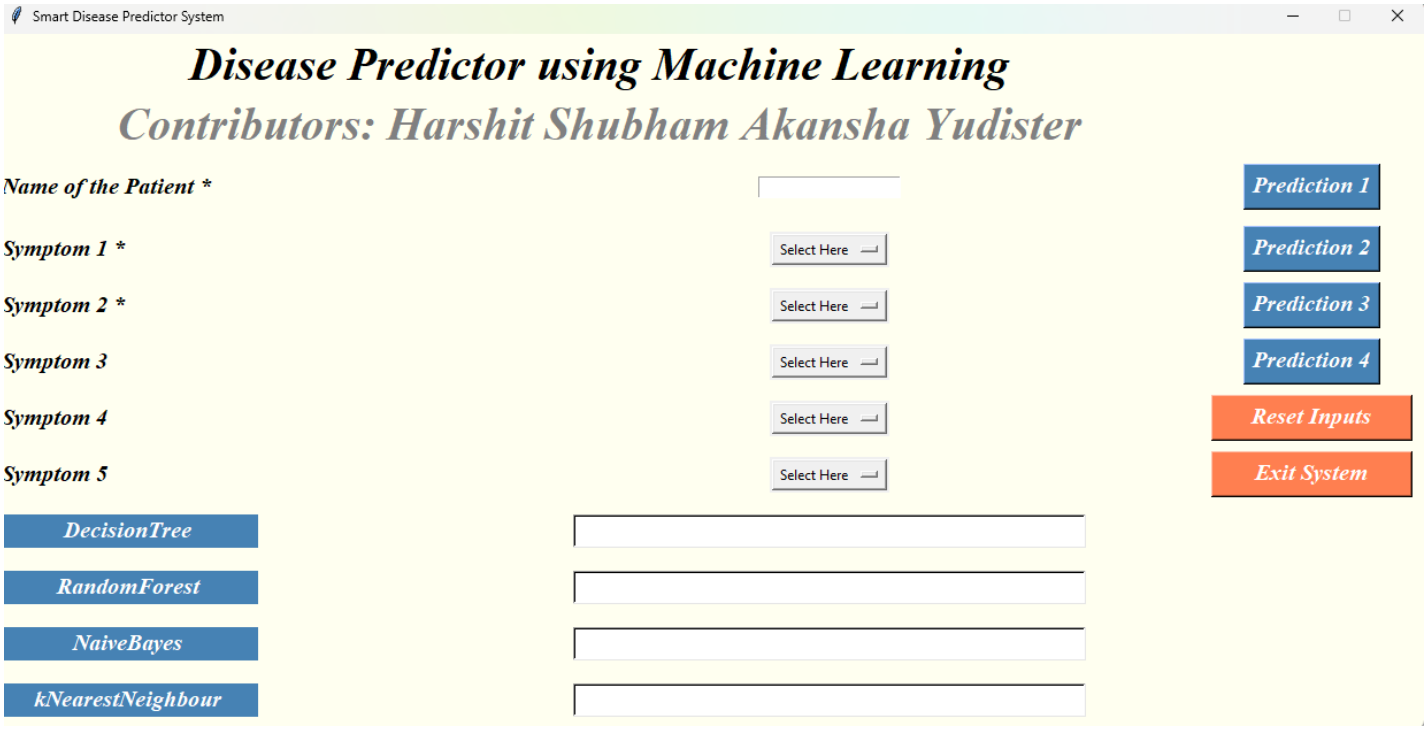


In [11]:
```
plotScatterMatrix(df, 20, 10)
```



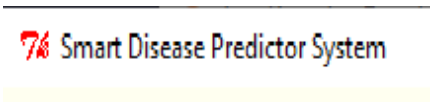
Testing Data :

```
#Reading the  testing.csv file
tr=pd.read_csv("testing.csv")

#Using inbuilt function replace in pandas for replacing the values

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

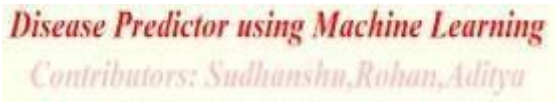


GUI made for this project is a simple tkinter GUI consisting of labels, messagebox, button, text, title and option menu

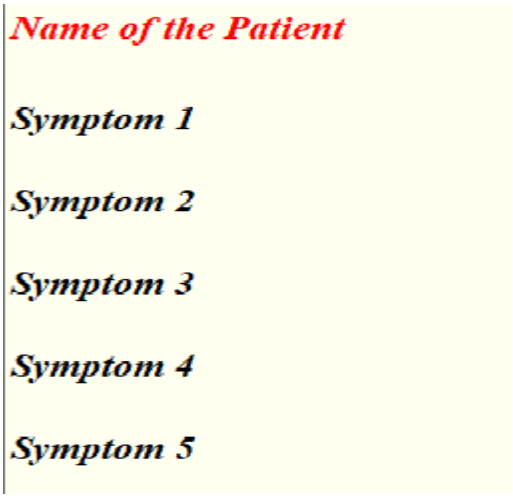
Root.title() is used to set the the title as Smart Disease Predictor System



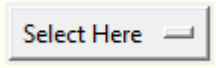
Label is used to add heading and contributors section.



Labels are further used for different sections



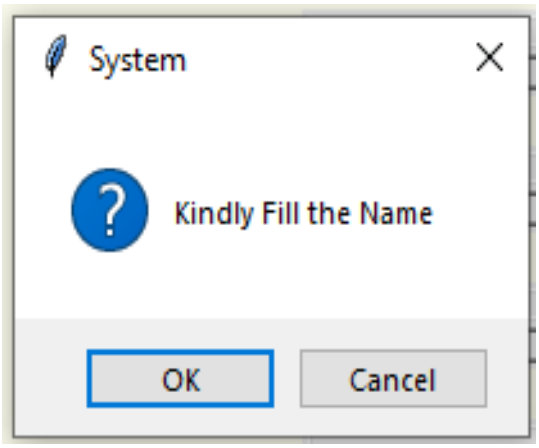
OptionMenu is used to create drop down menu



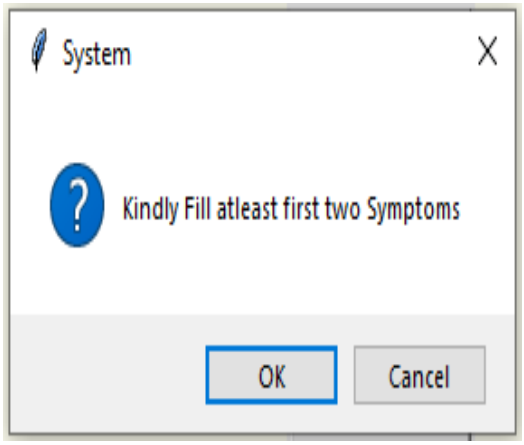
Buttons are used to give functionalities and predict the out come of models also two utilitybuttons namely exit and rest are also created.

Text is used to show output of the prediction using blank space.

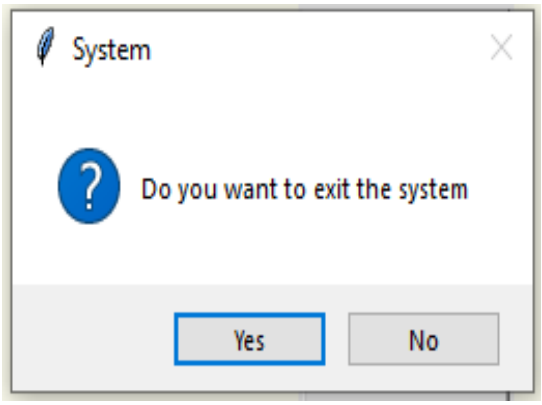
Messagebox are used at three different places, one- to restrain the to enter name



two- to ask for at least two symptoms,



three- to confirm to exit system.



Motivation Behind the Project :

Because of Trends Going on in AI:

Following the two climaxes of AI development in the 20th century, deep learning has set off the third wave of AI with its outstanding performance in automatic feature extraction. Since 2006, deep learning has made great breakthroughs in voice and vision recognition capabilities from its budding to maturity. A new technological breakthrough with a cycle of about 20 years has been gestating.

Four trends are becoming more and more obvious, according to the development process of AI technology.

- Accelerated integration of artificial intelligence and industry

  Generally speaking, the application of AI in industry is in its infancy, and there are still some difficulties that hinder the implementation of application scenarios. Therefore, AI must be closely integrated with the industry, not only to promote the implementation of AI application scenarios, but also to push forward the breakthrough innovations in basic data and platform technology, and to build a bridge that effectively connects with the traditional industry ecology.

- AI technology research: in-depth details

  At present, the new generation of AI has gradually shifted from the initial algorithm drive to the compound drive of data, algorithm, and computing power. Among them, the driving role of data-based applications has become increasingly prominent.

- A closed-loop without human intervention

  Today's AI is not an automatic and complete closed-loop system. After outputting wrong results, it cannot self-correct and immediately achieve better output. Instead, scientists need to retrain on better training datasets to achieve better results. The closed-loop human free intervention will be an important goal for the development of AI technology.

- Trusted AI

  To prevent artificial intelligence from being misused and abused, on the one hand, it is necessary to address the problematic symptoms from different levels such as laws and regulations, ethical norms, and industry consensus, but also to address the root cause from the level of technological innovation. Therefore, it is increasingly more important to embed ethics and governance in the entire life cycle of AI product design, research, development, deployment, and use.

Imaging Future of Artificial Intelligence

A digital transformation wave has swept over all walks of life around the world. Global enterprises have realized embracing emerging information technology is the key to improving business optimization, industrial upgrading, and value creation. In this process, the potential of AI, as a key enabling technology, is being uncovered in terms of computing power, algorithms, and data. Now, what are the successful landing cases? What new possibilities will it bring in the future?

It is undeniable that AI will continue to learn and advance with the times. What will happen in the future? There isn't a wink of a doubt that all aspects of our travel, medical care, and production will be further upgraded thanks to AI.

Perhaps dealing with robots will become our daily routine, perhaps brain-computer interfaces will come to fruition, helping people with disabilities to restore their lives and communication skills. Maybe AI will make people more creative, free humans from complicated or mindless tasks, and even replace humans in dangerous jobs. The technological development of AI will go hand in hand with the digitization and intelligent upgrading of the industry, building a future with unlimited possibilities.

Need of Futuristic AI Based HealthCare System:

Human Healthcare is one of the most significant subjects of society.The identification of the nature of the illness or other problems by examination of the symptoms i.e. diagnosis always stands first in overallcuring procedure of disease. Thus
, we can say that diagnosis and prediction of disease are most crucial aspects to be considered before thinking about exact procedure of curing the disease. This diagnosis procedure requires a lot of time as well as a lot of money. As a result , people belonging to poor financial background are not able to get accurate diagnose of disease which at the end may create life or death situation. So WeCare is basically a Human Disease Detection System which is supposed to input some data related like sugar levels etc. and in response to that it gives type of disease that person is having as an output. So , this System will be salutary to people who are not able to pay huge amounts to get diagnosed as well as the ones who requires immediate diagnosis.

Health information needs are also changing the information seeking behaviour and can beobserved around the globe. Challenges faced by many people are looking online for health information regarding diseases, diagnoses and different treatments. If a recommendation system can be made for doctors and medicine while using review mining will save a lot of time. In this type of system, the user face problem in

understanding the heterogeneous medical vocabulary as the users are laymen. User is confused because a large amount of medical information on different mediums are available.The idea behind recommender system is to adapt to cope with the special requirements of the health domain related with users.

REFERENCES

1. https://www.researchgate.net/publication/349054979_Human_Diseases_Detection_Based_On_Mach ine_Learning_Algorithms_A_Review
2. https://towardsdatascience.com/supervised-vs-unsupervised-learning-14f68e32ea8d Acharya, U. R., Fujita, H., Oh, S. L., Hagiwara, Y., Tan, J. H., & Adam, M. (2017). Application ofdeep convolutional neural network for automated detection of myocardial infarction using ECG signals. InformationSciences, 415–416, 190–198. https://doi.org/10.1016/j.ins.2017.06.027
3. Ahmed, S., Choi, K. Y., Lee, J. J., Kim, B. C., Kwon, G. R., Lee, K. H., & Jung, H. Y. (2019).Ensembles of Patch-Based
4. Classifiers for Diagnosis of Alzheimer Diseases. IEEE Access, 7, 73373–73383. 8. https://doi.org/10.1109/ACCESS.2019.2920011
5. Al-Zebari, A., & Sengur, A. (2019). Performance Comparison of Machine LearningTechniques on Diabetes
6. Disease Detection. 1st International Informatics and Software Engineering Conference:Innovative
7. Technologies for Digital Transformation, IISEC 2019 - Proceedings, 2–5. 12. https://doi.org/10.1109/UBMYK48245.2019.8965542
8. Bagga, P., & Hans, R. (2015). Applications of mobile agents in healthcare domain: A literature survey.
9. International Journal of Grid and Distributed Computing, 8(5), 55–72. 15. https://doi.org/10.14257/ijgdc.2015.8.5.05
10. Bargarai, F. A. M., Abdulazeez, A. M., Tiryaki, V. M., & Zeebaree, D. Q. (2020). Management of wireless
11. communication systems using artificial intelligence-based software defined radio. InternationalJournal of
12. Interactive Mobile Technologies, 14(13), 107–133. https://doi.org/10.3991/ijim.v14i13.14211
13. Chitra, K. and. (2018). Classification Of Diabetes Disease Using Support Vector Machine. 3(2), 1797–1801.
14. https://www.researchgate.net/publication/320395340
15. Cinarer, G., & Emiroglu, B. G. (2019). Classificatin of Brain Tumors by Machine Learning Algorithms. 3rd
16. International Symposium on Multidisciplinary Studies and Innovative Technologies, ISMSIT 2019 -
17. Proceedings. https://doi.org/10.1109/ISMSIT.2019.8932878 Daniels, M., & Schroeder, S. A. (1977). Variation among physicians in use of laboratory tests II. Relation to clinical productivity and outcomes of care. Medical Care, 15(6), 482–487. https://doi.org/10.1097/00005650-197706000-00004
18. Durai, V. (n.d.). Liver disease prediction using machine learning. 5(2), 1584–1588.

19. Fan, C. H., Hsu, Y., Yu, S. N., & Lin, J. W. (2013). Detection of myocardial ischemia episodeusing morphological
20. features. Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and
30. Biology Society, EMBS, 7334–7337. https://doi.org/10.1109/EMBC.2013.6611252
21. Grimson, J., Stephens, G., Jung, B., Grimson, W., Berry, D., & Pardon, S. (2001). Sharing healthcare records over the internet. IEEE Internet Computing, 5(3), 49–58. https://doi.org/10.1109/4236.935177
22. Hariharan, M., Polat, K., & Sindhu, R. (2014). A new hybrid intelligent system for accurate detection of
23. Parkinson's disease. Computer Methods and Programs in Biomedicine, 113(3), 904–913.35. https://doi.org/10.1016/j.cmpb.2014.01.004
24. Hashem, S., Esmat, G., Elakel, W., Habashy, S., Raouf, S. A., ElHefnawi, M., Eladawy, M., & ElHefnawi, M. (2018).
25. Comparison of Machine Learning Approaches for Prediction of Advanced Liver Fibrosis in Chronic
26. Hepatitis C Patients. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 15(3),861–868.
27. https://doi.org/10.1109/TCBB.2017.2690848
28. https://link.springer.com/article/10.1007/s12652-021-03612-z
29. https://www.tutorialspoint.com/python/python_gui_programming.htm
30. https://www.tutorialspoint.com/scikit_learn/scikit_learn_introduction.htm
31. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8950225/#:~:text=While%20the%20individual's%20ability%20restricts,health%20care%20may%20be%20developed.
32. https://www.analyticsinsight.net/top-10-ai-breakthroughs-in-healthcare-to-note-down/
33. https://www.brookings.edu/research/risks-and-remedies-for-artificial-intelligence-in-health-care/
34. https://www.google.com/search?q=project+management+techniques+and+tools&oq=Project+management+tecj&aqs=chrome.2.69i57j0i13i512l9.6959j0j7&sourceid=chrome&ie=UTF-8
35. https://www.nibusinessinfo.co.uk/content/five-stages-project-management
36. https://www2.deloitte.com/us/en/pages/life-sciences-and-health-care/articles/future-of-artificial-intelligence-in-health-care.html

Appendix 1 :

Jargons Used in ML :

Representation	Evaluation	Optimization
Instances	Accuracy/Error rate	Combinatorial optimization
K-nearest neighbor	Precision and recall	Greedy search
Support vector machines	Squared error	Beam search
Hyperplanes	Likelihood	Branch-and-bound
Naive Bayes	Posterior probability	Continuous optimization
Logistic regression	Information gain	Unconstrained
Decision trees	K-L divergence	Gradient descent
Sets of rules	Cost/Utility	Conjugate gradient
Propositional rules	Margin	Quasi-Newton methods
Logic programs		Constrained
Neural networks		Linear programming
Graphical models		Quadratic programming
Bayesian networks		
Conditional random fields		

Classification

Classification is a part of **supervised learning** (learning with labeled data) through which data inputs can be easily separated into categories. In machine learning, there can be **binary classifiers** with only two outcomes (e.g., spam, non-spam) or **multi-class classifiers** (e.g., types of books, animal species, etc.).

Clustering

Clustering is a form of **unsupervised learning** (learning with unlabeled data) that involves grouping data points according to features and attributes.

Clustering can be used to organize customer demographics and purchasing behavior into specific segments for targeting and product positioning. It can also analyze housing quality and geographic locations to create real estate valuations and plan the layout of new city developments. It can classify information by topics within libraries or web pages and compile an easily accessible directory for users.

Regressions

Regressions create relationships and correlations between different types of data. For example, each profile picture has an image with pixels that belong to a person. With **static prediction** (one that stays the same over time), machine learning acknowledges that a certain pixel arrangement corresponds to a given

name and allows for **facial recognition** (for example, when Facebook recommends tags for the photos you've just uploaded).

Regressions can also be useful when predicting outcomes based on data in the present. For a long time, statistical regression has been used to solve problems, such as predicting the recovery of cognitive functions after a stroke or predicting customer churn in the telecommunications industry. The only difference is that now many of these regression analyses can be done more efficiently and quickly by machines.

Deep Learning

Deep learning is similar to machine learning—in fact, it's more of an application of machine learning that imitates the workings of the human brain. Deep learning networks interpret **big data** (data that is too large to fit on a single computer)—both unstructured and structured—and recognize patterns. The more data they can "learn" from, the more informed and accurate their decisions will be. Here are some examples of deep learning in practice:

Neural Networks

Neural networks are closely related to deep learning. They create sequential layers of neurons that deepen the understanding of data collected from a machine to provide an accurate analysis.

A neural network consists of layers of **nodes**, which receive stimulation from "trigger" data. This data then is assigned a weight through coefficients, as some data inputs may be more significant than others.

Natural Language Processing

Natural language processing is the subfield of AI that processes human languages. It is a very important term in the field of data science and machine learning. The challenge is that often human speech is not literal. There are figures of speech, words, or phrasing specific to certain dialects and cultures, and sentences that can take on different meanings with grammar and punctuation. Similar to human conversations, natural language processors need to use the **syntax** (arrangement of words) and **semantics** (meaning of that arrangement) to come up with correct interpretations.

Machine Vision

Machine vision, or **computer vision**, is the process by which machines can capture and analyze images. This allows for the diagnosis of skin cancer by

looking at X-rays and other medical imagery, and for the detection of real-time traffic and vehicle types for self-driving cars, like Tesla's new models.

There are many different ways that machines can "see": representing colors numerically, decomposing images into different parts, and identifying corners, edges, and textures. As the machines gather and code more information, they begin to view the larger picture.

Machine Learning Engineer
With all these exciting technological advances, who is responsible for deploying ML within companies? In many cases, the responsibility first lies with the machine learning engineer, a data-driven software engineer focused on building the systems that can eventually learn and perform work autonomously. These engineers usually need to be familiar with different code bases, distributed computing, data wrangling, and computer science.

accuracy

The number of correct classification **predictions** divided by the total number of predictions. That is:

Accuracy=correct predictionscorrect predictions + incorrect predictions

For example, a model that made 40 correct predictions and 10 incorrect predictions would have an accuracy of:

Accuracy=4040 + 10=80%

Binary classification provides specific names for the different categories of *correct predictions* and *incorrect predictions*. So, the accuracy formula for binary classification is as follows:

Accuracy=TP+TNTP+TN+FP+FN

where:

- TP is the number of **true positives** (correct predictions).
- TN is the number of **true negatives** (correct predictions).
- FP is the number of **false positives** (incorrect predictions).
- FN is the number of **false negatives** (incorrect predictions).

Compare and contrast accuracy with **precision** and **recall**.

activation function
#fundamentals

A function that enables **neural networks** to learn **nonlinear** (complex) relationships between features and the label.

Popular activation functions include:

- **ReLU**
- **Sigmoid**

active learning

A **training** approach in which the algorithm *chooses* some of the data it learns from. Active learning is particularly valuable when **labeled examples** are scarce or expensive to obtain. Instead of blindly seeking a diverse range of labeled examples, an active learning algorithm selectively seeks the particular range of examples it needs for learning.

AdaGrad

A sophisticated gradient descent algorithm that rescales the gradients of each **parameter**, effectively giving each parameter an independent **learning rate**.

anomaly detection

The process of identifying **outliers**. For example, if the mean for a certain **feature** is 100 with a standard deviation of 10, then anomaly detection should flag a value of 200 as suspicious.

artificial intelligence

A non-human program or **model** that can solve sophisticated tasks. For example, a program or model that translates text or a program or model that identifies diseases from radiologic images both exhibit artificial intelligence.

Formally, **machine learning** is a sub-field of artificial intelligence. However, in recent years, some organizations have begun using the terms *artificial intelligence* and *machine learning* interchangeably.

attention

Any of a wide range of **neural network** architecture mechanisms that aggregate information from a set of inputs in a data-dependent manner. A typical attention mechanism might consist of a weighted sum over a set of inputs, where the **weight** for each input is computed by another part of the neural network.

attribute

Synonym for **feature**.

In machine learning fairness, attributes often refer to characteristics pertaining to individuals.

attribute sampling

A tactic for training a **decision forest** in which each **decision tree** considers only a random subset of possible **features** when learning the **condition**. Generally, a different subset of features is sampled for each **node**. In contrast, when training a decision tree without attribute sampling, all possible features are considered for each node.

Appendix 2 :

List of jargons in Medical:

- Benign: Not cancerous
- Malignant: Cancerous
- Anti-inflammatory: Reduces swelling, pain, and soreness (such as ibuprofen or naproxen)

- Body Mass Index (BMI): Body fat measurement based on height and weight
- Biopsy: A tissue sample for testing purposes
- Hypotension: Low blood pressure
- Hypertension: High blood pressure
- Lesion: Wound, sore, or cut
- Noninvasive: Doesn't require entering the body with instruments; usually simple
- Outpatient: Check in and check out the same day
- Inpatient: Plan to stay overnight for one or more days
- In remission: Disease is not getting worse; not to be confused with being cured
- Membrane: Thin layer of pliable tissue that serves as a covering or lining or connection between two structures
- Acute: Sudden but usually short (e.g., acute illness)
- Angina: Pain in the chest related to the heart that comes and goes
- Gastroesophageal Reflux Disease (GERD): Heartburn
- Cellulitis: Inflamed or infected tissue beneath the skin
- Epidermis: Outermost layer of skin
- Neutrophils: Most common type of white blood cell
- Edema: Swelling
- Embolism: Blood clot
- Sutures: Stitches
- Polyp: Mass or growth of thin tissue
- Compound fracture: Broken bone that protrudes through the skin
- Comminuted fracture: Broken bone that shatters into many pieces

Appendix 3:

List of symptoms in Medical and used in Project:

- back_pain
- constipation
- abdominal_pain
- diarrhoea
- mild_fever
- yellow_urine
- yellowing_of_eyes
- acute_liver_failure

- fluid_overload
- swelling_of_stomach
- swelled_lymph_nodes
- malaise
- blurred_and_distorted_vision
- phlegm
- throat_irritation
- redness_of_eyes
- sinus_pressure
- runny_nose
- congestion
- chest_pain
- weakness_in_limbs
- fast_heart_rate
- pain_during_bowel_movements
- pain_in_anal_region
- bloody_stool
- irritation_in_anus
- neck_pain
- dizziness
- cramps
- bruising
- obesity
- swollen_legs
- swollen_blood_vessels
- puffy_face_and_eyes

- enlarged_thyroid
- brittle_nails
- swollen_extremeties
- excessive_hunger
- extra_marital_contacts
- drying_and_tingling_lips
- slurred_speech
- knee_pain
- hip_joint_pain
- muscle_weakness
- stiff_neck
- swelling_joints
- movement_stiffness
- spinning_movements
- loss_of_balance
- unsteadiness
- weakness_of_one_body_side
- loss_of_smell
- bladder_discomfort
- foul_smell_of urine
- continuous_feel_of_urine
- passage_of_gases
- internal_itching
- toxic_look_(typhos)
- depression
- irritability

- muscle_pain
- altered_sensorium
- red_spots_over_body
- belly_pain
- abnormal_menstruation
- dischromic _patches
- watering_from_eyes
- increased_appetite
- polyuria
- family_history
- mucoid_sputum
- rusty_sputum
- lack_of_concentration
- visual_disturbances
- receiving_blood_transfusion
- receiving_unsterile_injections
- coma
- stomach_bleeding
- distention_of_abdomen
- history_of_alcohol_consumption
- fluid_overload
- blood_in_sputum
- prominent_veins_on_calf
- palpitations
- painful_walking
- pus_filled_pimples

- blackheads
- scurring
- skin_peeling
- silver_like_dusting
- small_dents_in_nails
- inflammatory_nails
- blister
- red_sore_around_nose
- yellow_crust_ooz

Appendix 4:

List of Diseases in Medical and used in Project:

- Fungal infection
- Allergy
- GERD
- Chronic cholestasis
- Drug Reaction
- Peptic ulcer diseae
- AIDS
- Diabetes
- Gastroenteritis
- Bronchial Asthma
- Hypertension
- Migraine
- Cervical spondylosis
- Paralysis (brain hemorrhage)
- Jaundice
- Malaria
- Chicken pox
- Dengue
- Typhoid
- hepatitis A
- Hepatitis B

- Hepatitis C
- Hepatitis D
- Hepatitis E
- Alcoholic hepatitis
- Tuberculosis
- Common Cold
- Pneumonia
- Dimorphic hemmorhoids(piles)
- Heart attack
- Varicose veins
- Hypothyroidism
- Hyperthyroidism
- Hypoglycemia
- Osteoarthristis
- Arthritis
- (vertigo) Paroymsal Positional Vertigo
- Acne
- Urinary tract infection
- Psoriasis
- Impetigo