IJRAR.ORG E-ISSN: 2348-1269, P-ISSN: 2349-5138

INTERNATIONAL JOURNAL OF RESEARCH AND ANALYTICAL REVIEWS (IJRAR) | IJRAR.ORG

An International Open Access, Peer-reviewed, Refereed Journal

DISEASE PREDICTION USING SYMPTOMS BASED ON MACHINE LEARNING ALGORITHMS

S. NITHYA[1], GAJENDRAN.K[2], ISHWARAYA PRASATH.V[3], KARTHIKEYAN.B[4], SURYA.M[5]
1 ASSISTANT PROFESSOR 2 UG SCHOLAR
BIOMEDICAL ENGINEERING
VSB ENGINEERING COLLEGE, KARUR, TAMILNADU, INDIA

ABSTRACT

Disease prediction is a critical area of research in healthcare, as early identification of diseases can lead to better outcomes and improved quality of life for patients. With the increasing availability of health data and advancements in machine learning algorithms, there has been growing interest in using these tools to predict diseases based on symptoms. In this study, we explore the use of machine learning algorithms to predict diseases based on symptoms reported by patients. We collected data from electronic health records and other medical sources and used feature engineering techniques to extract relevant features from the data. We evaluated several machine learning algorithms, including decision trees, random forests, support vector machines, and neural networks, to identify the best-performing model for disease prediction. We used performance metrics such as accuracy, precision, recall, and F1 score to evaluate the models' performance. People are currently suffering from a variety of diseases. Many people are unsure if the symptoms they are experiencing are indicative of a certain disease, and hence they are unable to take the required safeguards. Anticipating the disease during prodromal stage lowers the likelihood of complications. People will not be able to visit a doctor every time they experience a symptom. It may sometimes become a serious ailment if not treated. A model is suggested that uses a variety of symptoms as input to predict the illness. For disease prediction, the suggested method utilizes Decision trees, Naive Bayes, and Random forest classifiers. The ultimate result will be the mode of all these machine learning models. Users will be given a graphical user interface (GUI) to choose their symptoms. The final result will be shown on the interface using all three machine learning techniques, and feature extraction will be done depending on their symptoms.

Keywords: Neural networks, Tuberculosis, Naive Bayes, and Random forest

1. INTRODUCTION

Disease refers to an abnormal condition or disorder that affects the body or mind of an organism, resulting in impaired functioning of the affected system or organ. Diseases can be caused by a variety of factors, including infectious agents such as bacteria, viruses, and parasites, genetic mutations, environmental factors, and lifestyle choices.

Some common examples of diseases include:

Infectious diseases: These are diseases caused by microorganisms such as bacteria, viruses, fungi, and parasites, and can spread from one person to another. Examples include the flu, HIV/AIDS, tuberculosis, malaria, and COVID-19. Chronic diseases: These are long-term diseases that often develop over time and can be caused by a combination of genetic, environmental, and lifestyle factors. Examples include heart disease, diabetes, cancer, and chronic obstructive pulmonary disease (COPD).

Autoimmune diseases: These are diseases where the immune system attacks healthy cells and tissues in the body, leading to inflammation and damage. Examples include rheumatoid arthritis, lupus, and multiple sclerosis.
Mental health disorders: These are disorders that affect mental and emotional health, including depression, anxiety disorders, bipolar disorder, and schizophrenia.
Genetic disorders: These are disorders caused by mutations or abnormalities in the genes, leading to a range of health problems such as cystic fibrosis, sickle cell anemia and Huntington's disease
Lifestyle diseases: These are diseases that are often caused by unhealthy lifestyle choices such as poor diet, lack of exercise, smoking, and excessive alcohol consumption. Examples include obesity, type 2 diabetes, and hypertension. These are just a few examples of the many different types of diseases that can affect humans. Disease prevention and control measures can help to reduce the incidence and impact of many of these diseases. Diseases can affect any part of the body and can manifest in a range of symptoms, from mild to severe. Some diseases may be acute, meaning they develop rapidly and resolve within a short period, while others may be chronic, lasting for months or years. Effective disease management and prevention measures are critical to maintaining good health and quality of life. These can include vaccinations, medications, lifestyle changes, and environmental interventions. Early detection and treatment of diseases are also important in preventing complications and reducing the risk of long-term disability or death.

There are many different types of diseases that can affect humans, including infectious diseases, chronic diseases, autoimmune diseases, genetic disorders, mental health disorders, and lifestyle diseases. Understanding the causes and symptoms of diseases can help individuals take steps to prevent them or seek appropriate treatment when necessary. Diseases are a deviation from normal functioning of the body or mind, which is typically caused by an infection, genetic disorder, environmental factors, lifestyle choices, or a combination of these factors. Diseases can affect different parts of the body, such as organs, tissues, cells, or even the entire body. Some common types of diseases include infectious diseases, genetic disorders, autoimmune diseases, cancer, and cardiovascular diseases.

Infectious diseases are caused by pathogens such as bacteria, viruses, fungi, or parasites. These diseases can be transmitted from one person to another through direct or indirect contact, and can range from mild to severe. Genetic disorders are caused by abnormalities in a person's DNA, and can be inherited or occur spontaneously. These diseases can affect various parts of the body and can be present from birth or develop later in life. Autoimmune diseases occur when the body's immune system attacks its own tissues, leading to inflammation and damage. Examples of autoimmune diseases include rheumatoid arthritis, lupus, and multiple sclerosis.

Cancer is a group of diseases that involve the uncontrolled growth and spread of abnormal cells in the body. Cancer can occur in various parts of the body and can be caused by genetic factors, lifestyle choices, or exposure to environmental factors such as radiation or chemicals. Cardiovascular diseases refer to conditions that affect the heart and blood vessels, such as heart disease and stroke. These diseases can be caused by factors such as high blood pressure, high cholesterol, smoking, and obesity.

Severe disease generally refers to a medical condition that is serious and may result in significant harm or even death if left untreated. The severity of a disease can be determined by various factors, such as the symptoms experienced by the patient, the extent of organ damage or dysfunction, the progression of the disease, and the likelihood of recovery. Examples of severe diseases include advanced stages of cancer, severe infections like sepsis or pneumonia, heart failure, stroke, and advanced-stage neurological disorders like Alzheimer's disease or Parkinson's disease.

It's important to seek medical attention promptly if you suspect you may have a severe disease, as early detection and treatment can greatly improve your chances of recovery and minimize the risk of complications. Autoimmune diseases occur when the body's immune system attacks its own tissues, leading to inflammation and damage. Examples of autoimmune diseases include rheumatoid arthritis, lupus, and multiple sclerosis. Diseases are a deviation from normal functioning of the body or mind, which is typically caused by an infection, genetic disorder, environmental factors, lifestyle choices, or a combination of these factors. Infectious diseases are caused by pathogens such as bacteria, viruses, fungi, or parasites. These diseases can be transmitted from one person to another through direct or indirect contact, and can range from mild to severe.

2. OBJECTIVE

The objective of disease prediction using symptoms based on machine learning algorithms is to develop a model that can accurately predict the likelihood of a patient having a particular disease based on their reported

symptoms. Machine learning algorithms can learn the complex relationships between symptoms and disease diagnoses from large datasets of labeled data, and can use this knowledge to make predictions on new, unseen data.

3. LITERATURE SURVEY

CLASSIFICATION AND FUNCTIONAL ANALYSIS OF MAJOR PLANT DISEASE USING VARIOUS CLASSIFIERS IN LEAF IMAGES

Plant disease interrupts the normal or ordinary condition of a plant and it alters the essential functionality of a plant. Which intern impacts the productivity of the crops Speedy observation, recognition and categorization of the plant pathogens will increase the crop yield more than 60% of the total productivity. Disease analysis is more evident on the leaves when compare to the other parts of the plants. Automated methods are most commonly available in different image processing techniques to detect the pathogen attack which can be made more efficient by combining multiple domains, that utilizes computer vision technologies. Most modern techniques or technologies are analyzed to identify the various diseases on several crops or crop types. The paper summarizes about types of plants, types of plant diseases and the standard methodologies or technique that would help gaining knowledge about Computer Vision and its applications on plant disease identification and classification. Performance of the Classifiers is analyzed to recognize and classify the better method that typically works among different plant groups and different types of pathogen attack.

IDENTIfiCATION OF APPLE LEAF DISEASES BASED ON DEEP CONVOLUTIONAL NEURAL NETWORKS

Mosaic, Rust, Brown spot, and Alter aria leaf spot are the four common types of apple leaf diseases. Early diagnosis and accurate identification of apple leaf diseases can control the spread of infection and ensure the healthy development of the apple industry. The existing research uses complex image preprocessing and cannot guarantee high recognition rates for apple leaf diseases. This paper proposes an accurate identifying approach for apple leaf diseases based on deep convolutional neural networks. It includes generating sufficient pathological images and designing a novel architecture of a deep convolutional neural network based on Alex Net to detect apple leaf diseases. Using a dataset of 13,689 images of diseased apple leaves, the proposed deep convolutional neural network model is trained to identify the four common apple leaf diseases. Under the hold-out test set, the experimental results show that the proposed disease identification approach based on the convolutional neural network achieves an overall accuracy of 97.62%, the model parameters are reduced by 51,206,928 compared with those in the standard Alex Net model, and the accuracy of the proposed model with generated pathological images obtains an improvement of 10.83%. This research indicates that the proposed deep learning model provides a better solution in disease control for apple leaf diseases with high accuracy and a faster convergence rate, and that the image generation technique proposed in this paper can enhance the robustness of the convolutional neural network model.

A SURVEY PAPER ON PLANT DISEASE IDENTIFICATION USING MACHINE LEARNING APPROACH

Agriculture plays an important role in farmer's life. Sometimes manual identification of disease is time consuming and need of labor is more. One of the most important facts that reduce the growth of plants is disease attack. Overall study about agriculture shows that quality and quantity of agricultural products may be reduced due to various factors of plant diseases. These diseases can be more easily identified by using machine learning approach as compared to manual method. Hence machine learning method can be used to identify the affected leaf images. Images captured by camera will be processed using different image processing technique. These techniques will help in identifying plant diseases thereby increasing the yield of plants. This survey paper describes plant disease identification using Machine Learning Approach and study in detail about various techniques for disease identification and classification is also done. Keywords

DENSE SEMANTIC LABELING OF SUB DECIMETER RESOLUTION IMAGES WITH CONVOLUTIONAL NEURAL NETWORKS

Semantic labeling (or pixel-level land-cover classification) in ultrahigh-resolution imagery (<10 cm) requires statistical models able to learn high-level concepts from spatial data, with large appearance variations. Convolutional neural networks (CNNs) achieve this goal by learning discriminatively a hierarchy of representations of increasing abstraction. In this paper, we present a CNN-based system relying on downsample-thenupsample

architecture. Specifically, it first learns a rough spatial map of high-level representations by means of convolutions and then learns to up sample them back to the original resolution by deconvolutions. By doing so, the CNN learns to densely label every pixel at the original resolution of the image. This result in many advantages, including: 1) the state-of-the-art numerical accuracy; 2) the improved geometric accuracy of predictions; and 3) high efficiency at inference time. We test the proposed system on the Vaihingen and Potsdam subdecimeter resolution data sets, involving the semantic labeling of aerial images of 9- and 5-cm resolution, respectively. These data sets are composed by many large and fully annotated tiles, allowing an unbiased evaluation of models making use of spatial information. We do so by comparing two standard CNN architectures with the proposed one: standard patch classification, prediction of local label patches by employing only convolutions, and full patch labeling by employing deconvolutions. All the systems compare favorably or outperform a state-of-the-art baseline relying on super pixels and powerful appearance descriptors.

RESEARCH OF HEART DISEASE PREDICTION BASED ON MACHINE LEARNING

The use of massive clinical data in the medical field for supporting medical decision support is an inevitable development trend. Medical decision support is based on a variety of data sources accumulated and acquired in real-time in the clinic, and various machine learning algorithms are used to achieve classification of patient disease types or prediction of disease risks. This paper assists in performing cardiac disease prediction starting from different heart disease types (coronary heart disease) and data sets, summarizing the currently adopted machine learning diagnosis and prediction methods, highlighting the characteristics and differences of these methods, and analyzing the challenges and future developments. The results show that machine learning techniques have a wide range of applications in cardiac diseases. However, each machine learning method can only be applied to a specific scope due to the non-uniformity of medical data. At the end of the article, the prediction of heart disease is summarized.

4. SYSTEM ANALYSIS

4.1 EXISTING SYSTEM

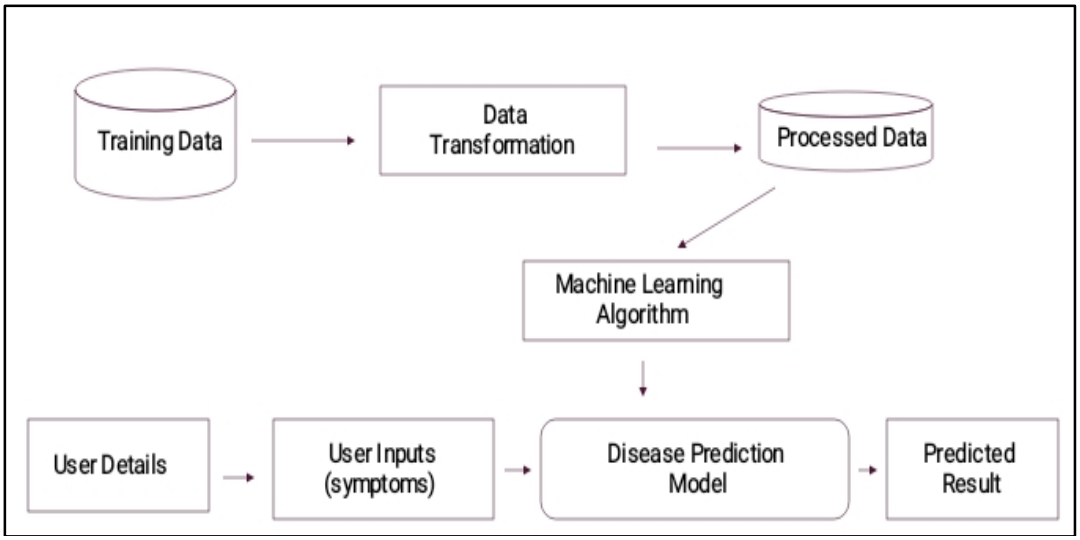
Disease prediction using symptoms based on machine learning algorithms is a rapidly growing field in healthcare. There are several existing systems that use machine learning algorithms for disease prediction based on symptoms. The general approach to developing such a system involves collecting data on the symptoms of various diseases and their corresponding diagnoses. This data is then used to train machine learning models to predict the likelihood of a particular disease based on the symptoms reported by a patient. To develop a new system for disease prediction using symptoms and machine learning algorithms, you would need to collect data on the symptoms of various diseases and their corresponding diagnoses. This data could be collected from electronic medical records, clinical studies, or other sources. Once you have collected the data, you would need to train machine learning models to predict the likelihood of a particular disease based on the symptoms reported by a patient. The disease prediction using symptoms based on machine learning algorithms has the potential to revolutionize healthcare by providing patients with more accurate and timely diagnoses. However, it is important to ensure that these systems are developed and used in an ethical and responsible manner, with appropriate safeguards in place to protect patient privacy and ensure the accuracy of the predictions. The existing systems only predict the diseases and various approaches used for predicting diseases is by using Machine Algorithms such as Naïve Bayes, Decision Tree, Random Forest, k-mean algorithm. Machine Learning and supervised learning algorithm which uses training data with the labels for the training of the models.

4.2 PROPOSED SYSTEM

Disease prediction using symptoms based on machine learning algorithms is a popular and effective way to diagnose and treat various medical conditions. In this proposed system, machine learning algorithms can be trained on a dataset of medical records containing information about patients' symptoms, diagnosis, and treatment. The goal is to create a model that can predict a patient's disease based on their symptoms. The first step in developing such a system would be to gather a comprehensive dataset of medical records containing information about patients'

symptoms, diagnoses, and treatments. This dataset would be used to train the machine learning algorithms to identify patterns in the data and make predictions about patients' health conditions. The next step would be to select a suitable machine learning algorithm for disease prediction. There are many different algorithms to choose from, including decision trees, support vector machines, neural networks, and random forests, among others. The choice of algorithm would depend on the specific requirements of the system and the type of data being analyzed. Once the algorithm has been selected, the dataset would be pre-processed to prepare it for analysis. This would involve cleaning the data, removing any duplicates or irrelevant data, and transforming the data into a suitable format for analysis. Feature selection techniques would be used to identify the most relevant symptoms for disease prediction. To train the machine learning algorithm on the pre-processed data. The algorithm would learn from the dataset and develop a predictive model that could be used to predict a patient's disease based on their symptoms. Finally, the system would be tested on a separate dataset to evaluate its accuracy and performance. The accuracy of the system would be evaluated using metrics such as precision, recall, and F1 score. A disease prediction system based on machine learning algorithms would be a powerful tool for healthcare professionals to diagnose and treat various medical conditions. In the suggested strategy, we use Machine Learning techniques to precisely forecast the ailment that the patient has been suffering from. When past healthcare records are used as a dataset, the results are more accurate. To train the model and predict user diseases based on the symptoms they enter, we use machine learning algorithm.

4.3 DATA FLOW DIAGRAM



5. SYSTEM TESTING

The purpose of testing is to discover errors. Testing is the process of trying to discover every conceivable fault or weakness in a work product. It provides a way to check the functionality of components, sub-assemblies, assemblies and/or a finished product It is the process of exercising software with the intent of ensuring that the Software system meets its requirements and user expectations and does not fail in an unacceptable manner. There are various types of test. Each test type addresses a specific testing requirement.

6. SYSTEM DESIGN

6.1 UML DIAGRAMS

UML stands for Unified Modelling Language. UML is a standardized general-purpose modelling language in the field of object-oriented software engineering. The standard is managed, and was created by, the Object Management Group. The goal is for UML to become a common language for creating models of object oriented computer software. In its current form UML is comprised of two major components: A Meta-model and a notation.

In the future, some form of method or process may also be added to; or associated with, UML. The Unified Modelling Language is a standard language for specifying, Visualization, Constructing and documenting the artifacts of software system, as well as for business modelling and other non-software systems. The UML represents a collection of best engineering practices that have proven successful in the modelling of large and complex systems. The UML is a very important part of developing objects oriented software and the software development process. The UML uses mostly graphical notations to express the design of software projects.

6.2 USE CASE DIAGRAM

A use case diagram in the Unified Modeling Language (UML) is a type of behavioral diagram defined by and created from a Use-case analysis. Its purpose is to present a graphical overview of the functionality provided by a system in terms of actors, their goals (represented as use cases), and any dependencies between those use cases. The main purpose of a use case diagram is to show what system functions are performed for which actor. Roles of the actors in the system can be depicted.

6.3 CLASS DIAGRAM

In software engineering, a class diagram in the Unified Modeling Language (UML) is a type of static structure diagram that describes the structure of a system by showing the system's classes, their attributes, operations (or methods), and the relationships among the classes. It explains which class contains information.

6.4 SEQUENCE DIAGRAM

A sequence diagram in Unified Modeling Language (UML) is a kind of interaction diagram that shows how processes operate with one another and in what order. It is a construct of a Message Sequence Chart. Sequence diagrams are sometimes called event diagrams, event scenarios, and timing diagrams.

6.5 DEPLOYMENT

Component diagrams are used to describe the components and deployment diagrams shows how they are deployed in hardware. UML is mainly designed to focus on the software artifacts of a system. However, these two diagrams are special diagrams used to focus on software and hardware components.

6.6 DATA FLOW DIAGRAM

The DFD is also called as bubble chart. It is a simple graphical formalism that can be used to represent a system in terms of input data to the system, various processing carried out on this data, and the output data is generated by this system.

The data flow diagram (DFD) is one of the most important modeling tools. It is used to model the system components. These components are the system process, the data used by the process, an external entity that interacts with the system and the information flows in the system.

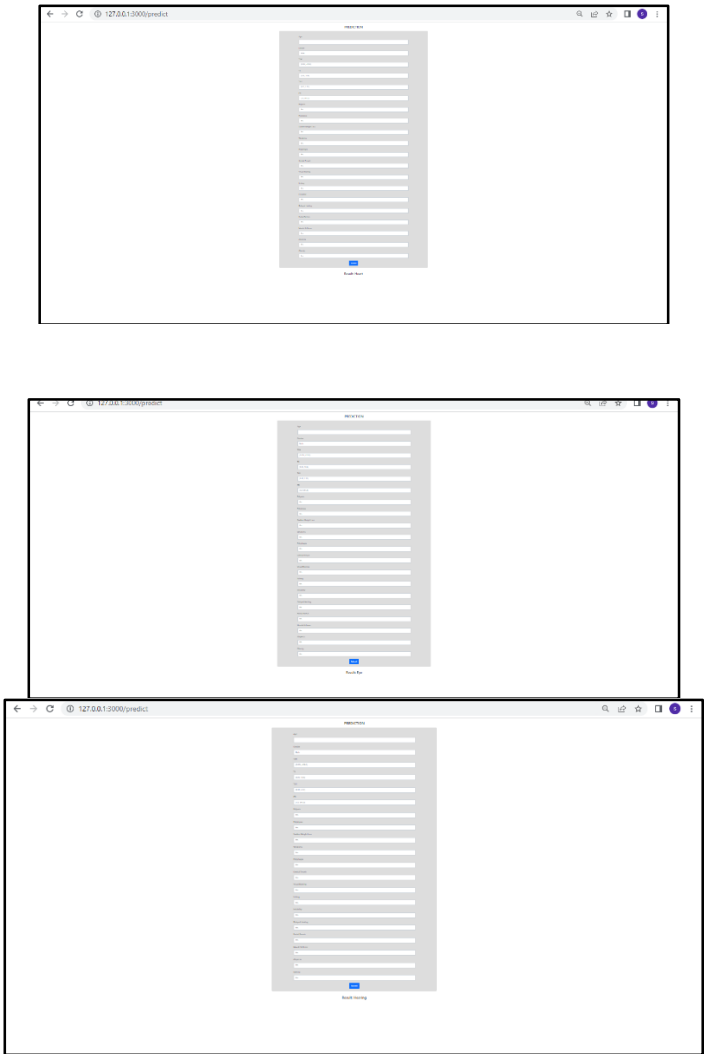
DFD shows how the information moves through the system and how it is modified by a series of transformations. It is a graphical technique that depicts information flow and the transformations that are applied as data moves from input to output.

DFD is also known as bubble chart. A DFD may be used to represent a system at any level of abstraction. DFD may be partitioned into levels that represent increasing information flow and functional detail.

7. CONCLUSION

Disease prediction using machine learning, is to everyone's daily lives, but especially to those in the healthcare industry, who use these systems on a daily basis to predict patient’s diseases based on their general information and symptoms. Because the health industry now plays such a large role in curing patients diseases, this is often quite helpful for the health industry to inform the user, and it's also useful for the user if he or she doesn't want to travel to the hospital or other clinics, because the user can learn about the disease he or she is suffering from simply by entering the symptoms and any other relevant information, and the health industry can benefit from this system.

8. SCREENSHOTS



9. REFERENCES

1) Bhanuteja Talasila, Saipoornachand Kolli, Kilaru Venkata Narendra Kumar and Poonati Anudeep, "Symptoms Based multiple Disease Prediction Model using Machine Learning Approach", *International Journal of Innovative Technology and Exploring Engineering*, August 2021.

2) Nishant Yede, Ritik Koul, Chetn Harde, Kumar Gaurav and Prof. C.S. Pagar, "General Disease Prediction Based On Symptoms Provided By Patient", *Open Access International Journal Of Science & Engineering (OAIJSE)*, June 2021.

3) Ibrahim Mahmood and Adnan Moshin Abdulazeez, "The Role of Machine Learning Algorithms for Diagnosing Diseases", *Journal of Applied Science and Technology Trends (JASTT)*, March 2021

4) P. Hamsagayathri and S. Vigneshwaran, "Symptoms Based Disease Prediction Using Machine Learning Techniques", *IEEE Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV)*, March 2021.

5) Rinkal Keniya, Aman Khakharia and Vruddhi Shah, "Disease Prediction From Various Symotoms Using Machine Learning", *Social Science Research Network(SSRN)*, October 2020.

6) Md. Ehisham Farooqui and Dr. Jameel Ahmad, "Disease Prediction System using Support Vector Machine and Multi Linear Regression", *International Journal of Innovative Research in Computer Science & Technology (IJIRCST)*, July 2020.

7) S Vijava Shetty, G A Karthik and M Ashwin, "Symptom Based Health Prediction using Data Mining", *IEEE International Conference on Communication and Electronics Systems (ICCES)*, February 2020

8) Shraddha Subhash Shirsath, "Disease Prediction Using Machine Learning Over Big Data", *International Journal of Innovative Research in Science*, vol. 7, no. 6, June 2018.

9) Sneha Grampurohit and Chetan Sagarnal, "Disease Prediction using Machine Learning Algorithms", *2020 International Conference for Emerging Technology (INCET)*, 2020
10) M. Maniruzzaman, M.J. Rahman, B. Ahammed and M.M. Abedin, "Classification and prediction of diabetes disease using machine learning paradigm", *Health Information Science and Systems*, vol. 8, no. 1, pp. 7, 2020.
11) A Gupta, M Gupta and P Chaturvedi, *Investing data with machine learning using python*, 2019
12) Sneha Grampurohit and Chetan Sagarnal, "Disease Prediction using Machine Learning Algorithms", *2020 International Conference for Emerging Technology (INCET)*, 2020
13) Dhiraj Dahiwade, Gajanan Patle and Ektaa Meshram, "Designing Disease Prediction Model Using Machine Learning Approach", *IEEE 3rd International Conference on Computing Methodologies and Communication (ICCMC)*, August 2018.
14) Harsh Patel and Purvi Prajapati, "Study and Analysis of Decision Tree Based Classification Algorithms", *International Journal of computer Sciences and Engineering (IJCSE)*, October 2018.
15) B. Dhomse Kanchan and M. Mahale Kishore, "Study of machine learning algorithms for special disease prediction using principal of component analysis", *IEEE International Conference on Global Trends in Signal Processing Information Computing and Communication (ICGTSPICC)*, 2016.