IMPROVEMENT IN EFFICIENCY FOR HEALTH INFORMATICS USING MACHINE LEARNING FOR DIABETES DISEASE

A Thesis Submitted in Partial Fulfillment of Requirements
For the Degree of

MASTER OF TECHNOLOGY

In

Software Engineering

By

RAJAT MANI TRIPATHI
(University roll no. 1200449005)

Under the Supervision of

Ms. UPASANA DUGAL

Assistant Professor
Department of Computer Science & Engineering BBD University, Lucknow
CSE Department BBDU, Lucknow

SCHOOL OF ENGINEERING
BABU BANARASI DAS UNIVERSITY
LUCKNOW, INDIA

May 2022

IMPROVEMENT IN EFFICIENCY FOR HEALTH INFORMATICS USING MACHINE LEARNING FOR DIABETES DISEASE

THESIS

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENT
FOR THEAWARD OF THE DEGREE OF

MASTER OF TECHNOLOGY
(Computer Science)

Submitted by

RAJAT MANI TRIPATHI

BABU BANARASI DAS UNIVERSITY
MAY 2022

CERTIFICATE

It is certified that the work contained in this thesis entitled " **IMPROVEMENT IN EFFICIENCY FOR HEALTH INFORMATICS USING MACHINE LEARNING FOR DIABETES DISEASE"** by Mr. Rajat Mani Tripathi (University roll no. 1200449005), for the award of **Master of Technology** from Babu Banarsi Das University has been carried out under our supervision and that this work has not been submitted elsewhere for a degree.

Signature……

Mrs. Upasana Dugal

Assistant Professor

Department of CSE

School of Engineering BBDU

Lucknow

Date : …….

Abstract

Researchers' need for a tool that would speed up and simplify the process of screening titles and abstracts for systematic reviews and meta-analyses prompted the creation of this instrument. Conducting a systematic review of the scientific literature on a regular basis is important for various activities such as systematic reviews and meta-analyses. To choose which research to include in a systematic review or meta-analysis, academics and practitioners are currently manually assessing hundreds of published articles. Due to the unequal nature of the data, this is both error-prone and wasteful. To keep up with the massive rise in the amount of text that is being made public, the future of systematic review will need collaboration with machine learning algorithms. The use of data mining in healthcare is on the rise as more organisations seek to optimise their operations. Data mining has become a need in the healthcare industry because to the glut of available data. Massive amounts of information are routinely generated and collected by healthcare organisations from their patients. Data mining and knowledge discovery can be automated to uncover some surprising patterns. Data may be pulled directly from electronic records without the need for additional software, eliminating the necessity for manual operations. Medical data can be preserved, lives can be saved, and the cost of medical services can be reduced using information technologies. Patients' physical and behavioural conditions may be analysed utilising data mining approaches by health-care providers. Investigating a wide range of options and creating connections between apparently unrelated pieces of information are two methods for discovering these patterns. The quantity of raw data generated by healthcare institutions is enormous, and it may be difficult to make sense of it all. A single, universal medical information system is only possible if data is gathered and stored in an orderly fashion and then integrated. Using data mining in the medical field, for example, researchers may examine data models that would otherwise be unavailable or obscured by normal analytic methods. An extensive diabetic health dataset is used in this study to increase the accuracy of a prediction of diabetes based on machine learning techniques. Biotechnology and public healthcare infrastructure advancements have resulted in a tremendous rise in the creation of vitally essential and sensitive health-related information. Intelligent data analysis tools for early and onset detection and prevention of several serious diseases have shown many remarkable patterns. Because it increases the risk of heart disease, renal disease, and nerve damage, diabetes mellitus is a life-threatening disorder.

ACKNOWLEDGEMENT

I would like to place on record my deep sense of gratitude to Ms. Upasana Dugal,

Assistant Professor, Department of Computer Science and Engineering, Banu Banarsi Das University, Lucknow, India for her generous guidance, help and useful suggestions.

I also wish to extend my thanks to Dr. Manuj Darbari, Mrs. Akanksha Singh, Miss. Hina Rabbani, Mr. Sunil Kumar Vishwakarma, Dr. Gaurav Kumar Srivastava and other faculty members, colleagues for attending my seminars and

for these insightful comments and constructive suggestions to improve the quality

of this research work.

I am extremely thankful to Dr. Praveen Kumar Shukla, HOD, Department of Computer Science Engineering, Babu Banarsi Das University, Lucknow, India for providing me infrastructural facilities to work, without which this work would

not have been possible.

Mr. Rajat Mani Tripathi

Babu Banarsi Das University

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

ML - Machine learning

WHO - World Health Organization

HI- Health Informatics

SVM - Support Vector Machine

AI - Artificial Intelligence

SDLC- Software Development Life Cycle

DL - Deep Learning

PriMIA - Privacy-Preserving Medical Image Analysis

DSS - Decision Support Systems

ECG – Electrocardiogram

MD- Medical Devices

DTC - Decision Tree Classifier

HCS - Healthcare Services

ELM - Excessive Learning Machine

PCA - Principal Component Analysis

KNN – K- Nearest Neighbor

MLP - Multi-Layered Perceptron

MLPN - Multi-Layer Perceptron Network

SLFN - Single Hidden-Layer Feed-Forward Neural Network

AL – Artificial Learning

TABLE OF CONTENTS

Chapter 1
Introduction

1.1 Introduction:

Blood glucose levels in diabetes patients are abnormally high, this is a sign of disease. When the pancreas, a human gland, is not able to produce sufficient insulin, diabetes develops (Type 1 diabetes) or w while the body's cells are incapable of operating insulin produced (Type 2 diabetes), diabetes develops (Type 2 diabetes), according to medical authorities (Type 2 diabetes) [2, 3]. When the pancreas, a human gland, is unable to produce enough insulin, type 1 diabetes develops (Type 1 diabetes). As a result of the digesting process, glucose is released into our circulation when we eat something. Insulin is a hormone that flows from the bloodstream to the cells, telling them to accept glucose and convert it to energy. he pancreas' beta cells create insulin. The cells' ability to absorb glucose is harmed as a result of the pancreas' failure to produce adequate insulin, and glucose is still present in the bloodstream. Consequently, the level of blood glucose/sugar in the blood rises to dangerously high levels [3], that is extremely damaging. When blood sugar levels in the body are too high, it can cause a range of problems, including extreme hunger, frequent urination & excessive thirst. Advancement of technology has resulted in gadgets producing massive amounts of data on a regular basis. There has been an explosion in the availability of data for academics all around the world. Because of the complexity, enormous amount, and heterogeneity of data, it is necessary to seek for, develop, and implement new software tools and processes in order to successfully manage, analyse, and display the data [1] and to effectively manage, analyse, and visualise the data According to [2,] the author has received results from Google Scholar for the phrase "Big data" during the years 2008 to 2015. These findings demonstrate how this discipline has developed over the years, as well as the rising number of papers in the field of big data. This exponential expansion in the field of big data began in 2012, and this area of study continues to draw an increasing number of academics.

The human race has made significant advances in, materials science, computer science, biotechnology, proteomics & genomics in recent years [6, 7], among other fields. This new generation of disruptive technologies is reshaping the landscape of medical practise. AI and

big data in particular are changing the face of illness and patient care, with the emphasis now turning to individualised diagnosis and therapy. As a result of this transition, health of public may become more predictive & preventative [6].

In artificial intelligence, Deep learning is a branch of machine learning that tries to create computer systems which can recognise patterns in training data & apply those patterns to prediction & classification tasks on new data [7]. For the purpose of creating models, machine learning brings together methods from the fields of statistics, data mining, and optimization.

An important subfield of machine learning, representational learning is concerned with creating an appropriate representation of the information retrieved from data automatically [7]. We're working with deep learning approaches when this representation has multiple layers (i.e., when it's a multi-level representation).

The dataset for Diabetes Disease has been chosen, and it will comprise patient information such as , glucose levels, ag, insulin levels & so on. It will be in the form of a spreadsheet with tabular data. Between 500 and 100 patients will be included in the dataset. Additional methods could be assessed and compared to the improvement, for example, support vector machine, logistic regression, or nave method. Depending on the method utilised, For the detection or prognosis of the illness, the test will produce a positive or negative result. Extreme machine learning approaches have the potential to enhance the accuracy of current systems dramatically. By utilising algorithms and algorithmic methodologies, machine learning technologies have the potential to increase treatment regimen precision and health-care outcomes.

A branch of sophisticated machine learning called brain simulation models how the human brain works, is one of the rapidly growing uses of deep learning in radiology and medical imaging. Deep learning systems are being utilised to find, recognise, and assess malignant tumours in photos captured by neural networks that can learn from data without being supervised.

As processing rates increase, Anomalies in photographs that are not visible to the naked eye can be detected using machine learning algorithm specifications and cloud infrastructures improve, assisting in the diagnosis and treatment of illness.

The healthcare business will continue to transform as a result of machine learning one of two machine learning applications currently in development is predictive analytics for predicting breast cancer recurrence using medical data and photographs. A diagnostic tool for diabetic retinopathy is also in development, as is a predictive analytics tool for predicting breast cancer recurrence. One of the fundamental goals of health informatics (HI) is to build a large web of connections betwixt distinct ideas at a high magnitude of abstraction. In the past, reading data through dataset associating has been problematic in the software engineering field since healthcare datasets are usually always inadequate and noisy. Machine learning (ML) is a rapidly growing field of computer science due to its ability to store massive amounts of data. It's a computer science subfield that's quickly growing as a subfield of computer science. Machine learning (ML) technology could be utilized to analyse data in a variety of ways and give information to help employees and physicians enhance their work quality; However, there is currently no means for developers to review data and create information. Diabetes is one of the most deadly viruses that can strike at any moment and in any area. Obesity, high blood glucose levels, and other factors can contribute to the development of diabetes. It causes crabs to have an irregular metabolism while also reducing quantity of sugar in their bloodstream by inhibiting insulin hormone. Diabetes is a disease in which the body's insulin production is insufficient to meet the body's needs. According to the World Health Organization (WHO), diabetes affects 422 million people globally. Only a small fraction of individuals impacted live in high-income countries, with the vast majority living in low- and middle-income countries. Some estimates suggest that by 2030, the entire amount of money in the globe will be $490 billion. In contrast, type II diabetes antiquated discovered to be ubiquitous in a number of countries, together with China & Canada. The population of India has risen to almost 100 million people, as a result, there are now 40 million diabetics. Diabetes is a primary cause of death worldwide, affecting people of all ages and backgrounds. Detecting and treating illnesses like diabetes early can assist to decrease the disease's progression & prolonging the patient's life. The purpose of this study is to look at diabetes prediction using a number of risk factors connected with the condition. The Pima Indian Diabetes Dataset is used in conjunction with a variety of Machine Learning classification and Methods for predicting the onset of diabetes using ensemble. Machine Learning is a method for teaching computers or machines to perform specific tasks. Deep learning is another name for it. Various Machine Learning Techniques deliver successful outcomes in the accumulation of information by using obtained datasets to develop diverse classification and ensemble models.

Diabetes mellitus is a collection of metabolic illnesses defined by hyperglycaemia as a consequence of abnormalities in insulin production, insulin action, or a combination of these problems [1]. The presence of insulin resistance (insulin action deficiency), in particular, is connected with type 2 diabetes [2]. Insulin resistance is defined as cells that react poorly to insulin, hence increasing their glucose consumption. The American Diabetes Association has established the following diagnostic criteria: (1) a glycated haemoglobin (HbA1c) level greater or equal to 6.5 percent; (2) a fasting blood glucose level of more than 126 mg/dL at rest; (3) a blood glucose level of 200 mg/dL or higher 2 hours following a 75 g glucose oral glucose tolerance test [1].

Diabetes is a major public health issue that affects individuals all over the world. In 2019, the International Diabetes Federation estimated that 463 million people worldwide had diabetes, with a predicted increase of 51 percent by the year 2045, according to their latest estimates. In the United States, it is also predicted that for every diagnosed diabetic, there is one undiagnosed individual [2].

Preventing type 2 diabetes from progressing further and developing problems such as diabetic retinopathy are among the most important things you can do to keep it from becoming worse. An early diagnosis, according to the findings of the ADDITION-Europe Simulation Model Study, lowers the absolute & relative risk of cardiovascular events & mortality [4]. A sensitivity study of data from United States revealed that a 2-year earlier diagnosis resulted in a 25 percent relative decrease in diabetes-related complication rates.

Various machine learning algorithms are suited for varied types of data and have different sizes of data, and each method has its own set of restrictions. Predictive analytics in healthcare is the topic of this research, which is written in English. A big dataset of healthcare data is gathered for the aim of the experiment, and multiple machine learning methods are applied to the dataset. The performance and accuracy of the algorithms that were used are addressed in relation to the nature of the dataset. The study's goal is to present enough information to the reader to enable them to comprehend how the healthcare sector may use big data analytics to enhance decision-making & illness prediction. Second, the performance assessment of machine learning algorithms in the context of predictive analytics for diabetic illness is being investigated. Tool used in this research is Python and its relevant packages.

1.2 OBJECTIVES

The primary goals of this research are to:

• Apply machine learning algorithms to health information datasets for illness prediction and detection.

• Develop a machine learning algorithm for disease prediction and detection.

• To determine the accuracy of a KNN-based algorithm that has been built.

In order to enhance the algorithm on the basis of accuracy assessment and comparison with prior results, an extreme machine learning method will be used.

Chapter 2

Literature Review

2.1 Literature Review

[1] Moreb et al. The SEMLHI framework can be used by researchers and developers to organise jobs using the SEMLHI technique. The framework is made up of four modules (artificial, software, learning, fitness informatics data & artificial learning algorithms) that allow developers & researchers to do a software engineering analysis on fitness informatics & give developers with a new blueprint for integrating fitness apps with programme operates & deployments of freeware. The author's unique method displays its qualities by allowing users to investigate & determine the purpose of system components and the artificial learning methods to be used to the dataset while evaluating user needs. Data from a Palestinian government-run hospital was used to build the author dataset, which was collected over a three-year period.

[2] Beam et al. Artificial intelligence and big data are redefining practically every element of modern life and changing the way we live. Netflix and Google know what movies people want to see and what information they want. Google has only recently begun to use artificial learning algorithms to replace a major amount of their non-artificial learning technology, according to reports, & there is a lot of hope that these approaches will be able to offer comparable results across a wide reach of businesses.

[3] et al. The data from all of the experiments has been organised so that artificial intelligence practitioners and researchers can acquire relevant artificial intelligence algorithms for fitness care more rapidly. We've compiled a list of the finest algorithms based on the papers we've read so far. In this study, no preference is given to any of the methodologies stated in the literature. The author was only concerned with the correctness of the algorithms. The authors expect that the results of this research could assist practitioners and researchers avoid

exhaustion that comes with trying to find the optimum algorithm for forecasting diseases listed.

[4] et al. Rahman et al. COVID-19 was eventually identified as a contagious and potentially lethal viral illness, and as a result of its rapid and extensive spread, It has become one of the most hazardous diseases on the planet. Despite the fact that a number of projects & clinical trials are now in progress, no effective vaccine or medicine to ensure its mitigation has been created. Countries all across the world are attempting to contain the spread of COVID-19 while also researching early identification and treatment options. Experts in the field of fitness care, researchers, and scientists in this topic have been researching both current and emerging technology. The circumstance necessitates the creation of a clinical decision support system which could supply medical workers with the tools they need to diagnose the illness early on.

[5] et al. Patel et al. Sickness prediction is greatly aided by artificial learning techniques. The Illness Prediction System employs predictive modelling to anticipate a user's illness based on a variety of symptoms entered into the system. The programme examines the user's detailed symptoms and calculates the likelihood of the illness occurring. Multiple Supervised Artificial Learning algorithms are used in the disease prediction technique. With the rise of big data in the biomedical and fitness-care industries, proper medical data analysis will undoubtedly benefit in the detection and in the future, disease treatment. The ultimate goal of this research is to help academics and practitioners choose an appropriate conclusion for use in fitness care work. Main goal of this paper is to impart up-to-date knowledge on current state of Supervised Artificial Learning algorithms in fitness care market. To determine which learning algorithm was best, the author produced a data table & evaluated the accuracy of learning algorithms for different disorders.

[6] Shams & co-authors Recently, COVID-19 pandemic has impacted nearly every element of human life around the world. There has been a lot of discussion about food estimation and diet during the current pandemic. In this study, The Support Vector Machine (SVM) is handed-down to look at the influence of the COVID-19 pandemic on a diet & predict how many people will die as a result of the pandemic. The data used in this investigation included energy intake (kcal), fat content, protein content & food supply amount (kg) for a variety of dietary categories. Furthermore, the authors are worried about the impact of obesity, cereals (except beer), animal products & Effects of plant-based diets on human health in the case of

a pandemic. here are also confirmed deaths and cases that have been recovered in the data, and as a percentage of each country's current population, active cases are calculated.

[7] Hari Kumar et al. In the agricultural and fitness care industries, Artificial intelligence (AI) has been used to increase crop yields, detect disease outbreaks, improve supply chain management, and increase operational efficiency & continuous monitoring, in addition to reducing water waste the major purpose is to create a standardised system, reliable product quality control systems and to find new methods to engage with and help society while staying within budget. The two most extensively utilised artificial intelligence approaches are artificial learning and deep learning. This type of model is used by individuals, businesses, and government agencies to forecast & learn from data. To deal with the diversity & complexity of data present in real-world applications, artificial learning models are increasingly being created. In the agriculture, fitness care, and management fields, artificial learning and artificial intelligence are extensively discussed, and social studies fields in the following article. Some of the most widely utilised applications are irrigation optimization, sales growth, illness prediction & sales forecasting & profit maximisation. Inventory management, security, fraud detection & portfolio management are some of the other applications.

[8] Mienye et al. For automated medical diagnosis, hospital data is becoming a more valuable source of information. These data, when correctly analysed, reveal hidden patterns & connections which can aid in diagnosis. Because most medical datasets are highly imbalanced, the majority of ML applications to these patient records have depended on directly using machine learning methods, this typically leads to poor performance. Labelling a big amount of medical data is a time-consuming and costly process. The creation of improved artificial learning approaches has been the focus of recent study, includes pipelines for pre-processing and to overcome these challenges, feature learning algorithms.

[9] Fetaji et al. Using three diabetes datasets, the researchers test the performance of three well-known ML algorithms for polygenic prediction. The public domain datasets account for about half of the datasets, while the remaining half is a cluster of facts from an experimentational research class. J48, Random Forest, and Nave Bayes are all artificial learning algorithms that have been examined and proven to be successful. AL is a technique for inspecting & anticipating data sets that are simply unduly large as humans to handle on daily grounds. Mainly correct in the field of pharmaceutical fact, where AI and problem-

solving are still in their early stages of development. To be more specific, this study looks at how artificial learning algorithms may be applied to the massive amount of data from fitness care business on the global diabetes epidemic.

[10] Khan et al. Artificial learning (ML) is seen as an operating method and by mining hidden patterns in software measures, an operational technique for finding problematic modules with shifting components (attribute. Machine learning approaches are also being used by a number of fitness care researchers on large dataset. This work uses a variety of machine learning approaches to predict software flaws using seven well-known datasets.

[11] Lalmuanawma et al. Researchers are encouraged to look into these tactics since there is evidence of the application of Artificial Intelligence (AI) and Machine Learning (ML) during earlier pandemics has inspired them to think of new ways to combat the upcoming Coronavirus epidemic. The goal of this research is to look into the role of artificial intelligence and machine learning as important tools in the SARS-CoV-2 and related pandemic screening, forecasting, contact tracing, and therapy development sectors.

[12] Khan et al. Early in the software development life cycle (SDLC), software defect prediction (SDP) remains a significant and major responsibility. Because of its potential to dramatically improve the overall quality of software systems, SDP has been the topic of extensive research for decades. The ability to predict inaccurate or Early in the software development process, erroneous artefacts can help the development team make better use of existing assets, enable them to produce high-quality software in a short period of time. Several canvassers have already published industrialised fault prediction models based on machine learning (ML) and statistical methods. Artificial learning (ML) is seen as a practical and operational way for locating faulty modules with moving components by mining hidden patterns in software measures (attributes). Machine learning approaches are also being used by a number of fitness care researchers on large datasets.

[13] Scott et al. To detect and diagnose sickness, artificial learning systems are being used, as well as predicting prognosis and therapy response Hundreds of new algorithms are in the works, but it's uncertain if they'll improve clinical decision-making and patient outcomes in the long run. If doctors are going to use algorithms, they need to understand that the most crucial concerns about their value, validity, safety, and practicability are all answered in the affirmative & ethical questions have been addressed. There are certain drawbacks to using algorithms. The author's ten-question checklist, which doctors can use to assess the pros and

cons of a certain algorithm, Professionals are not obliged to show that they comprehend potentially difficult statistical and computational principles, as non-experts would be asked to do.

[14] Mijwil et al. new types arise as chemical waste accumulates on author planet; there is a high level of hazardous contamination; daily life becomes more difficult; human psychology deteriorates, and other elements all have a role in the genesis of a variety of human diseases, including lethal disorders like COVID-19. Multiple illnesses appear differently in different people; some people could be aware of their condition, Others, on the other hand, may be unaware of their fitness level due to a lack of information, resulting in death or a chronic illness that lasts the rest of their lives.

[15] Zhao et al. Deep learning (DL) has become a hot study topic as a result of its rapid growth, Object identification, segmentation, artificial translation & speech recognition are just a few of the categories where we've redefined state-of-the-art performance. As low-cost sensors become more prevalent, data-driven artificial fitness monitoring is becoming increasingly common in contemporary production systems and Internet connections become more widely deployed. Deep learning, on the other hand, allows large amounts of data generated by industrial machinery to be analysed and interpreted. The primary goal of this research is to give an overview & a summary of recent deep learning research on artificial fitness monitoring that has recently been published.

Kumar et al. [16] Using a range of different supervised learning methodologies, a number of researchers have recently constructed a varied array of automated diagnostic models. Early detection of disease may help to reduce the number of people who die as a result of certain illnesses. This study uses machine learning methods to create an effective automated sickness diagnostic model, which are discussed in depth elsewhere the author examines, among other things, the influenza virus, cardiovascular disease, and diabetes in this research. The analysis in the proposed model is carried out using a firebase-deployed pre-trained artificial learning model based on the same dataset, which is then displayed in android application. An Android application is used to enter the data into the suggested model. The results of the analysis are then displayed on the Android application.

[17] Arvanitis et al. Patients, fitness care professionals, and the fitness care system can all benefit from digital fitness applications, which provide a variety of resources to patients, fitness care professionals & the fitness care system. The introduction of new technology is not

without risk & digital fitness apps are frequently designated as medical devices in the United States. Clinical validation is required to ensure their safety, this needs the use of big datasets in order to test its applicability in real-world clinical settings. Due to concerns regarding patient privacy, access to such databases may be difficult. Fauxs datasets that are realistic enough for evaluating digital applications are being developed as a viable option which could help with their acceptability.

[18] Kaissis et al. To procession AI systems on enormous, transnational datasets for zenith-implementation medicinal imaging, it's critical to innovate in privacy-preserving artificial learning. This eliminates the need for data transfer when training models on sensitive data. PriMIA (Privacy-Preserving Medical Image Analysis), a free open-source software framework for differentially private, securely aggregated federated learning, is discussed in the following article & using medical imaging data to make encrypted inferences. To recognise paediatric chest X-rays the assessment is founded on a real-world commentary in which a deep convolutional neural network was trained. PriMIA is used to compare the classification performance of the resulting model against that of locally trained, non-securely trained models. A gradient-based model inversion approach is utilized to recover valuable data from useless data. The author gauges the author framework's staging & seclusion confirms theoretically & empirically, proving that it protects in place, likely prevent reconstruction of useful data from useless data. To protect the data and model from being compromised, using secure multi-party computing, In an end-to-end encrypted remote inference scenario, the trained model is successfully deployed, as explained in the previous section. Due to concerns regarding patient privacy and business interests, it may be challenging to obtain medical data for the purpose of training artificial intelligence systems.

[19] Ojo et al. The goal of this research was to collect data using artificial intelligence technologies to aid policymakers in making decisions which will guarantee the achievement of Fitness Sustainable Development Goals. Agglomerative The findings were visualised using a dendrogram, based on personnel data, which used hierarchical clustering to categorise states (community fitness workers, number of doctors, midwives & nurses). The only people who are regularly scattered across all states are community fitness professionals, according to the exploratory research, and the number of prenatal care hospitals in the North-western states is the lowest. In a hospital, the availability of doctors has little bearing on the implementation of maternal fitness programmes, with community fitness workers, nurses, and midwives being the most important determinants. In order to identify the most important factors that influence

the obtainability of tender fitness shipment amenities in a sanatorium, researchers looked at the following factors, to obtain a feature relevance score, the Random Forest model was used.

[20] Rahman et al. Medical data mining holds a lot of promise for disclosing hidden patterns in big volumes of pharmaceutical data. These specimens could be used to aid in the diagnosis of clinical issues. Raw medical data, on the other hand, is widely dispersed, diverse in type, and massive in size. These facts must be gathered in a systematic and orderly manner. The information gathered can be integrated to create a hospital information system if needed. Data mining is a simple method for discovering new & in enormous amounts of data, there are hidden patterns. Data mining and statistics are both aimed at identifying patterns and structures in massive amounts of data. Data mining is limited to dealing with diverse geographies, whereas statistics is limited to dealing with heterogeneous numbers. With the purpose of knowledge discovery, a wide range of fitness care disciplines have been recognised as potential application areas for these approaches in fitness care databases. This document provides a fast overview of a number of data mining techniques, having a focus on medical diagnosis, including artificial neural networks.

[21] Toh et al. Artificial learning approaches, when used in fitness care, take use of the growing quantity of fitness data made available by the Internet of Things to enhance serene aftermath. These methods have a wide range of potential uses, additionally come with a lot of drawbacks. Natural language processing of medical data, medical imaging & genetic information are the three main disciplines where artificial learning is applied. The diagnosis, detection, and prediction of illness are studied in a variety of areas. While there is a large armature of medical gadgetry that spawns' facts, there is often a dearth of underpinning armature to efficiently employ those facts. Multitude about ways to gather pharmaceutical data generates formatting issues and may lead to a rise in noise. The chapter presents a summary of AL, moreover as a foundational understanding of the actions presupposed in with course condition of the automation in the fitness area.

[22] Akhil et al. Artificial intelligence (AI) is a rapidly expanding area. The objective of AI is to recognising patterns & constructing inference from data. Artificial intelligence-assisted precision medicine is a type of individualised treatment. Artificial learning approaches have made significant advances in the field of fitness care. Artificial learning and its applications in fitness care are the focus of this research. Artificial intelligence will revolutionise the way fitness care is offered in a few years ML and AI (machine learning & artificial intelligence)

will be applied to improve fitness care in the future, However, to solve the issues that patients and clinicians face during illness diagnosis, Decision support systems (DSS) based on machine learning (ML) & artificial intelligence (AI) will be required.

[23] Manne et al. Artificial intelligence (AI) is becoming more common in a variety of fitness-related industries, Artificial intelligence is gaining traction in fitness-related businesses of all sizes, types & specialisations interested in learning more about how it has progressed and how it is serving patient requests and while simultaneously lowering expenses and increasing efficiency. A look at many research papers that utilised AI models in various fitness-related areas, including radiography, drug creation, dermatology & others field, Artificial intelligence's impact on fitness-care management & limitations of employing artificial intelligence in fitness care are examined in this paper.

et al. Gavhane et al. [24] Given the alarming rise in the number of children and teenagers suffering from heart attacks, the author must create a system that can detect the signs of a heart attack before it happens and so avoid it from happening in the first place. Because the average person cannot afford to have expensive tests like the ECG performed on a regular basis, to aid in heart disease prevention, it is necessary to develop a method that is both convenient & accurate in predicting the likelihood of developing heart disease. As a result, the authors propose developing an app which could predict the likelihood of developing a heart ailment based on basic symptoms like pulse rate, gender & age. Neural networks artificial learning method was chosen for usage in the proposed system since it has been demonstrated to be the most accurate and reliable technique.

[25] Debnath et al. TAs a outcome of this viewpoint, the importance of evidence-based prediction tools in a variety of therapeutic contexts is highlighted, using familiar tools for helping hospital frontline staff & administrators, in making well-informed decisions on patient care & handling the COVID-19 epidemic's hospital volume is also a possibility.

[26] Kovacevic et al. Electronics and artificial intelligence have advanced as a result of developments in both fields, medical devices (MDs) have become more intricate (AI). Given how different MD management practises are now than they were decades ago, it's reasonable to enquire about the author's long-term goals. The aim of this research is to show how machine learning (ML) approaches can help with the management of new-born incubators in fitness centres.

[27] Rani et al. Diabetes is a chronic disease that has the potential to destroy the global fitness-care system. According to the International Diabetes Federation, 382 million people worldwide suffer with diabetes. By 2035, this number will have more than tripled to 592 million. Diabetes is a condition in which the blood glucose level is unusually high. As a result of the elevated blood glucose levels, the patient experiences frequent urination, increased thirst, and increased hunger. Diabetic retinopathy, renal failure, amputations, heart failure, and stroke are the most common complications of diabetes. When authors eat, their bodies convert the nutrients in their food into sugars, also known as glucose. Normally, the author's pancreas should be producing insulin at that moment. Insulin works as a solution, allowing glucose into cells later utilising it for energy. It's made by the pancreas. However, in the case of diabetes, this strategy is ineffective.

[28] Thenabadu et al. Diabetes is a fatal chronic disease which affects the entire body. Millions of people are affected by this disorder, and Every year, a large number of individuals die as a result of disease's devastating repercussions. A diabetic's blood glucose level is abnormally high. Diabetes that goes undiagnosed can lead to nerve & heart, kidney damage & blood vessel illness, poor wound healing, hearing loss, and a variety of skin issues, to name a few. In order to live a healthy lifestyle, diabetes must be detected early. Artificial Learning systems, which have just been developed, are being used to address this essential issue.

[29] Baig et al. Diabetes is one of the most deadly metabolic diseases which can affect anyone alive today. Diabetes is a long-term condition that decreases fitness. Diabetes affects roughly 400 million people worldwide, according to the World Fitness Organization a large number of people are infected with this ailment every day, and many are completely unaware that they are afflicted. Type I diabetes & type II diabetes are the two types of diabetes. Diabetes & diabetes mellitus are not interchangeable terms. Diabetes is diagnosed through a physical examination and a blood sugar test. To find a cure for this ailment, more research is required. Artificial intelligence and cloud computing will be key components of the investigation and treat illness in the general public in a timely manner. Diabetes is primarily a condition that affects the elderly and obese. Diabetic foot syndrome, heart attack, kidney failure, hypertension, and diabetic foot syndrome are just a few of the major health problems that diabetes can cause.

[30] Lyngdoh et al. The goal of this research is to look at five supervised machine learning algorithms for diabetic illness prediction: Decision Tree Classifier (DTC), K-Nearest

Neighbours, Nave Baye, Support Vector Machine Learning (SVM) (SVM) & Random Forest. Furthermore, after amalgamating every single one of the risk factors existents in time the trial, author lighted on this, after categorising & cross validating the dataset, accuracy remained steady. With the KNN classifier, the authors achieved the greatest and 76 percent accuracy is the highest level of consistency, The other classifiers, On the other side, they were all over 70% accurate. One of the authors' main goals is to get the best diabetes sickness prediction outcomes possible with the data given. When it comes to diabetes risk detection, traditional techniques of chronic disease management rely on rule-engines or score estimates, which are not nearly as successful as machine learning approaches when it comes to recognising patient health issues [4]. Machine learning and feature engineering advancements in recent years [5], [6] have opened the door to new possibilities for improving the early diagnosis & treatment results of people with chronic diabetes. Several recent research in diabetes management have identified the most often used kinds of technologies, that include insulin self-management applications, wearable blood glucose monitors, automated SMS risk warnings & virtual diagnostician coaching [3, 4]. Using a literature review, Su [7] shown that self-monitoring and frequent intervention resulted in a drop in patient glucose levels, demonstrating the usefulness of RPM approaches for diabetes treatment. One of the most significant shortcomings of present systems is that they are comprised of separate diabetes prediction models that do not interact with real-world patients or data. A notable deficiency is that there are currently no unified healthcare device vendor-independent frameworks which could acquire multi-faceted patient readings in order to facilitate simple monitoring & management [8].

Chapter 3
Problem Statement

3.1 Problem Statement

Insulin resistance is a disorder that affects the way your body processes glucose. It is also referred to as diabetes mellitus in certain circles. Insulin insufficiency & inefficient insulin usage are two conditions that occur when the pancreas is unable to create enough insulin or when human body is unable to use the insulin produced. Diabetes occurs when the pancreas produces insulin, which aids in the movement of glucose from the bloodstream to the cells for use as energy. The most frequent complication of untreated diabetes is hyperglycaemia, or high blood sugar, when combined with other health problems, may lead to significant health problems, particularly in the nerves and blood vessels. According to 2021 figures, persons aged 18 and older were diagnosed with diabetes & according to 2020 estimates, diabetes alone was responsible for 1.5 million fatalities. It is now much easier to diagnose diabetes in its early stages, thanks to the fast rise of machine learning (ML) & deep learning (DL) classification algorithms in a variety of fields, including health research.

The ability to capture, share, and distribute data has become more crucial as a result of the pervasive influence of digital disruption on every business, including healthcare. ML , big data & AI may every single one of them be employed during future to deal with the problems brought on by massive volumes of data.

Healthcare organisations could benefit from machine learning to fulfil increased demand for medical services, streamline procedures, and cut costs.

Machine learning innovation at the point of care could help doctors diagnose and treat illnesses faster, more precisely, and with more personalised treatment than previously conceivable.
A study of machine learning in healthcare indicates how technological advancements could lead to more effective, comprehensive treatment practises, perhaps improving patient results.

3.2 Research Gap

Diabetes Mellitus is a chronic condition that, if left untreated for an extended period of time, may be fatal. AI is assisting professionals in the healthcare business to a significant degree by allowing them to extract meaningful information and patterns from data that is accessible in a variety of forms, including survey data, electronic health records, and laboratory data, among others. Diabetes, if detected at an early stage, has the potential to save many lives while also reducing healthcare expenses. In recent years, there has been an increase in the use of artificial intelligence in diabetes decision-making, diagnosis, and prediction. Diabetes prediction has been the subject of several papers, and it continues to be a study issue of interest due to the availability of fresh data and approaches. The greatest benefit is the improved precision with which disease can be predicted. For early sickness diagnosis, our health-care system demands specialised smart innovation and treatment in these difficult times so that it can be treated, and other health-care therapies can be considered. Since Covid-19 has radically altered the medical landscape, this work has not previously been considered, and there is a critical need for it now. As a result, by utilising an appropriate dataset, this study has the potential to give considerable benefits to the health sector's growth.

One of the most critical issues is the accuracy of data gained through testing methodologies. It's possible that a false positive or false negative will lead to a misdiagnosis of a disease.

Artificial intelligence will become more widespread in healthcare industry due to the complexity and volume of data. Artificial intelligence is already being used by payers, healthcare providers, and companies in the health sciences industry in diverse ways. Diagnostic and therapeutic suggestions, patient association & fidelity & managerial tasks are among the bulk typical application areas. Artificial intelligence can do preventive medicine chores moreover as or more than humans in many circumstances. Concerns with enactment, on the other hand, for a long time, extensive automation of healthcare professional employment will be postponed.

3.3 Justification and Discussion

Machine learning has risen to a prominent position in the field of healthcare services (HCS) as a result of its potential to enhance illness prediction in the HCS field. In the HCS field, machine learning techniques & artificial intelligence have previously been utilized and studied. Diabetes has recently emerged as a prominent public chronic illness affecting people all over the globe. It is quickly expanding as a result of unhealthy lifestyles, increased consumption of fast food, and a general lack of health knowledge. Thus, there is a need for a framework that can efficiently track and monitor people's diabetes and health status from the perspective of an application view. Incidence and prevalence of diabetes are increasing gradually over the world, placing a significant strain on health-care systems around the world. It is predicted that the prevalence of all kinds of diabetes in people would rise by 69 percent in developing nations and by 20 percent in developed countries between 2010 and 2030. In the examination of massive datasets, machine learning is a relatively new study area. Early detection of diseases such as diabetes and other ailments is critical in today's world. When it comes to the advancement of health informatics research and development, The most crucial aspects for standing out from the competition are machine learning innovation and algorithm refining. The upgrade has the benefit of improving the correctness parameter, resulting in a boost in efficiency.

Chapter 4

Methodology

4.1 Methodology

A method of extreme machine learning that can be applied to health informatics prediction, is a viable solution for accuracy improvement that can be implemented. Excessive learning machine (ELM) is a novel feedforward neural network learning approach that consumes a significant quantity of input and contains a single hidden layer. The empirical risk minimization hypothesis underpins ELM and may be learned in a single iteration, as opposed to other methods. A number of rounds and local reduction are avoided by using this technique.

Figure 4.1 depicts a flow diagram of the process described above.

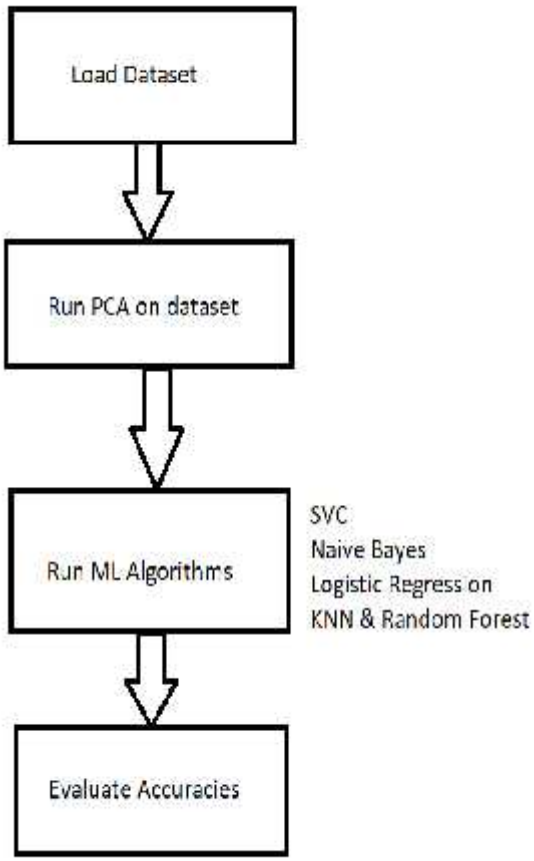

Figure 4.1: Flow Chart

To improve disease prediction in health-care systems, the author recommends merging Software Engineering with Machine Learning methods and in the reference base paper [1], lower the time necessary to anticipate disease as a result of concerns such as a shortage of appropriate hospitals or beds to handle an increasing number of patients. By merging software and machine learning techniques, we can improve our results, the author hopes to solve the problem of predicting disease in a shorter amount of time. SEMLHI is the notion proposed by the author in his work (where SE stands for software, ML for machine learning, and HI for health information). SEMLHI, according to the author, is composed of four components.

1) Data on Health Informatics: To predict any disease, Machine Learning models must be built from datasets, and these datasets commonly contain missing data, null values, and non-numeric data, all of which might reduce the accuracy of ML prediction. To solve this challenge, The author utilises PREPROCESSING to remove all missing and null values from health-care data, & the non-numeric input is subsequently converted to numeric data using

Python. PREPRO SKLEARN Frequently, there will be unused columns or attributes in this dataset, & the author will use a dimensionality reduction method known as PCA to eliminate these unwanted columns or attributes. PCA (principal component analysis) is a technique for removing unimportant variables from a dataset and keeping just the most relevant qualities that are required to create a successful prediction.

In this lesson, the author uses a variety of machine learning methods, such as Multinomial Naive Bayes, Linear SVC, Logistic Regression, Random Forest, , and Kernel Neural Networks. This approach will train itself using publicly available datasets before constructing a train model, which will then be used to fresh test data in order to do prediction. By using the aforementioned techniques, you may teach a computer to learn and anticipate without the need for human assistance.

3) Machine Algorithm Model: Once the author has constructed the models described above, he or she may use the model to forecast whether or not patient lab results are positive or negative.

4) Software: By applying software quality checks, software verification procedures &UNIT TESTING developers may verify that the above modules are reliable.

Author proposes work by applying classification, clustering, and regression to a variety of dataset sizes, and the author uses the Palestine Hospital dataset to apply this notion, which is not available on the internet and has not been released by the authors, so the author uses the INDIAN DIABETES dataset instead of the Palestine Hospital dataset.

The extreme machine technique is used to the health dataset as a case study. In the 1990s, a new learning technique for single hidden layer feedforward neural networks called the Extreme Learning Machine was developed (ELM). This solution eliminates the sluggish training pace and over-fitting difficulties associated with the conventional neural network learning technique when compared to the standard neural network learning technique. Extreme learning machines are made up of one or more layers of hidden nodes, with hidden node parameters learned by trial and error. They are utilized in a variety of applications, including regression, classification, sparse approximation, grouping, feature learning & compression. A multi-layered perceptron (MLP) is a type of neural network that is commonly

used in DL. It is also investigated as a multi-layer perceptron network (MLPN), in some faction (MLP). An MLP is frequently presumptuous as "vanilla" neural-network, significantly slighter complicated from today's complex models. On the other hand, the approaches it pioneered cleared the path for the future building of much more complicated neural networks. A multi-layered perceptron is made up of a linked network of neurons that communicate with one another, similar to how the human brain works. Each neuron has been assigned a numerical value. It is possible to partition the network into three major levels in order to simplify it. Along with SFLN, the ELM algorithm represents a single hidden-layer feed-forward neural network (SLFN). The SLFN's efficiency should be high enough to allow for more learning than is now possible when considering data such as threshold level, weight, and activation function while modelling a system

Excessive learning machine (ELM) is a Google-developed, very efficient learning technique for training single-layer feed-forward neural networks (SLFNs). As unsupervised learning algorithms have improved in recent years, The use of ELM and autoencoders to extract features from unlabelled data has emerged as an attractive new method.

The extreme learning machine (ELM) is often used in batch learning, sequential learning & incremental learning applications due to its quick and efficient learning speed, short convergence period, high generalisation ability & ease of installation. Throughout evolution of the classic ELM, ELM algorithms that are more effective have been developed in a number of ways. In addition, the original ELM's field of applications has been expanded to encompass semi-supervised & unsupervised learning, among other things, in addition to supervised learning.

Electronic Learning Machine (ELM), a unique machine learning technique that consists of layers of hidden nodes (also known as neurons), where the input weights are supplied at random. Due to the fact that it is a feedforward neural network, data only flows in one way through the sequence of layers. In order to address certain challenges, it makes use of random projection and early perceptron (a mathematical model of a neuron). ELM may be used for a variety of tasks such as classification, logical regression, clustering, and others. ELM is based on the structure of "generalised" single hidden-layer feedforward neural networks, in that the hidden node parameters are created at random via a random number generation process. In terms of the usual optimization approach, the ELM issue might be represented as an optimization problem, which would be similar to the formulation of the optimization problem

for the Support Vector Machine (SVM). However, as compared to ELM, SVM tends to provide a solution that is suboptimal.

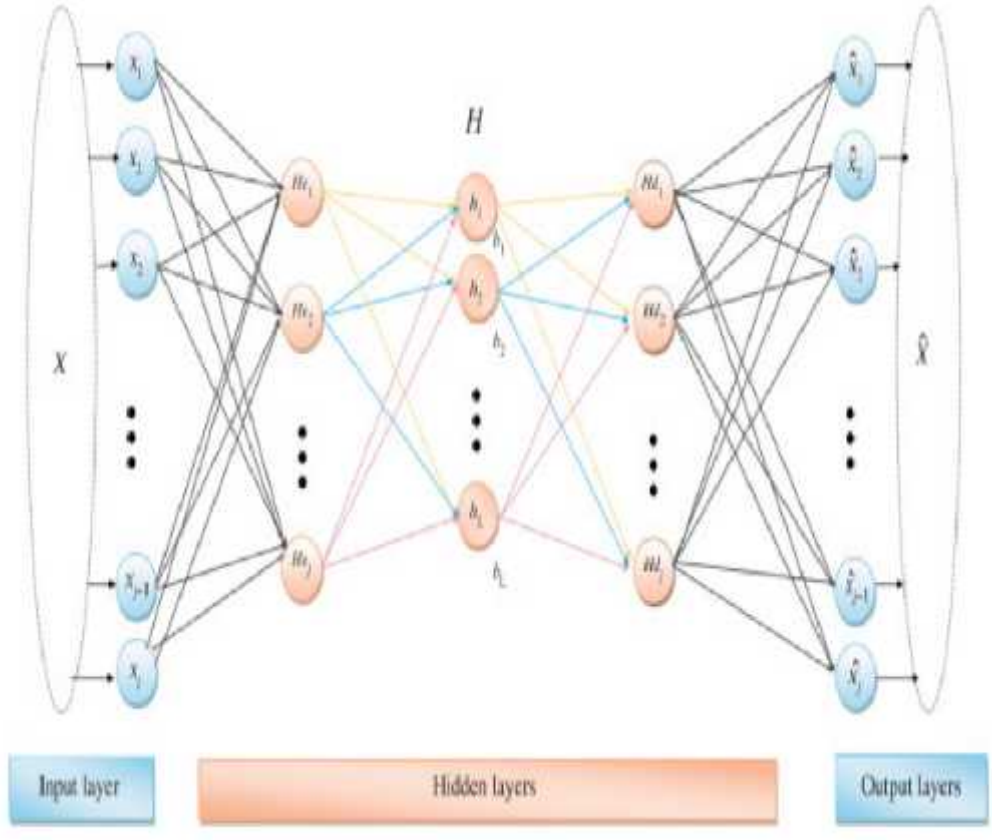

Figure 4.2: ELM Layers

Chapter 5

Results and Discussions

5.1 Results and Discussions

The user interface for the project execution is shown in the following screen figure 5.1. You may use the buttons to upload datasets, to execute algorithms, to pre-process data, and to do accuracy assessments.

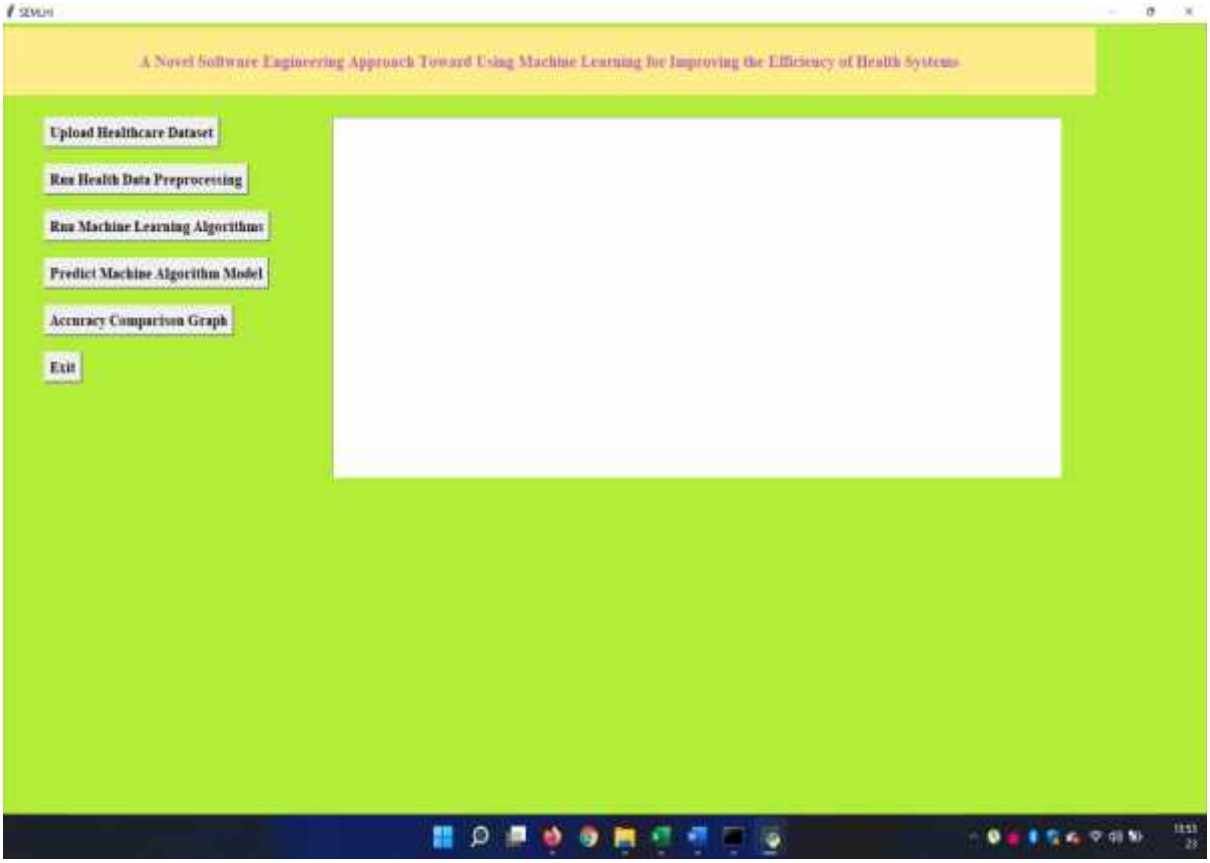

Figure 5.1: Graphical User Interface

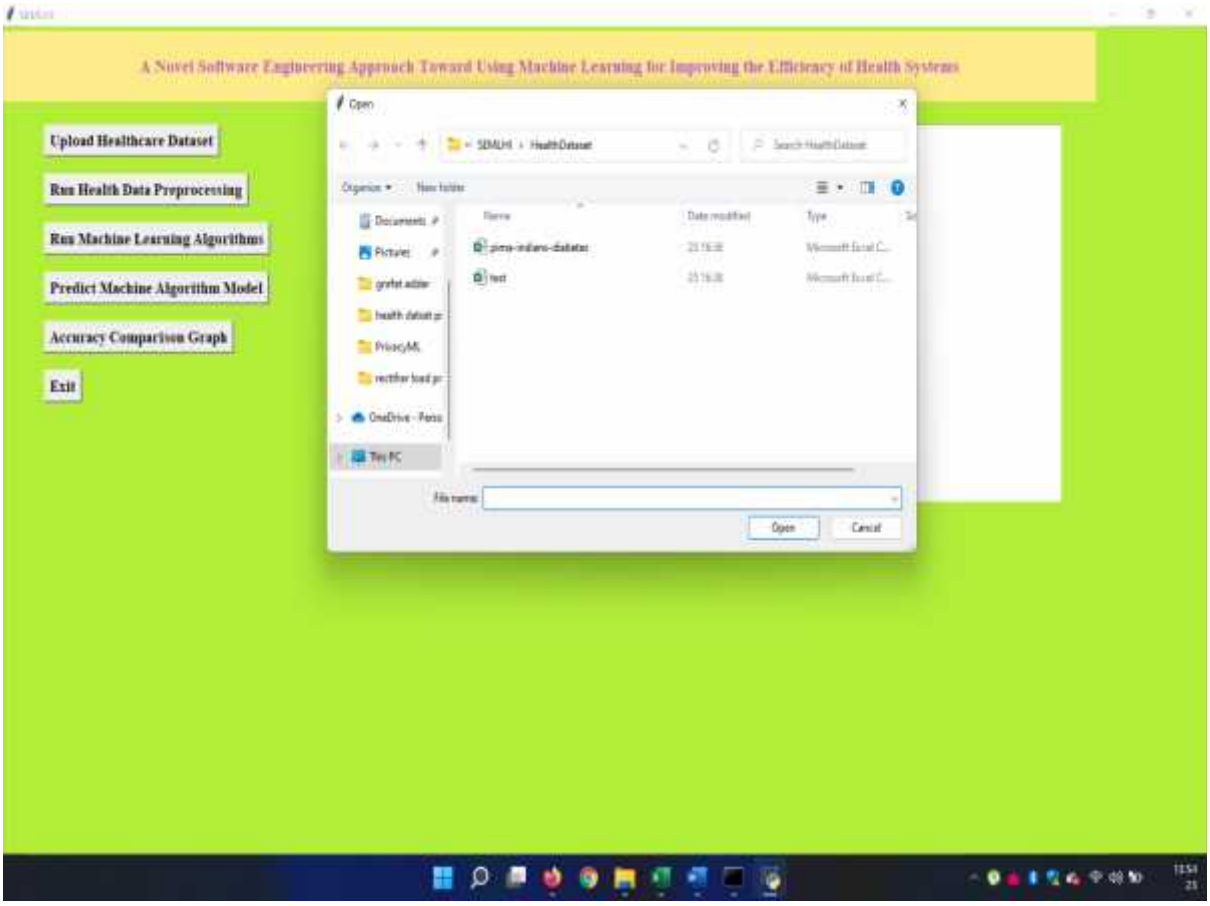

Figure 5.2: Dataset Loading Screen

Depending on the platform, a dataset can be read in either excel or csv format after it is published, as shown in screen picture 5.2. The message, that indicates that the dataset has been loaded, is displayed on the screen below. Loading is the process of storing data in a variable of the array type.

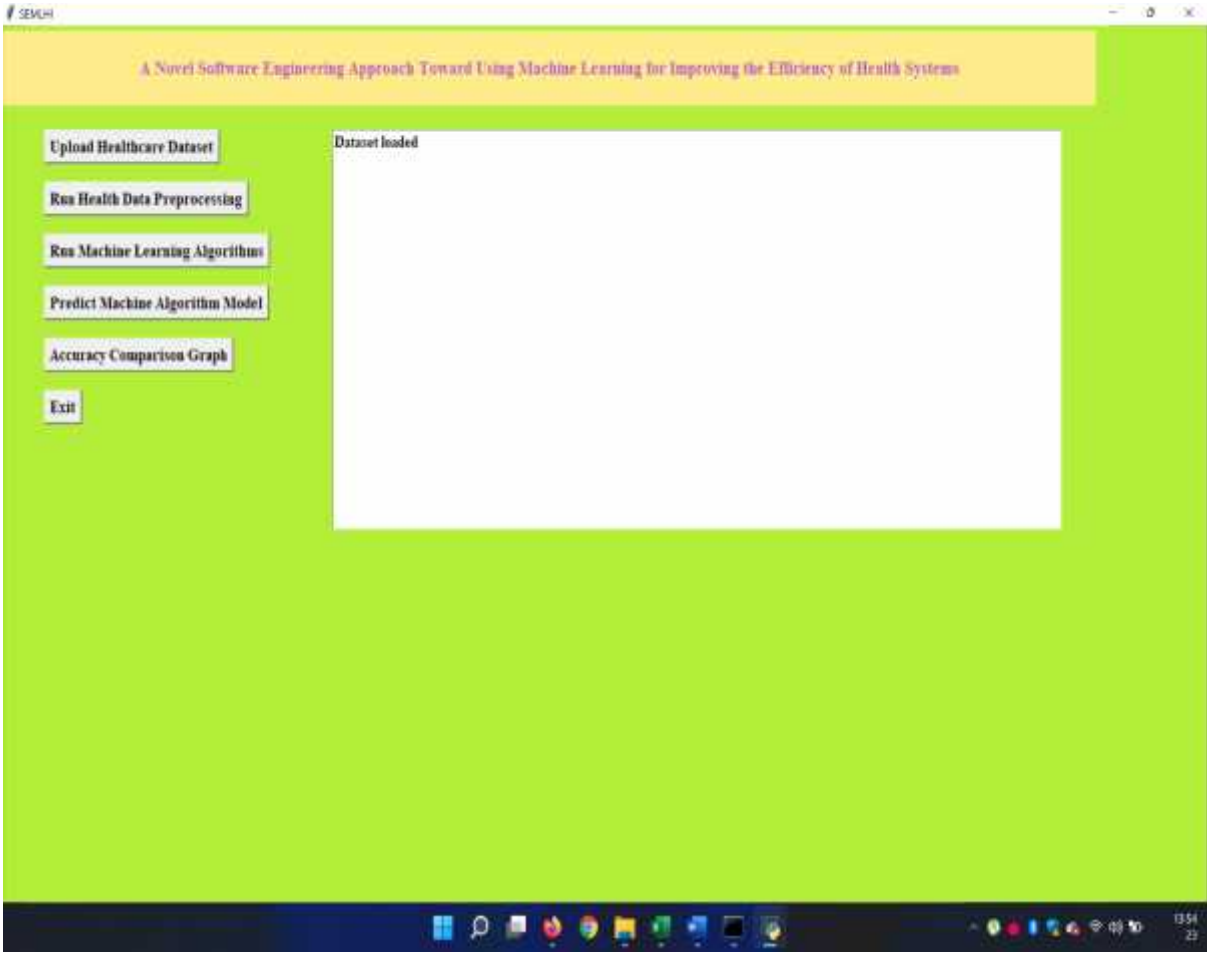

Figure 5.3: Dataset is Loaded

The dataset is loaded on the above page, and the dataset is shown in the following screen, where the final label 'Class' has values of 0 &1, where 0 indicates the non-appearance of disease & 1 indicates the appearance of illness in lab values.

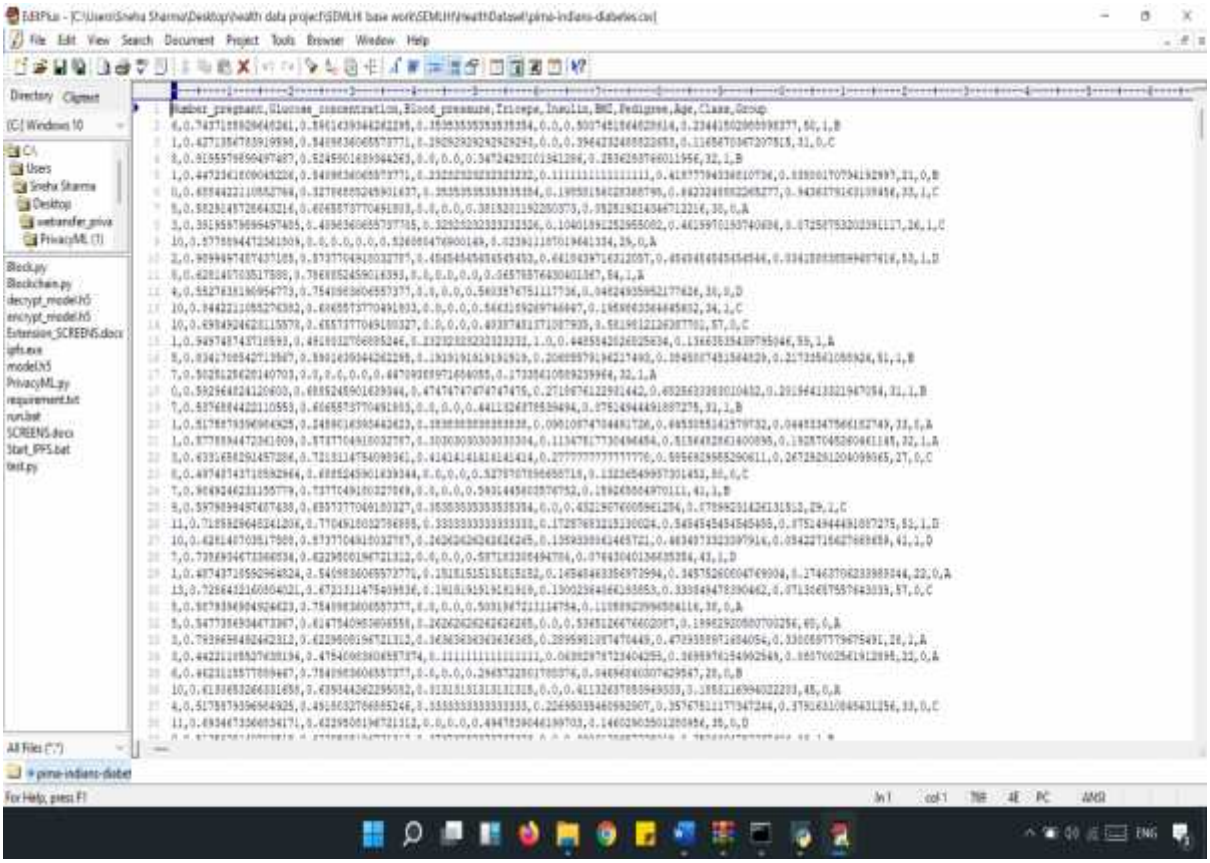

Figure 5.4: View of Dataset

ML method will train using the above lab report values and the Class Value (which includes 0 or 1), and then construct a model using the above lab report values and the Class Value (which contains 0 or 1), as shown in the above dataset screen. Created a train model and applied it to the test data below to predict the class label. For example, in the following example dataset, there is no Class label field, thus ML will guess it based solely on the lab results. Take a look at the test values shown below.

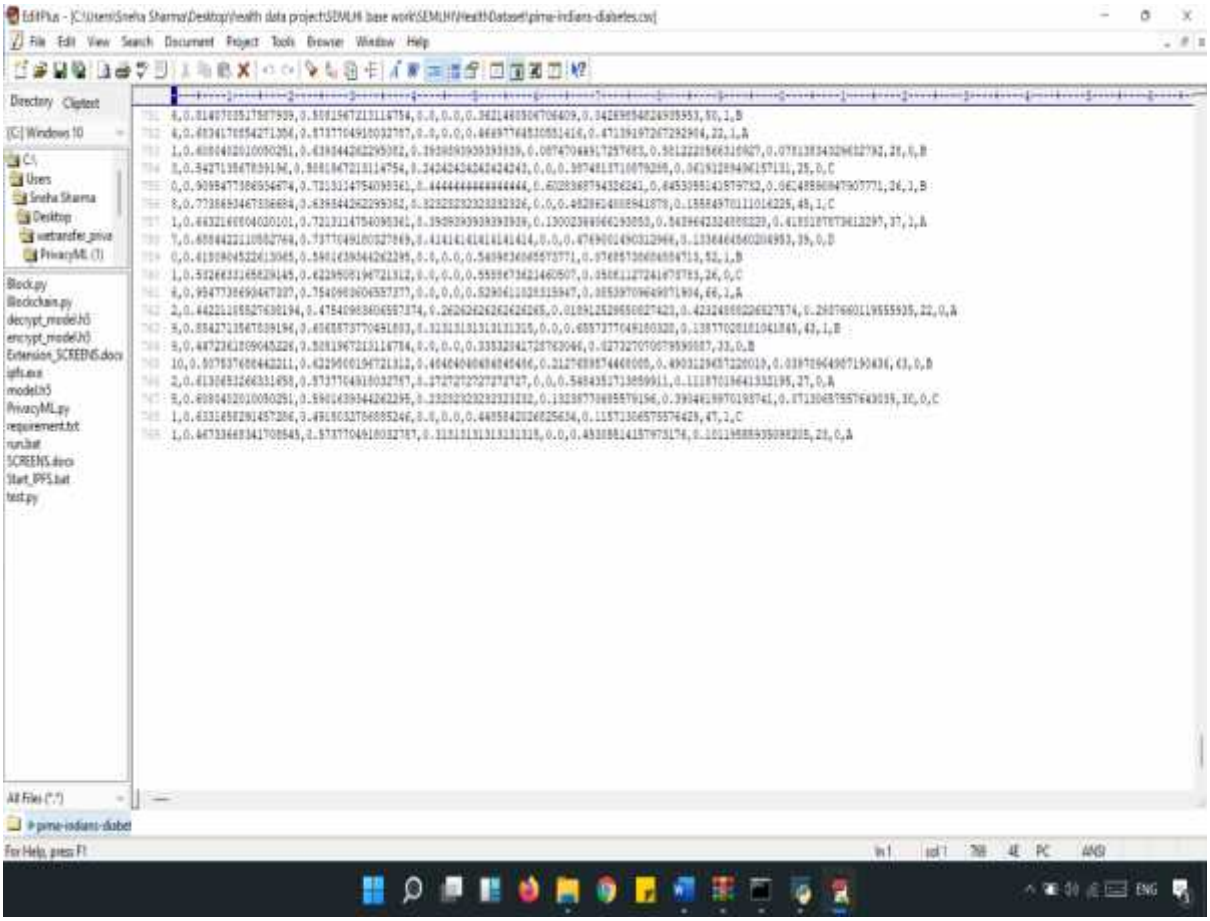

Figure 5.5: Dataset test used

There is no Class designation in the test lab report data shown above. Pre-processing, which involves applying PCA to the dataset, is the next step, as shown in the accompanying screen. Principal component analysis (PCA) is a statistical technique that transforms a set of correlated variables into a set of uncorrelated variables using an orthogonal transformation. PCA is an abbreviation for principal component analysis. In machine learning, PCA is the most commonly used method for exploratory data analysis and visualisation, and it is also the most widely used tool in data science.

Figure 5.6: Principal Component Analysis Output

The names of the columns are shown at the top of the graph, and positive column values are the only ones that matter, and the ML system will only train with these. Pre-processing has been done on the data from the tests and trains & the records have been tallied on the screen below.

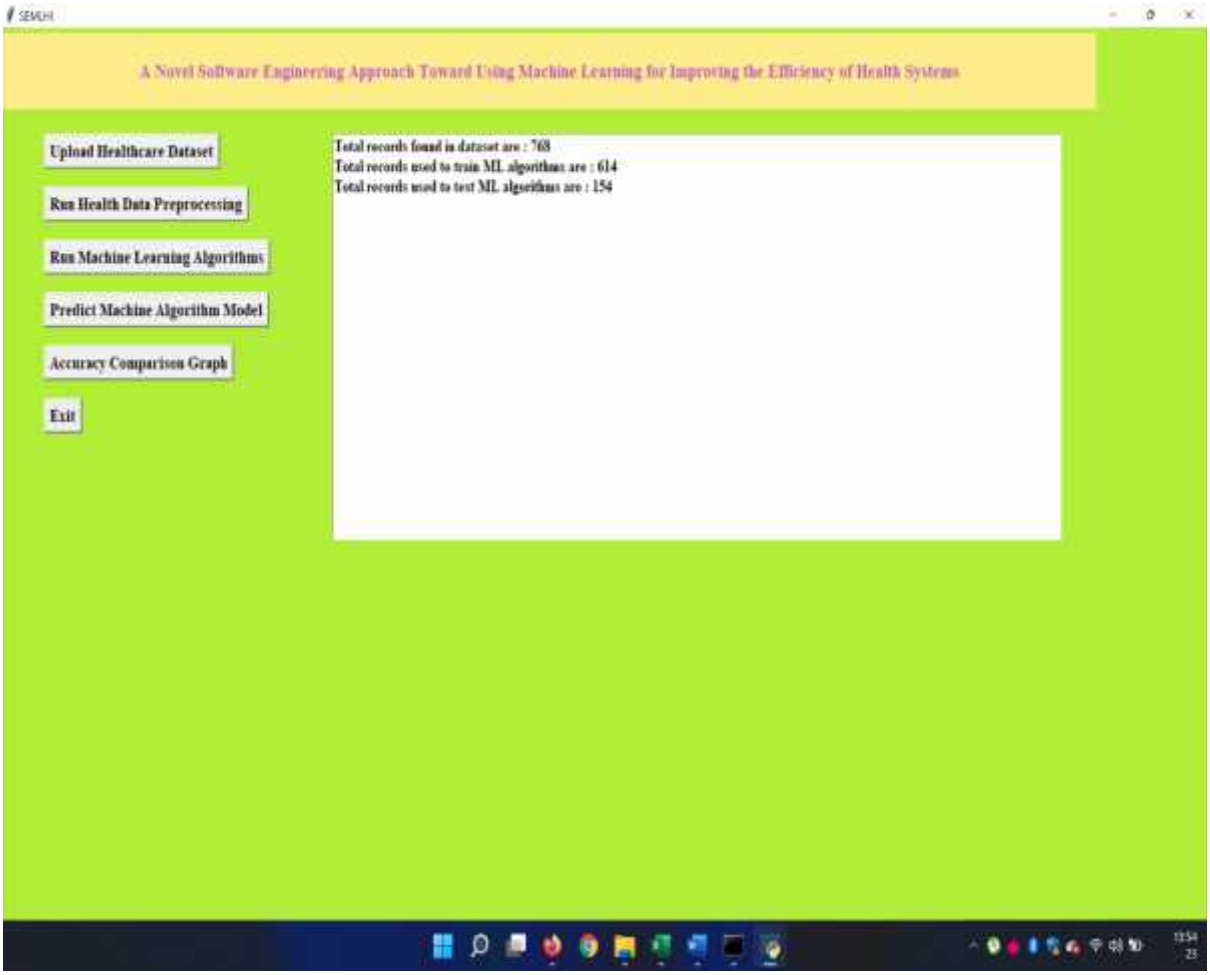

Figure 5.7: Storing of records

According to the screen above, a total of 768 records were obtained after pre-processing and PCA, and the application used 614 records to train machine learning algorithms & produce models, the trained model was then tested with 154 data to measure prediction accuracy. After collecting both test data & train, the following phase is to use machine learning techniques that were discussed in the previous semester's mid-semester implementation subject. It will compute accuracy, which will be shown in the right-hand column as seen below.

The following are the algorithms that have been implemented:

The KNN (K-Nearest Neighbor) approach is one of the key basic ML algorithms available. It is rooted on supervised learning method. Using the assumption that the latest instance/data & previous instances are equivalent, K-NN algorithm reserves a latest instance to the classification which is bulk homogenous to the current ones. This technique saves all available information & categorises each new data point based on its similarity to the previous data point using the K-NN algorithm. As a result, when fresh data is generated, it may be swiftly

sorted into an appropriate category utilising the K-NN algorithm and a simple algorithmic approach.

The method of categorization known as Naive Bayes is based on Bayes' Theorem and is based on the notion that predictors are unrelated to one another. To put it another way, a Naive Bayes classifier claims that the existence of one feature in a class has no bearing on presence of any other feature.

Random forests are a type of machine learning technique for solving problems that involve both regression and classification, can be used to address regression and classification problems. Ensemble learning is a strategy for tackling complicated situations to be successful, it necessitates the cooperation of a large number of classifiers.

Logistic regression is a statistical classification method that uses supervised learning to estimate the likelihood of a target variable occurring. It is required to collect a set of independent variables before anticipating a categorical dependent variable. When dealing with categorical dependent variables, logistic regression is employed to forecast the dependent variable's outcome. As a result, in order to be useful, the output must be categorical or discrete. Instead of delivering actual number, it sends probability values ranging from 0 to 1, if the value is boolean or numeric or in the middle.

It's a nonparametric clustering method that doesn't assume anything about the size or structure of the data clusters. A type of SVC is linear SVC. Because low-dimensional data is ideal, high-dimensional data may demand pre-processing methods such as principal component analysis. The extreme machine learning algorithm will now influence all algorithms that employ it.

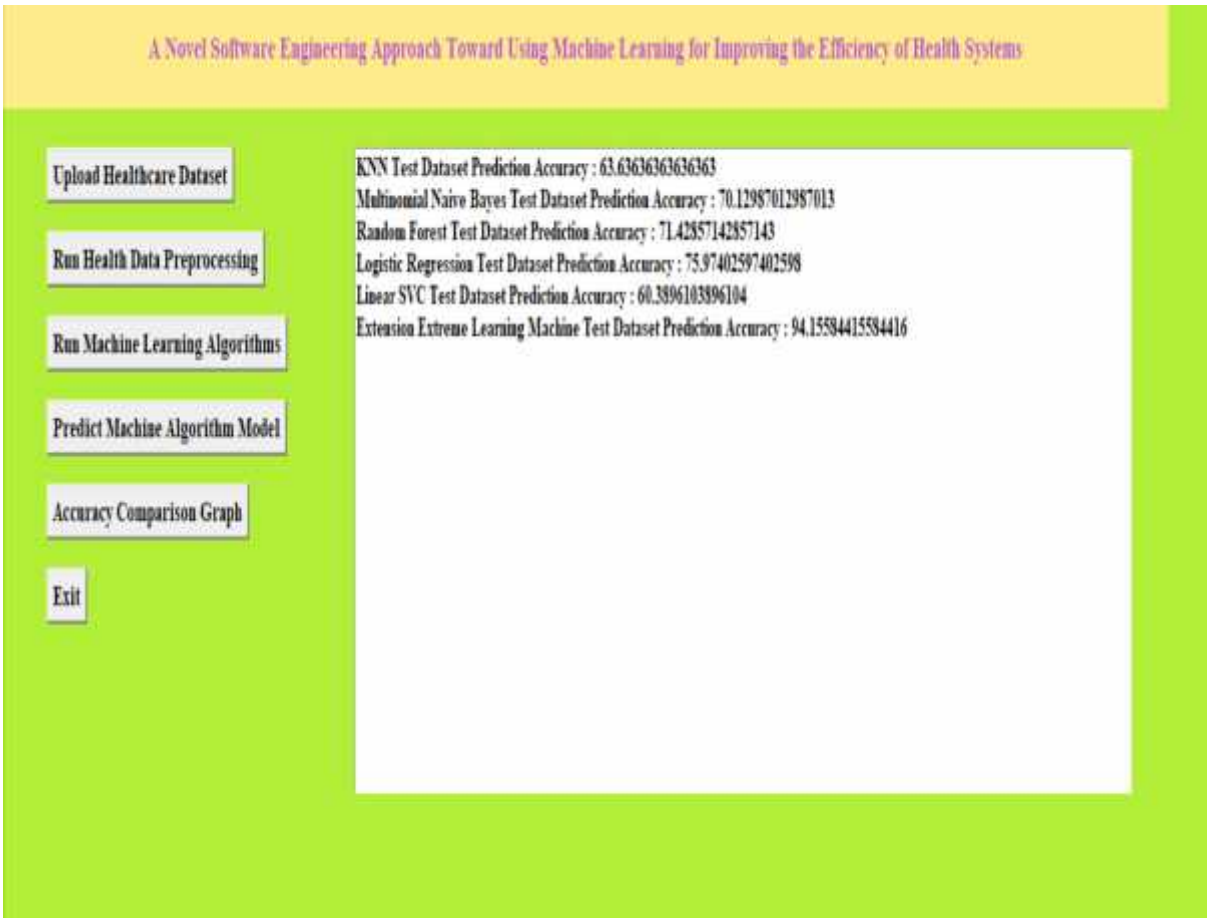

Figure 5.8: Machine Learning Algorithms

The extreme machine learning result is seen in the preceding screen, and it demonstrates that the machine learning algorithm has the maximum % of accuracy available to it.

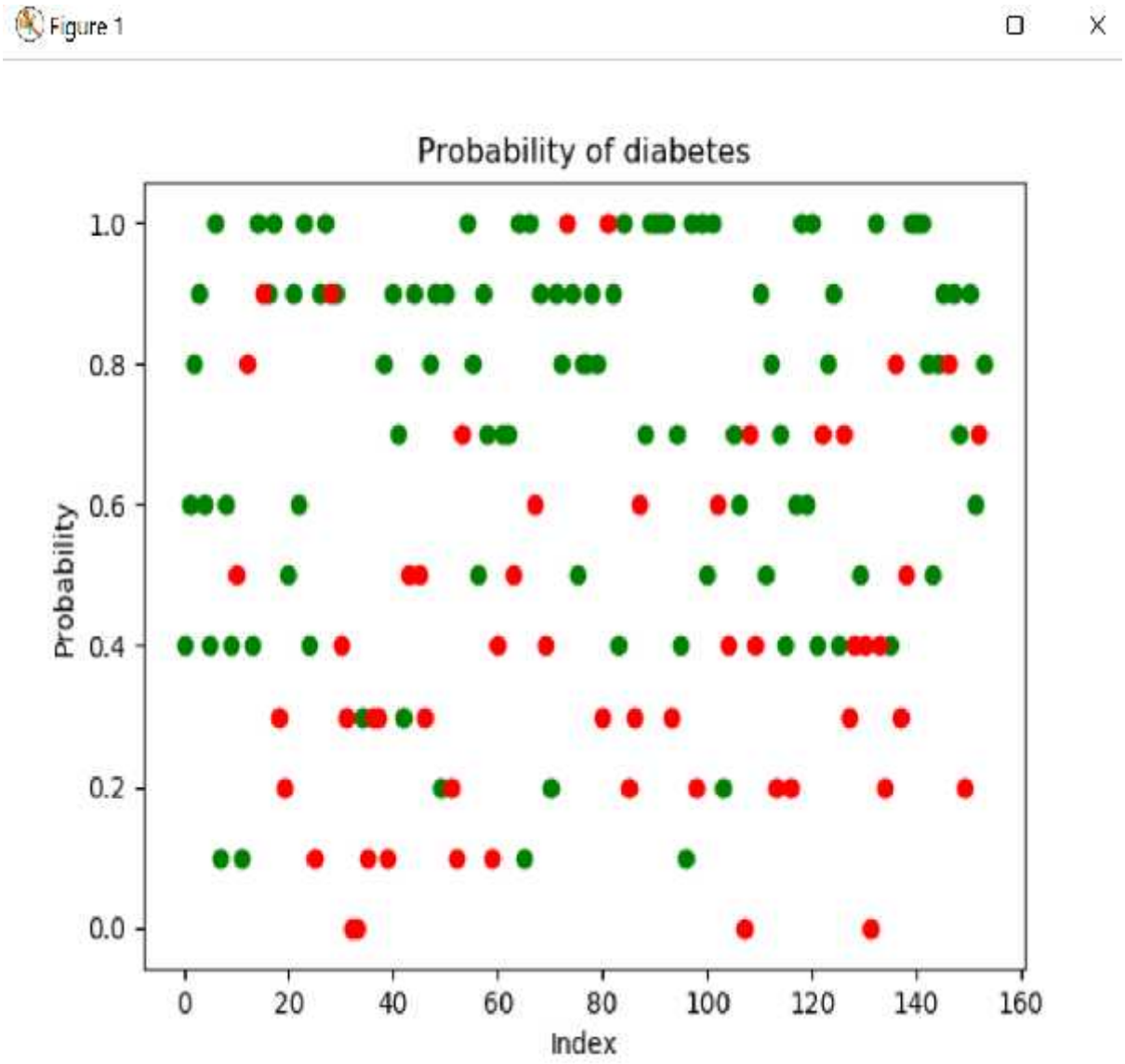

Figure 5.9: Chart for Probability

The likelihood of developing diabetes is shown in the above screen, with red indicating a positive forecast and green indicating a negative prediction.

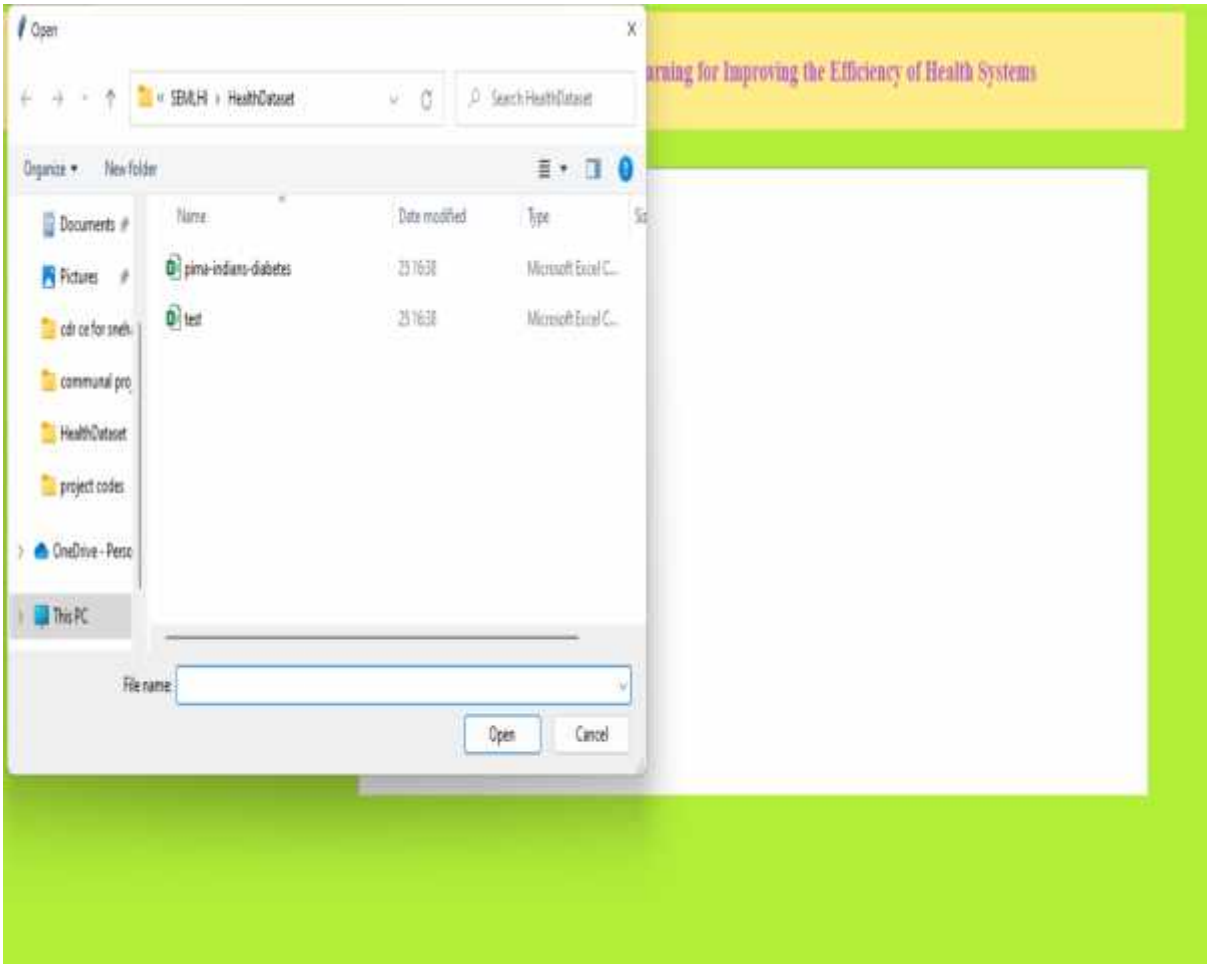

Figure 5.10: Test Data Loading

When the test method is executed, the screen shown above appears. It is necessary to input the test data into the test procedures.

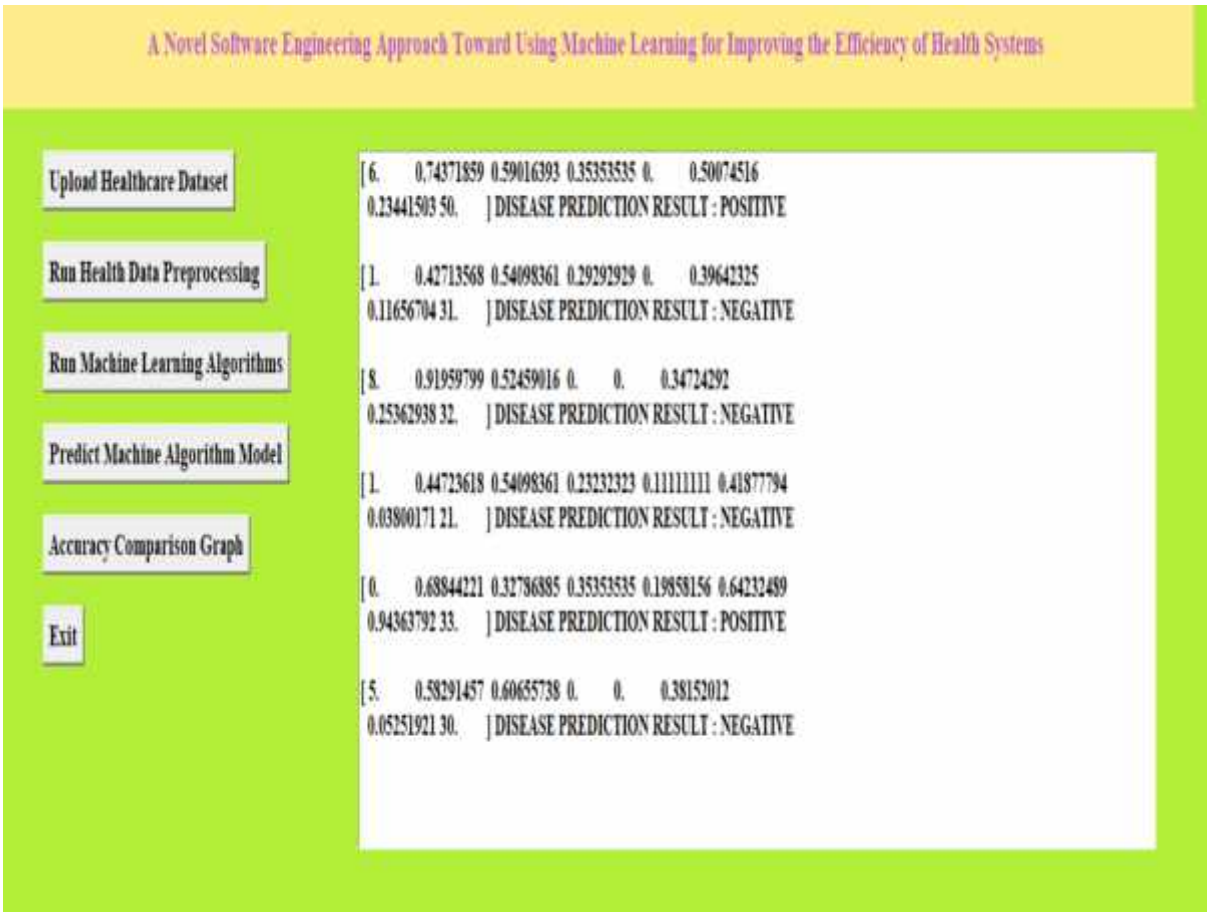

Figure 5.11: Test Result Output

The results of the diabetes positive and negative prediction results for the test loaded data are shown on the screen above.

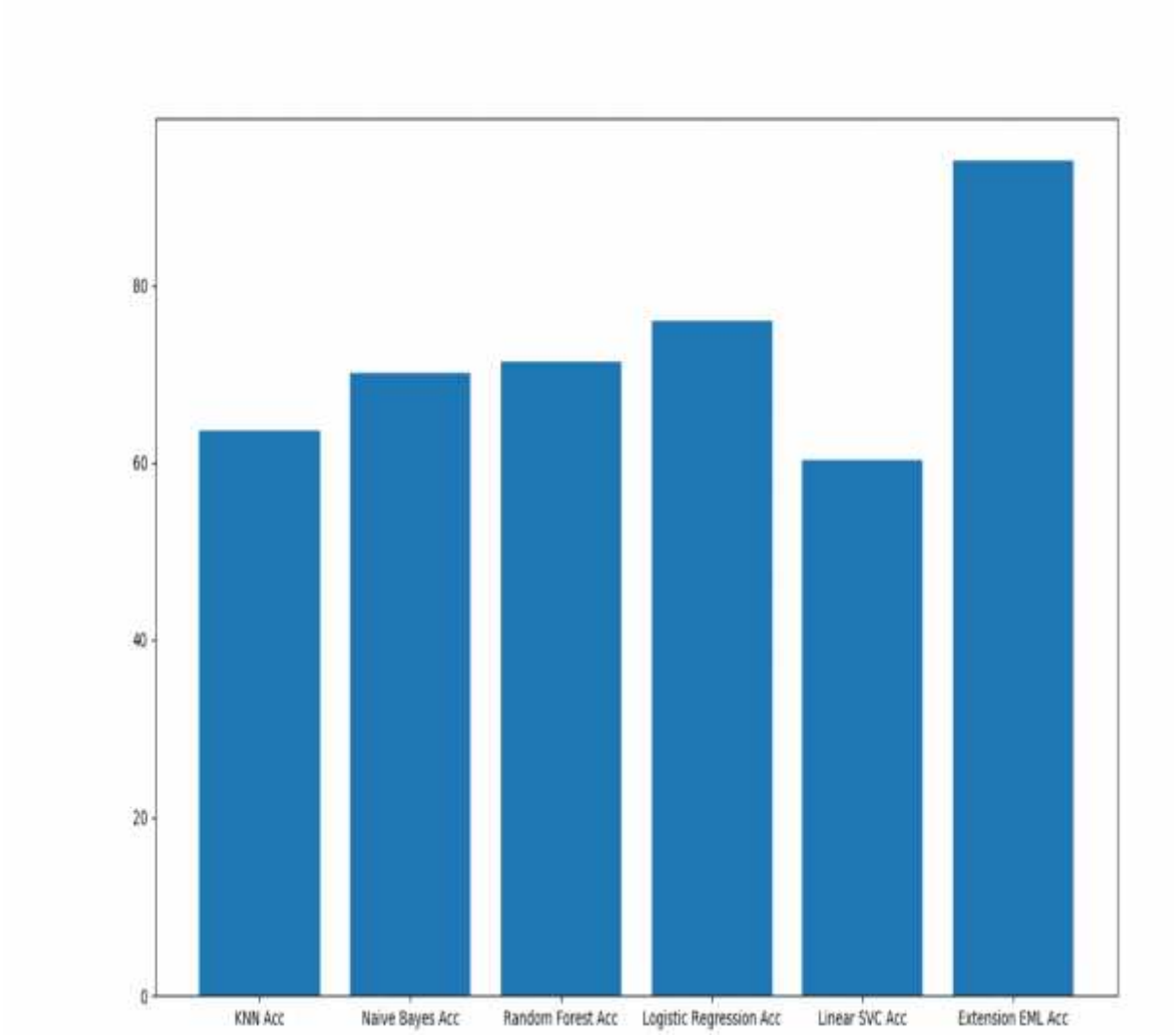

Figure 5.12: Output Comparison for Accuracy

The following diagram compares the extreme learning algorithm to a variety of other algorithms. On the hospital dataset, it shows that the extreme learning algorithm performs the best for polygenic disease divination.

Chapter 6
Conclusion

6.1 Conclusion

In order to improve the accuracy of diabetes prediction, in this paper, an extreme machine learning technique is utilized to a health dataset. Among other things, in batch learning, sequential learning & incremental learning applications, the extreme learning machine (ELM) is widely used, due to its quick & efficient learning pace, rapid convergence, high generalisation ability & ease of implementation. In terms of performance and efficiency. The extreme learning machine (ELM) is a highly efficient and effective learning approach for single hidden layer feed forward neural networks. When compared to other traditional neural network approaches, It offers the advantage of being less prone to over-fitting issues & requiring less time to become adept. According to the empirical risk minimization hypothesis, only one iteration of the learning process is required for ELM to be successful. When finding the minimum, the technique avoids the requirement for local minimization and several iterations, which would otherwise be required. Because of its greater generality, durability & controllability, as well as its high rate of learning, it has been used in a wide reach of industries & applications. Throughout the evolution of classic ELM, ELM algorithms that are more effective have been developed in a number of ways. Furthermore, the traditional ELM's field of applications has been expanded to encompass semi-supervised learning and unsupervised learning, among other things, in addition to supervised learning.

6.2 Future Work

Automation systems employed in a wide range of technologies will have higher processing capabilities in the future as a result of the future application of Extreme Machine Learning. The dataset may be expanded to include a variety of conditions such as heat-related disorders, lung concerns, & covid -19 prognosis sets.

Testimonials

[1] M. Moreb, T. A. Mohammed and O. Bayat, "A Novel Software Engineering Approach Toward Using Machine Learning for Improving the Efficiency of Health Systems," in IEEE Access, vol. 8, pp. 23169-23178, 2020, doi: 10.1109/ACCESS.2020.2970178.

[2] Beam, Andrew & Kohane, Isaac. (2018). Big Data and Machine Learning in Health Care. JAMA. 319. 10.1001/jama.2017.18391.

[3] Ferdous, Munira & Debnath, Jui & Chakraborty, Narayan. (2020). Machine Learning Algorithms in Healthcare: A Literature Survey. 1-6. 10.1109/ICCCNT49239.2020.9225642.

[4] Rahman, Atta & Sultan, Kiran & Naseer, Iftikhar & Majeed, Rizwan & Musleh, Dhiaa & Gollapalli, Mohammed & Chaabani, Sghaier & Ibrahim, Nehad & Siddiqui, Shahan & Khan, Muhammad. (2021). Supervised Machine Learning-based Prediction of COVID-19. Computers, Materials and Continua. 69. 21-34. 10.32604/cmc.2021.013453.

[5] Patel, Yash. (2021). Machine Learning in HealthCare.

[6] Shams, Mahmoud & Elzeki, Omar & Abdelfatah, Mohamed & Abou El-Magd, Lobna & Darwish, Ashraf & Hassanien, Aboul Ella. (2021). Impact of COVID-19 Pandemic on Diet Prediction and Patient Health Based on Support Vector Machine. 10.1007/978-3-030-69717-4_7.

[7] Harikumar & Mustafa, Malik & Sanchez, D. & Sajja, Guna & Gour, Sanjeev & Naved, Mohd & Jawarneh, Malik. (2021). IMPACT OF MACHINE learning ON Management, healthcare AND AGRICULTURE. Materials Today: Proceedings. 10.1016/j.matpr.2021.07.042.

[8] Mienye, Domor & Sun, Yanxia & Wang, Zenghui. (2021). Improved Machine Learning Algorithms with Application to Medical Diagnosis.

[9] Fetaji, Bekim & Fetaji, M. & Ebibi, Mirlinda & Ali, Maaruf. (2021). Predicting Diabetes Using Diabetes Datasets and Machine Learning Algorithms: Comparison and Analysis. 10.1007/978-3-030-90016-8_13.

[10] Khan, Bilal & Naseem, Rashid & Shah, Muhammad Arif & Wakil, Karzan & Khan, Atif & Uddin, M. Irfan & Mahmoud, Marwan. (2021). Software Defect Prediction for Healthcare

Big Data: An Empirical Evaluation of Machine Learning Techniques. Journal of Healthcare Engineering. 2021. 1-16. 10.1155/2021/8899263.

[11] Lalmuanawma, Samuel & Hussain, Jamal. (2020). Applications of Machine Learning and Artificial Intelligence for Covid-19 (SARS-CoV-2) pandemic: A review. Chaos Solitons & Fractals. 10.1016/j.chaos.2020.110059.

[12] Khan, Bilal & Naseem, Rashid & Shah, Muhammad Arif & Wakil, Karzan & Khan, Atif & Uddin, M. Irfan & Mahmoud, Marwan. (2021). Software Defect Prediction for Healthcare Big Data: An Empirical Evaluation of Machine Learning Techniques. Journal of Healthcare Engineering. 2021. 1-16. 10.1155/2021/8899263.

[13] Scott, Ian & Carter, Stacey & Coiera, Enrico. (2021). Clinician checklist for assessing suitability of machine learning applications in healthcare. BMJ Health & Care Informatics. 28. e100251. 10.1136/bmjhci-2020-100251.

[14] Mijwil, Maad & Salem, Israa & Abttan, Rana. (2021). Utilisation of Machine Learning Techniques in Testing and Training of Different Medical Datasets. Asian Journal of Computer and Information Systems. 9. 29-34. 10.24203/ajcis.v9i4.6765.

[15] Zhao, Rui & Yan, Ruqiang & Chen, Zhenghua & Mao, Kezhi & Wang, Peng & Gao, Robert. (2019). Deep learning and its applications to machine health monitoring. Mechanical Systems and Signal Processing. 115. 10.1016/j.ymssp.2018.05.050.

[16] Kumar, Naresh & Das, Nripendra & Gupta, Deepali & Gupta, Kamali & Bindra, Jatin. (2021). Efficient Automated Disease Diagnosis Using Machine Learning Models. Journal of Healthcare Engineering. 2021. 1-13. 10.1155/2021/9983652.

[17] Arvanitis, Theodoros & White, Sean & Harrison, Stuart & Chaplin, Rupert & Despotou, George. (2021). A Method for Machine Learning Generation of Realistic Synthetic Datasets for Validating Healthcare Applications. 10.1101/2021.02.11.21250741.

[18] Kaissis, Georgios & Ziller, Alexander & Passerat-Palmbach, Jonathan & Ryffel, Théo & Usynin, Dmitrii & Trask, Andrew & Lima, Ionésio & Mancuso, Jason & Jungmann, Friederike & Steinborn, Marc-Matthias & Saleh, Andreas & Makowski, Marcus & Rueckert, Daniel & Braren, Rickmer. (2021). End-to-end privacy preserving deep learning on multi-institutional medical imaging. Nature Machine Intelligence. 3. 1-12. 10.1038/s42256-021-00337-8.

[19] Ojo, Adebola & Olanrewaju, Ahmed. (2019). Knowledge Discovery in Medical Database using Machine Learning Techniques. International Journal of Computer Applications. 178. 14-21.

[20] Rahman, Mohammed & Akhter, Nahida & Saleh, Sabbir. (2017). Health Data Mining: Machine Learning Approach. 3. 89-96.

[21] Toh, Christopher & Brody, James. (2021). Applications of Machine Learning in Healthcare. 10.5772/intechopen.92297.

[22] Akhil, Jabbar & Samreen, Shirina & Aluvalu, Rajanikanth. (2018). The Future of Health care: Machine Learning. International Journal of Engineering and Technology(UAE). 7. 23-25. 10.14419/ijet.v7i4.6.20226.

[23] Manne, Ravi & Kantheti, Sneha. (2021). Application of Artificial Intelligence in Healthcare: Chances and Challenges. Current Journal of Applied Science and Technology. 40. 78-89. 10.9734/CJAST/2021/v40i631320.

[24] Gavhane, Aditi & Kokkula, Gouthami & Pandya, Isha & Devadkar, Prof. (2018). Prediction of Heart Disease Using Machine Learning. 1275-1278. 10.1109/ICECA.2018.8474922.

[25] Debnath, Shubham & Barnaby, Douglas & Coppa, Kevin & Makhnevich, Alexander & Kim, Eun Ji & Chatterjee, Saurav & Tóth, Viktor & Levy, Todd & Paradis, Marc & Cohen, Stuart & Hirsch, Jamie & Zanos, Theodoros & Becker, Lance & Cookingham, Jennifer & Davidson, Karina & Dominello, Andrew & Falzon, Louise & Mcginn, Thomas & Mogavero, Jazmin & Osorio, Gabrielle. (2020). Machine learning to assist clinical decision-making during the COVID-19 pandemic. Bioelectronic Medicine. 6. 14. 10.1186/s42234-020-00050-8.

[26] Kovačević, Živorad & Gurbeta Pokvic, Lejla & Spahić, Lemana & Badnjevic, Almir. (2019). Prediction of medical device performance using machine learning techniques: infant incubator case study. Health and Technology. 10. 10.1007/s12553-019-00386-5.

[27] Rani, KM. (2020). Diabetes Prediction Using Machine Learning. International Journal of Scientific Research in Computer Science Engineering and Information Technology. 294-305. 10.32628/CSEIT206463.

[28] Thenabadu, Thamindu & Ilmini, Kalani. (2020). Diabetes Prediction System using Machine Learning.

[29] Baig, Muhammad & Nadeem, Muhammad. (2020). Diabetes prediction using machine learning algorithms. 10.13140/RG.2.2.18158.64328.

[30] Lyngdoh, Arwatki & Choudhury, Nurul & Moulik, Soumen. (2021). Diabetes Disease Prediction Using Machine Learning Algorithms. 10.1109/IECBES48179.2021.9398759.