

The Applications of Machine Learning in Healthcare

Semester: Fall 2024/2025

**Done By: Jad Majed Soubra id:12031634**

**Instructor: Samira Klaylat**

# Abstract

The early disease prediction mechanism is a key to reducing the cost of treatment and improving the speed and efficiency of diagnostics in the medical field. The following paper, addresses the implementation and evaluation of the naïve bayes machine learning probability based algorithm and the random forest classifier and comparing both models in the end. And this study focuses on the relationship of different key factors including, age, cholesterol level, chest pain type, resting blood pressure, and many other features with the diagnosis of a heart disease as the target variable. To implement this experiment, the use of a dataset obtained from kaggle was used in both naïve bayes and random forest classifiers which are python based models for predicting the diagnosis of a heart disease. Also a survey was conducted that focuses on the social part of this study that aims to understand how the outside world thinks when it comes to the use of machine learning models for disease prediction in healthcare, which areas it’s used in, and the advantages and disadvantages of using these models. The Naïve bayes classifier achieved an 86% accuracy rate which is less than that of the random forest classifier at 83% with the performance metrics validating a good diagnosis. The results stated that the naïve bayes classifier is really effective when small datasets are used. However, the good predictions can put that model under the risk of overfitting that limits the use of the naïve bayes classifier encouraging the use of larger datasets and other models that work well for complex datasets. Also survey results indicated the importance of machine learning’s use in healthcare in general especially when it comes to early disease predictions and how crucial the data privacy and security problems can be therefore also indicating the need to address those issues in future studies. The research study done demonstrates the efficiency of the use of ml models in early disease prediction and shows the need to work on more complex and real-life datasets to ensure the minimization of overfitting and to also address the privacy and security issues that those models face.

Table of Contents

[Abstract 1](#_Toc186890536)

[Table of figures 4](#_Toc186890537)

[Acknowledgement 6](#_Toc186890538)

[1.INTRODUCTION 7](#_Toc186890539)

[1.1 Artificial Intelligence: 9](#_Toc186890540)

[1.2. Machine Learning: 10](#_Toc186890541)

[1.2.1. Supervised Learning: 10](#_Toc186890542)

[1.2.2 Unsupervised Learning: 11](#_Toc186890543)

[1.2.3 Reinforcement learning: 12](#_Toc186890544)

[1.3 Deep Learning: 13](#_Toc186890545)

[Background and literature review: 14](#_Toc186890546)

[2.1 Medical Diagnostics using machine learning: 14](#_Toc186890547)

[2.2 Patient Monitoring Using Predictive Analysis: 15](#_Toc186890548)

[2.3. Personalized Medicine and Treatment : 16](#_Toc186890549)

[3.Methodology 17](#_Toc186890550)

[3.1. Studies and Benchmarks 17](#_Toc186890551)

[**3.1.1 Data Dictionary:** 20](#_Toc186890552)

[**3.1.2 conducting the experiment :** 21](#_Toc186890553)

[3.2 survey study 23](#_Toc186890554)

[4.Results 24](#_Toc186890555)

[4.1: Model experiment results: 25](#_Toc186890556)

[4.2: survey results: 28](#_Toc186890557)

[4.2.1 Demographic findings : 29](#_Toc186890558)

[4.2.2 Machine learning focused questions: 30](#_Toc186890559)

[5.Discussion 34](#_Toc186890560)

[5.1 plot analysis: 35](#_Toc186890561)

[**5.1.1 insights and analysis:** 35](#_Toc186890562)

[**5.1.2 Obstacles and limitations:** 35](#_Toc186890563)

[**5.2 survey analysis:** 36](#_Toc186890564)

[**5.2.1: audience and perceptions:** 36](#_Toc186890565)

[ **5.2.2: limitations** 37](#_Toc186890566)

[Conclusion 37](#_Toc186890567)

[References 39](#_Toc186890568)

# Table of figures

**Figure 1*.*1** illustrating concepts of AI, ML, Deep learning…………………………… 9

**Figure 1*.*2** the supervised learning cycle ………………...…………………………… 11

**Figure 1*.*3** the workflow of unsupervised learning …………………………………… 12

**Figure 1*.*4** the workflow of reinforcement learning ………………………………….. 13

**Figure 3*.*1** publications by year in using the machine learning early disease detection approach …………………………… …………………………………………………….18

**Figure *3.*2** Top ten countries that contributed to the machine learning early disease detection literature. ……………………………………………………………………… 19

**Table 3.3** this Table describes the dataset with 303 entries and 14 columns with target as the target variable ……………………………………….…………………………… 20

**Figure 3.4:** importing the necessary python libraries …………………………………22

**Figure 3.5** python code for basic data exploration.……………………………………22

**Figure 3.6** creating the predictions for both models and evaluating their performance ……………………………. .……………………………………………………………… 23 **Figure 4.1:** The naïve bayes predictions *………………….*…………………………… 25

**Figure 4.2:** the confusion matrix of the naïve bayes model……………………………26

**Figure 4.3:** the random forest predictions…………………………… ………………… 27

**Figure 4.4** : the confusion matrix of the random forest classifier…………….……… 27

**Figure 4.5** : the accuracy score of both models………………………………………… **28**

**Figure 4.6:** the output that counted the points at 0 and 1 for both models...........….. 28

**Figure 4.7:** age groupof respondents……………………………………………………29

**Figure 4.8:** gender of respondents…………………………………………………………29

**Figure 4.9:** people with a medical condition diagnosis…………………………………...30

**Figure 4.10:** people with a medical background ………………………………………….30

**Figure 4.11 :** people being familiar with machine learning used in healthcare…………31

**Figure 4.12:** uses of machine learning in healthcare…………………………………......31

**Figure 4.13:** trusting a machine learning model to assist in diagnosing diseases……..32

**Figure 4.14:** improving early disease diagnosis…………………………………………32

**Figure 4.15:** anonymous data sharing to help ml models provide better predictions….33

**Figure 4.16:** the concerns for sharing data with ml models………………………………33

**Figure 4.17:** graph showing thebenefits of machine learning for disease prediction...34

**Figure 5.1:** the formula for the accuracy score…………………………………………….35

# Acknowledgement

Throughout the process of writing this paper I have received numerous forms of support and assistance.

I would like to express my gratitude for my instructor Dr. Samira Klaylat who assisted me with her expertise guiding me through the process of writing this paper and being responsive when it comes to her assistance in answering questions I had for this research. Thanks to you I had a brighter vision to complete the work I was assigned to do.

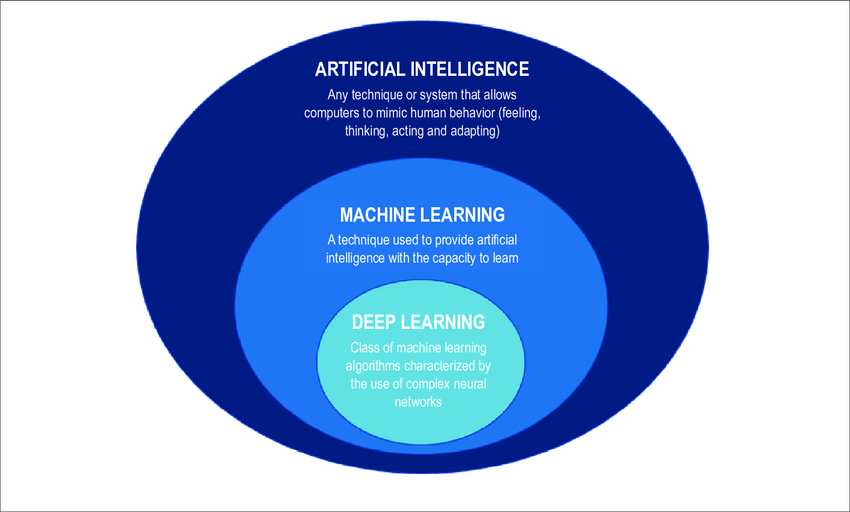
I would like to also acknowledge my siblings who also provided me with assistance as they are experienced in research. Your tips assisted me in more ways that I could count.

I would like to thank my parents for their moral support in writing this research. You were always there for me no matter what during my research providing me with the positive and motivating thoughts throughout the process.

# 1.INTRODUCTION

Nowadays, Artificial intelligence and machine learning which is considered a game-changing subset of AI, are changing the world in two simple but yet complex methods: AI is making today’s world easier by increasing the productivity of its users by completing complex tasks more efficiently and easily. However, there is also a dark side in that realm that lets people depict AI as villain robots that are ready to take over the world. As an example, “Devin AI, developed by Cognition Labs, represents a groundbreaking advancement in AI technology, designed to autonomously handle coding, web design, and even project management tasks. Its ability to generate code, debug existing applications, and learn from its own mistakes through machine learning algorithms promises to streamline software development significantly” (Hardika, Valencia, Adzani, Barus, & Fami, 2024). Despite Devin AI being a massive productivity booster and the fact that it “has the capacity of taking entire projects from concept to deployment thereby improving software development processes while freeing up human developers for strategic thinking and innovation. For example in the SWE-bench test 13.86% of GitHub issues were resolved correctly by Devin without any assistance given whatsoever, which was far much higher than GPT-4” (Durrani & Jain) , the major amount of people working in the technology field were afraid of losing their positions in the companies they work in because of the belief that this model will substitute current software engineers. On the bright side, what if Artificial intelligence can be used in a way to save lives, and to help treat millions of patients all around the world through making hospitals more productive by enhanced AI accurate decision making tools as “the introduction of AI into previously human-only workflows is motivated by this objective of improving decision accuracy by leveraging the complementary strengths of the human DM and the AI. At a minimum, we expect humans aided by AI to perform better (or at least not worse) than humans who make decisions unaided. Many studies have been able to achieve this benchmark primarily because they involved situations in which the human was offered AI advice by an AI that exhibited higher accuracy than could be produced through human-only performance.” (Steyvers & Kumar, 2024). Also AI assisted decision making isn’t only limited to the assistance of treating ill patients in medical centers it can also be used to aid the same process in flexible manufacturing systems for example, the (AI-1) expert system tool that “is built to support the initial FMSs design, i.e. the simulation models in this research project. The aim of the AI-1 is to ensure that the design objectives can be met. In attempting to determine simulation models that are efficient with respect to FMS design objectives, the AI-1 analyses the results of the simulation model, and uses a number of heuristics in much the same way a human designer would approach the problem. These are identified as: operational heuristics, which are used to assess production level considerations; economic heuristics, which consider the financial options, and social and human heuristics, which are used to access the number of operators and maintenance technicians and workload of workers, etc.” ( Chan, Jiang, & Tang). This was a brief example to prove the broad and various uses of AI based decision making yet this is only considered as the tip of the iceberg. When it comes to the medical field, “These decisions have a direct impact on patients' lives, on 40 caregivers, as well as on society at large. As one example, 250.000 Americans die from medical errors each year” ( Makary , M. A. Makary and M. Daniel, & Daniel, 2016). “A medical error costs hospitals on average 939$, making $1 billion costs for US hospitals alone” (G. David, 2013). “Additionally, the absence of evidence for clinicians as well as biased research were identified as reasons for errors in medical decisions” (V. Saini et al, July 2017). . According to the survey that was made in (J. Cortada, jan. 2021), “37% of healthcare organizations lack the data they need for their decision-making.” (J. Cortada, jan. 2021). In due time, in case high-quality datasets are provided with accurate insights, we see a high potential for AI-driven digital decision support systems (DDSS). The Cambridge Dictionary defines AI as “the study of how to produce computers that have some of the qualities of the human mind, such as the ability to understand language, recognize pictures, solve problems, and learn” (Artificial Intelligence, Press, Cambridge, 2014. Accessed: Feb. 15, 2021. [Online]. Available). To understand the meaning of (DDSS) in a more detailed manner, Sauter defines DDSSs as “computer-based systems that bring together information from various sources, that assist in the organization and analysis of information and facilitate the evaluation of assumptions underlying the use of specific models” (Sauter, 2017). In the medical field, DDSSs can enhance healthcare by improving certain decisions taken in the mentioned areas that lack the ability of having efficient medical care. Therefore, DDSSs have a deal-breaking role of high importance in making our places of habitat, regions, cities, settlements, and communities smart and sustainable. Ironically, instead of making the situation more bearable and better, DDSSs currently have a major role, therefore coming up with new decision making problems. “A useful and robust DDSS should draw upon existing state of the art clinical decision making processes which subsequently inform treatment allocation. Successful implementation and adoption into clinical practice requires that the DDSS is developed in partnership with the end user group through early stakeholder involvement”. (Fraser Philp, 2021).

DSSs can also play a high role when it comes to treating patients with psychological illness as a study done on AI and decision support systems shows that “52.7% of people with depression are not correctly diagnosed by their general practitioner. This evidence shows that we collect medical data but struggle to make use out of it. Digital Decision Support Systems (DDSS) and artificial intelligence (AI) could be one way to address diagnostic uncertainty by assisting medical professionals in making sense of data.” (Markus Bertl, 2020). “Mounting evidence suggests that demand for such systems is given. 56% of U.S. adults are willing to share their health data with tech companies like Google. The big data market for health data is booming” (Dash, 2019) and is estimated to reach 7 billion USD by 2021. However, considerable doubt exists. 85.9% of office-based physicians use electronic health records in U.S.” (Office-based Physician Electronic Health Record Adoption., 2019) A similar trend can be observed in the European Union (eHealth, 2019). Artificial Intelligence is basically categorized by the Cambridge Dictionary as “the study of how to produce computers that have some of the qualities of the human mind, such as the ability to understand language, recognize pictures, solve problems, and learn” (Artificial Intelligence, Press, Cambridge, 2014. Accessed: Feb. 15, 2021. [Online]. Available). Even though AI and machine learning could seem extremely similar, however their differences lay wide as machine learning is only a subset of artificial intelligence that includes machine learning and deep learning. To understand those differences more, we should be able to define each branch of AI in a more detailed manner.



**Figure 1*.*1** illustrating concepts of AI, ML, Deep learning

## Artificial Intelligence:

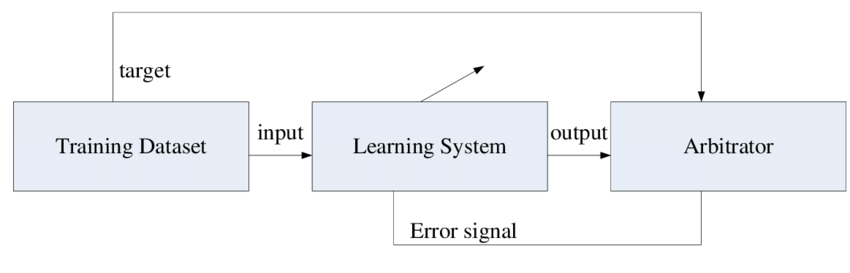
Artificial intelligence is a field in technology that simply studies and mimics human behavior by using several methods of perceptions that depend on the heuristics perceived by its sensors. “Aristotle attempted to formalize ‘right thinking’ (logic) through his syllogisms (a three-part deductive reasoning). Much of the work in the modern era was inspired by this and the early studies on the operation of mind helped to establish contemporary logical thinking. Programs which enable computers to function in the ways” (AN Ramesh, 2004). “The application of machine learning dates back to the 1950s when Alan Turing proposed the first machine that can learn and become artificially intelligent”. (Turing, 1950). Artificial Intelligence (AI) is the umbrella term that refers to any computerized intelligence that learns and imitates human intelligence. (Lee E.J., 2017).  Ever since the concepts of Artificial intelligence and machine learning were utilized in today’s world a lot of applications for these technologies also came to place. ranging from “security services through face detection” (Wati D.A.R., 2017) , to “increasing efficiency and decreasing risk in public transportation” (Ellis K., 2014) , and “recently in various aspects of healthcare and biotechnology” (Siddiqui M.K., 2020). AI has achieved massive success throughout its applications due to “their higher levels of decision-making, accuracy, problem-solving capability, and computational skills “ (Miotto R., 2018). Also the large scale organizations and medical facilities are starting to implement AI- based approaches to “achieve increased efficiency in the organization of electronic health records” (Rao S.R., 2011), “identification of irregularities in the blood samples” (Siddiqui M.K., 2020), bones (Tian L., 2020),and organs (Rajpurkar P, 2017),using medical imaging and monitoring, as well as in robot-assisted surgeries (Kaouk J.H., 2019). AI was mostly known for machines that follow the autopilot mechanism especially when AI was used in the manufacturing of specialized robots, self-driving vehicles. In addition to those inventions a new project that would change the lives of millions. “Where in July 2019, the American billionaire Elon Musk revealed the new objectives for his Startup “Neuralink” (“Neuralink Launch Event”): to develop a cerebral implant that will help an individual to control different technological devices, such as a computer, solely using the electrical activity of neurons.” (Fourneret, 2020).

## Machine Learning:

As displayed in **Fig.1.1**, Machine learning (ML) is basically a subset of the huge realm of AI yet with a more specific and complex set of algorithms. “In simple words learning means either acquiring new knowledge or enhancing or updating individual’s skills. Learning new knowledge is the combination of various processes such as acquisition of significant concepts, understanding their meanings and relationships to each other and to the area concerned. Skill enhancement can be interpreted in biological terms as reinforcing a pattern of neural connections for performing the desired function.” (Ku. Chhaya A. Khanzode, January-April 2020). Machine learning has 3 general and important methods of learning including: Supervised learning, Unsupervised learning, reinforcement learning.

## Supervised Learning:

“Supervised learning is the machine learning task of learning a function that maps an input to an output based on example input-output pairs. It infers a function from labelled training data consisting of a set of training examples. The supervised machine learning algorithms are those algorithms which needs external assistance. The input dataset is divided into train and test dataset” (Mahesh, 2018). “In training process, samples in training data are taken as input in which features are learned by learning algorithm or learner and build the learning model” (Sandhya N. dhage, 2016) .When the process is tested, the compilation engine is used by the trained model to set predictions for the test data. The below figure describes the process of the supervised learning workflow Several algorithms are used in this type of learning including linear regression which is “one of the most used algorithms developed to predict a quantitative variable using one or more independent variables with the assumption that the independent variables and the outcome variables have a linear regression with each other” (R. G. , May 28 2019 ). (SVM) is another one of its most valuable algorithms. “Support Vector Machines (SVM) are known as classification algorithms that use supervised learning to classify features in two group problems by finding the largest margin hyperplane to separate the data and providing the best fit to organize it “. (Lee E.J., 2017)



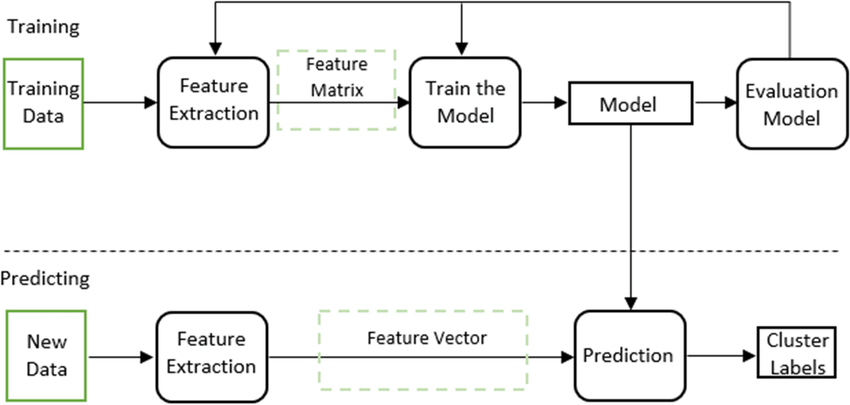
**Figure 1.2** the supervised learning cycle

As shown in **Fig 1.2** the sample training dataset that targets the arbitrator for execution results sends an input to the learning model to predict the test results for that dataset depending on the experiment done, then the arbitrator which is also known as the execution engine returns the data tested and if the data is invalid the arbitrator might send an error signal to the model.

## 1.2.2 Unsupervised Learning:

In unsupervised learning the machine simply receives inputs x1, x2, . . ., but obtains neither supervised target outputs, nor rewards from its environment. It may seem somewhat mysterious to imagine what the machine could possibly learn given that it doesn’t get any feedback from its environment. However, it is possible to develop of formal framework for unsupervised learning based on the notion that the machine’s goal is to build representations of the input that can be used for decision making, predicting future inputs, efficiently communicating the inputs to another machine, etc. (Ghahramani, September 16, 2004). This method of learning uses several types of algorithms including the Deep Belief Networks, Convolutional Neural Networks, and the K-means algorithm which is “the most common unsupervised learning algorithm that is used as a clustering method to identify the mean between groups within unlabeled datasets and create groups based on the mean” (Alloghani M., 2020). A Deep Belief Network (DBN) is a “multi-layer network consisting of intra-level connections useful for data retrieval that typically uses unsupervised learning and has many hidden layers tasked with feature detection and finding correlations in the data”. (Coates A., 2011) .

“A Convolutional Neural Network (CNN) is a multilayer network that relies on feature recognition and identification and is useful for anomaly detection, image recognition, and identification” (Leijnen S., 2020).

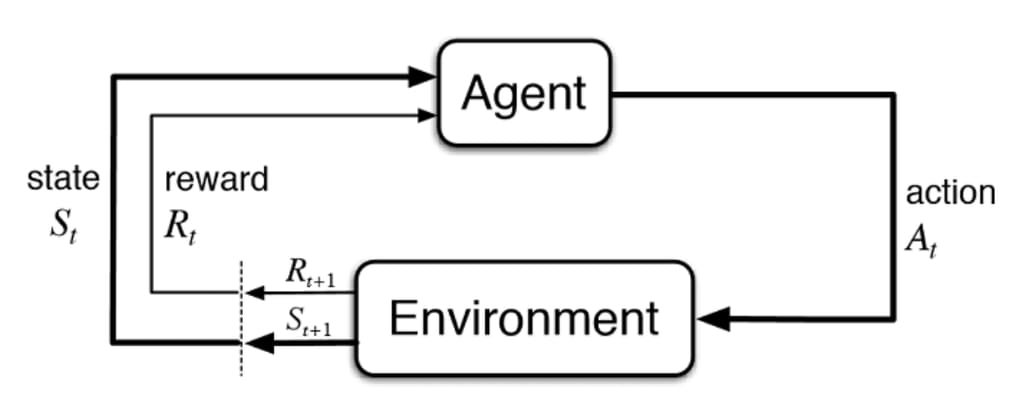


**Figure 1.3**: the workflow of unsupervised learning

In the following figure the two phases of Unsupervised Learning are shown, the training phase and the prediction phase with the training phase beginning with input that does not contain any target values often referred to as training data. In the feature extraction step the input entered is processed resulting in having the useful data needed to generate the output as the feature matrix which is used to train the model using clustering to evaluate the results of the trained model. Whereas in the prediction phase the model uses the previously trained data to generate a prediction of new data. Similar to the training phase feature extraction, this phase follows the same method that ends up generating a feature vector that is passed to the prediction model using the method of clustering.

## Reinforcement learning:

This type of learning uses neither the methods of Unsupervised learning nor supervised learning. This type of learning basically relies on positive feedback from the user in order to improve the learning experience on the model that is being used. This learning method uses rewards as part of the positive feedbacks provided. “Reinforcement learning methods have the potential to influence their environment, are geared towards optimizing the error criterion, and have been described as the closest form of learning as seen in humans and animals” (Sutton R.S., 2018). In this method of learning, the algorithms used are less complicated than those of the other types of learning. Recurrent Neural Network or (RNN) is a neural network that uses delayed inputs and previous output sequences to input it into a new step. (RNN) “is useful for time series prediction, translation, speech recognition, rhythm learning, and music composition” (Leijnen S., 2020). However, this method of learning isn’t considered to be practical in the medical and healthcare field due to its dependency on the reward mechanism.



**Figure 1.4**: workflow of reinforcement learning

Here in the following figure the environment inputs the feedback necessary therefore also giving the agent a state to act upon and based on previously generated output. Then the feedback and the previous feedback and input would cause the agent to return an action which is also referred to as an output based on the state, and the reward factor.

## Deep Learning:

Deep learning is also a subset of Machine Learning itself as shown in **Fig1.1.** Where the use of complex neural networks is utilized. “It is an emerging approach and has been widely applied in traditional artificial intelligence domains, such as semantic parsing” (] A. Bordes, 2012), transfer learning (D.C. Ciresan, 2012), , computer vision (D. Ciresan, 2012), Natural Language Processing. (T. Mikolov, 2013). “Deep learning is a class of machine learning which performs much better on unstructured data. Deep learning techniques are outperforming current machine learning techniques. It enables computational models to learn features progressively from data at multiple levels. The popularity of deep learning amplified as the amount of data available increased as well as the advancement of hardware that provides powerful computers.” (Amitha Mathew, 2021).

Back to the Healthcare sector, this field faces major obstacles in deriving insights, gathering information, analyzing broad complex data from sources that claim themselves as credible. “By using cellular network technologies,

wearable sensor devices can transmit health data to hospital databases and then to cloud storage systems. Data collected from these sources can then be analyzed for medical purposes” (Qureshi, Din, Jeon, & Piccialli, 2020) .Many researchers have used machine learning for disease detection and pattern recognition. (Shailaja, Seetharamulu, & Jabbar, march 29-31 2018). In spite of the many factors stating that machine learning can be used effectively, the amount of studies that provided a demonstration stating that the accuracy of healthcare data and patient diagnostics can be enhanced, minimizing risks in the medical field by improving the decision making process, and early disease detection can be achieved through the integration of machine learning into healthcare is thin in number. Therefore, this paper aims to take a deep dive into exploring the applications of machine learning in the healthcare sector, shining the light on its effective role in the enhancement of the concept of early disease detection, by exploring recent studies, implementations, survey studies, limitations, and recommendations for the future. The following research and experiment is dedicated to recognize the areas where machine learning can escalate its efficiency in healthcare, the obstacles to face and prevail over to minimize them, focusing on the early prediction accuracy for diseases.

# Background and literature review:

While the use of machine learning and artificial intelligence in medicine has its roots in the earliest days of the field. (Addis., 1956), but only in the last preceding years the integration of machine learning models into the medical field was encouraged and researched.

## 2.1 Medical Diagnostics using machine learning:

Coronavirus Resource Center at Johns Hopkins University of Medicine has reported a total of 23,638 deaths as worldwide COVID-19 infections surpass 500,000 (as of 5 PM EST on March 26, 2020). And COVID-19 was considered as one of the most dangerous pandemics due to the high death rate it had. ”To manage the spread of the COVID-19 infection among people rapidly, all the governments around the world applied severe actions, such as the quarantine of hundreds of millions of citizens worldwide” (al, 2020). However, distinguishing between individuals that tested as positive and those who tested as negative depending on several symptoms of the disease itself remained as a difficult task to be completed overall. Then a solution that dated back to 1983 that was invented by “Kary Mullis, a technician at the Cetus Corporation, assigned to improve the synthesis of oligonucleotides” (Kaunitz, 2015) and develop various essential methods for DNA manipulation , “along with the advancement in medical diagnosis, nucleic acid detection-based approaches have become a rapid and reliable technology for viral detection. Among nucleic acid tests, the polymerase chain reaction (PCR) method is considered as the ‘gold standard’ for the detection of some viruses and is characterized by rapid detection, high sensitivity, and specificity.” (Ardebilib, 2020). Simultaneously, these viral testing methods despite their practicality, they also consume a large amount of both time and money. It is also considered to be difficult to treat a large scale of the population with those viral tests. However, a study done by (Jordi Laguarta, Ferran Hueto, and Brian Subirana) states that the following issues can be addressed effectively by solutions discovered and implemented using Machine Learning “by presenting the dataset, model architecture and performance of a zero-cost, rapid and instantly distributable COVID-19 forced-cough recording AI pre-screening tool achieving 98.5% accuracy, including 100% asymptomatic detection rate. An orthogonal set of biomarkers may be developed to diagnose COVID-19, Alzheimer’s and perhaps other conditions.” (Jordi Laguarta, 2020) This type of diagnosis assists the patient by providing and delivering real-time results after the user just inputs a voice recording of the cough then the model processes the audio based on both the user input of the cough sound and patterns learned by the AI Model to derive accurate results. Another remarkable advantage of this diagnosis approach is the ability to be cost effective by nearly making it accessible to anyone throughout the world with no additional cost through a mobile app or a website. The way this system works is by taking some sample training datasets which basically consists of audio recording files that contain cough sounds to train a neural network using the supervised learning based binary classification method that uses Artificial Neural Networks (ANN) that solves regression and classification based problems. The aim of this approach is to educate the network so that it could distinguish between infected individuals and those with a negative infection status.

## 2.2 Patient Monitoring Using Predictive Analysis:

“Remote Patient Monitoring (RPM) has witnessed remarkable advancements in recent years, thanks to the integration of Artificial Intelligence (AI), Machine Learning (ML), Deep Learning (DL), Natural Language Processing (NLP), and Computer Vision (CV) technologies.” (Gaidaa Maher Dogheim, 2023) Machine Learning also played a major role in predicting the occurrences of diseases based on patterns the AI models previously learned when encountering patients with certain diseases. This method of patient monitoring is really effective especially when it comes to preventing the occurrence of diseases before it’s too late using early prevention of diseases. “Through the analysis of extensive patient data, including medical records, laboratory results, genetic information, and imaging data, machine learning algorithms can unveil intricate patterns and identify subtle indicators that may evade human clinicians. By detecting these early signs of diseases such as cancer, cardiovascular diseases, diabetes, and neurodegenerative disorders, these algorithms offer a vital advantage in improving patient outcomes.“ (Mohamed Said Ibrahim D. S., 2023). Those analytics can be processed using various methods that extract the datasets originating from the patient to analyze this data in an efficient manner. One of those trendy methods is the use of wearable devices that support the use of machine learning to analyze the data based on multiple criteria ranging from (heartbeat rate, calorie consumption, metabolism, physical activity, etc….) such as (Smart watches and health bands that are mainly used by athletes to track their heartrate, steps, sleep schedule and much more. “The first study investigating the use of a modern wearable activity tracker on post-surgical mobility recovery during hospitalization was recently published by the Mayo Clinic” (Cook DJ, 2013). The following research displays an occurring relationship between the duration of the hospital stay, the steps counted in early recovery, and elderly cardiac surgery patient disposition. “The use of accelerometers and pedometers to measure physical activity has also enhanced treatment for individuals participating in pulmonary rehabilitation” (R. B. , 2009). “A recent meta-analysis of activity monitor-based counseling studies with diabetes patients concluded that activity monitor-based counseling had a beneficial effect on physical activity, blood glucose, systolic blood pressure, and body mass index”. (Vaes AW, 2013). On the other hand, credible evidence proving the efficiency of using those wearable activity monitors in detecting chronic diseases such as (osteoarthritis, cardiovascular diseases, type 2 diabetes mellitus, and chronic obstructive pulmonary disease) (Allet L, 2010). Studies have found modest short-term activity improvements and weight loss resulting from monitor and pedometer use. (Barwais FA, 2013) (Freak-Poli RL, 2011) (Tudor-Locke C, 2009).However, the following results were not sustained over extended periods of time. “Self-management technology and physiological measurement have expanded beyond hospital settings to remote and direct use by patients with chronic diseases”. (Patel S, 2012). But in some cases, some patients face difficulties when it comes to understanding the use of wearable activity tracking since patients who suffer from certain health conditions face due to the patient’s needs that change during the time of their illness. which encourages greater patient-physician collaboration, expanded patient social networks, and increased patient use of personal data for tracking their health outcomes. In order to benefit the most from the use of those wearable devices, the approach of patient driven healthcare is taken. This approach “encourages greater patient-physician collaboration, expanded patient social networks, and increased patient use of personal data for tracking their health outcomes”. (Wicks P, 2014) . A good method to enhance their uses is by integrating machine learning into these devices. Machine Learning in that case is used “for predicting the chronic diseases using continuous and real-time data obtained from multiple sensors. Combining data from different sensors can help, not only to improve the detection performance of chronic disease care, but also to identify patterns and correlations between different physiological signals.“ (ADITI SITE, 2021) “Depending on the data and the information to be extracted from the data, various machine learning algorithms for analysis and preprocessing are used. Support vector machines are used for analyzing the Electro Cardio Gram (ECG) and Electro encephala gram (EEG) data for diabetes classification” (ADITI SITE, 2021). “Neural network algorithms have been used for the prediction task for disease diagnosis using data from wearable medical sensors, multilayer perceptions are used for classification of different types of diabetes and for behavioral analysis for Parkinson’s disease. Similarly, artificial neural networks (ANN) are used for the analysis of EEG, Electromyogram (EMG), voice signals. Along with this there are many pre-processing methods and feature selection” (ADITI SITE, 2021).

## 2.3. Personalized Medicine and Treatment :

“In the last ten years, many advances have been made in the treatment and diagnosis of immune-mediated diseases. In particular, an increasing number of new monoclonal antibodies and small molecules have been developed for the treatment of these conditions. Concurrently, many new genetic or serological markers have been discovered to increase our capability in the early diagnosis of autoimmune diseases.

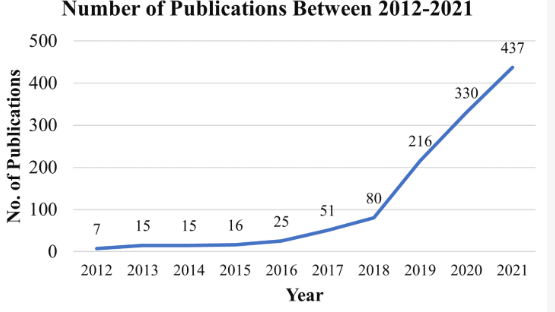
In the same period, advances in artificial intelligence and machine learning have allowed great improvements in the treatment and follow-up of some diseases, such as cancer, but experience in autoimmune systemic diseases is still very limited”. (Marco Sebastiani). Machine learning can identify disease risks, and then provide insights based on the risks identified, then the derived insights could be used to give recommendations and personalized treatment for the ill individual through the use of clustering in post genomic data analysis. “Unsupervised classification has many applications in post-genomics. In particular, clustering plays a crucial role in the analysis of gene expression data” (Eisen, 1998) Clustering can also be applied directly to sequence the data, for example to group genes based on shared coregulatory regions (Bilu, 2002). It serves as a data-mining tool to analyze both proteomics and metabolomics data (Goodacre, 1998).and can be applied in the context of protein comparison and structure prediction. (Kaplan, 2004).

This literature review is aimed to inform the reader how machine learning influences the realm of healthcare using studies previously done by other researchers.

# 3.Methodology

## 3.1. Studies and Benchmarks

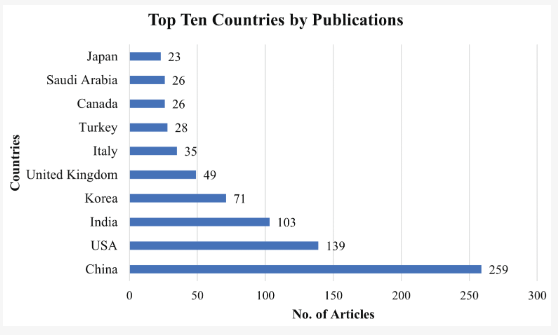
As we discussed in our literature review, machine learning can be very effective in the early diagnosis or detection of diseases in order to treat the medical conditions as soon as possible before the illness deteriorates even more while having the highest level of cost efficiency, and accuracy in predictions. “the most recent developments and studies in machine-learning algorithms have proved that machine learning can make a substantial impact on the detection and diagnosis of several diseases” (Pushpa Singh, 23 April 2021). The effectiveness of machine learning algorithms exceeds far beyond individual patients. By aggregating and analyzing data from large populations in details, these algorithms can uncover demographic shifts and patterns that hold immense value for public health initiatives. (Mohamed Said Ibrahim S. S., 2023 ). However, the publication of using the machine learning models to detect those diseases varies from one research journal to another. (Md Manjurul Ahsan, 15 march 2022).



**Figure 3.1** publications by year in using the machine learning early disease detection approach

As we observe in **Figure 3.1** according to the data fetched from the Scopus and WOS abstraction databases which are basically databases of reviewed literature including books, scientific articles etc.. the publication growth by journals from the year 2012 till 2016 remained at a level close to neutral due to the limitations of technology back then. However in the year 2017, a drastic escalation occurred in the number of publication of this disease detection method. And the numbers of those publications keep increasing progressively till the present day.

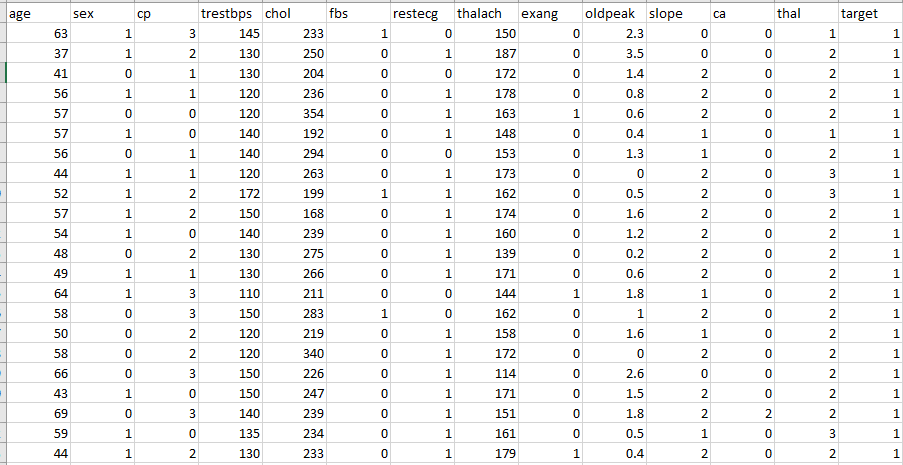
In addition to that, a lot of countries in today’s world are beginning to implement this machine learning concept in their studies and research. Mainly the first world countries are the ones who began deep diving into implementing this method. As the below figure states.



**Figure 3.2** Top ten countries that contributed to the machine learning early disease detection literature.

Here in **figure 3.2**, the top 3 countries that are implementing this diagnosis approach are china that has the highest number of publicized research articles in early diagnosis at 259 articles conducted in this domain, USA being in second place at more than half of China’s published articles at 139, and thirdly India at 103 articles. (Md Manjurul Ahsan, 15 march 2022).

One of the most efficient disease detection algorithms is the Naïve Bayes algorithm which is a probability based algorithm. Briefly, Bayesian theory and probability are named after a British 18th century mathematician, Thomas Bayes. Bayesian logic can show the result of a patient’s test with a pre-test probability (of the population), to predict or determine the chance of finding a disease (Mostafa Langarizadeh, Nov 2016). Here we use the Naïve Bayes supervised approach and the random forest classifier and then compare them in benchmarking a program that will be built based on a dataset obtained from kaggle which is a website that provides datasets for experimenting purposes. To test the theory of predicting the occurrence of a heart disease based on certain symptoms and factors assessed from anonymous patients. In this experiment, I’ll be using Python 3.10 on windows 10 as the interpreted programming language to create my program and test it on the Google Colab interpreter which is an online hosted Jupyter notebook service intended for research and benchmarking programs.



**Table 3.3** this Table describes the dataset with 303 entries and 14 columns with target as the target variable

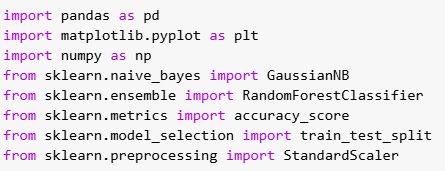
### **3.1.1 Data Dictionary:**

Here this section acts as a legend for the dataset to understand the meaning of each acronym in the dataset columns. In some cases here we used values as 1,2,3 since those values represent categorical records(yes/no) as an example. And encoding those categorical columns was necessary since classification problems need to conduct predictions based on the numerical fields(encoded) fields.Also note that the data was previously pre-processed. Note that the dictionary is not provided by me but by the publisher of this dataset on kaggle under the username: **desalegngeb.**

1. age: age in years
2. sex: sex
   * 1 = male
   * 0 = female
3. cp: chest pain type
   * Value 0: typical angina
   * Value 1: atypical angina
   * Value 2: non-anginal pain
   * Value 3: asymptomatic
4. trestbps: resting blood pressure (in mm Hg on admission to the hospital)
5. chol: serum cholestoral in mg/dl
6. fbs: (fasting blood sugar > 120 mg/dl)
   * 1 = true;
   * 0 = false
7. restecg: resting electrocardiographic results
   * Value 0: normal
   * Value 1: having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05 mV)
   * Value 2: showing probable or definite left ventricular hypertrophy by Estes' criteria
8. thalach: maximum heart rate achieved
9. exang: exercise induced angina
   * 1 = yes
   * 0 = no
10. oldpeak = ST depression induced by exercise relative to rest
11. slope: the slope of the peak exercise ST segment
    * Value 0: upsloping
    * Value 1: flat
    * Value 2: downsloping
12. ca: number of major vessels (0-3) colored by flourosopy
13. thal:
    * 0 = error
    * 1 = fixed defect
    * 2 = normal
    * 3 = reversable defect
14. target (the target variable indicating the occurrence of a heart disease):
    * 0 = no disease,
    * 1 = disease

### **3.1.2 conducting the experiment :**

Using the datasets present in **Table3.3,** we will use the pandas library for converting the datasets from Lists to data-frames which takes the form of a table in python for manipulating data. We will also use matplot lib to plot the dataset and visualize it’s correlations, predictions and compare them in the next section of this paper. Finally the SciKit-learn model selection library would be used in splitting the data accordingly by using the columns with the dropped target as input and the dropped target itself as the output, evaluating the model, and training the model based on the dataset provided.

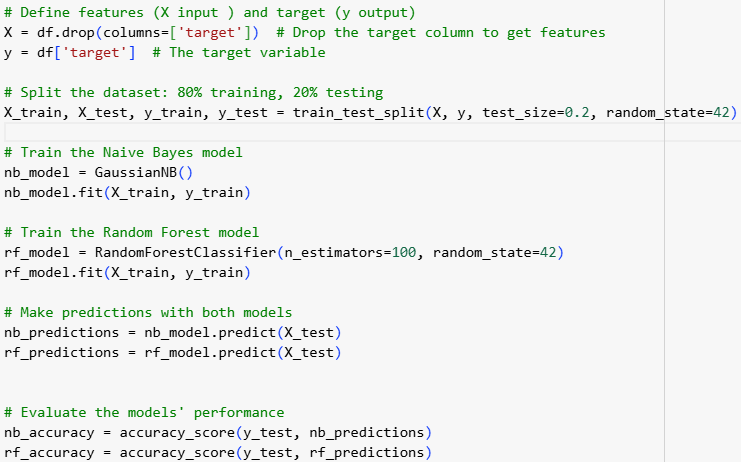


**Figure 3.4:** importing the necessary python libraries

As we can see we imported pandas, standard scaler, and numpy to perform additional data exploration on this dataset including: handling missing values and normalizing numerical columns. We used the sklearn naïve bayes classifier and the sklearn ensemble random forest classifier to predict the target variable and then compare both models by also evaluating them using the accuracy score classification performance metric.



**Figure 3.5** python code for basic data exploration.

The dataset used was converted to a pandas data frame in order to allow the manipulation of data in python as shown. We check for null values and drop them and normalize the numerical columns in the dataset and we show the first 5 rows after the changes we made. 

**Figure 3.6** creating the predictions for both models and evaluating their performance

The input variable is taken by dropping the target column and including all other columns. However the output only takes the target column. Those 2 values will be divided into 4 train-test variables. Then we initialize naïve bayes and random forest classifiers and we train them using the x and y train variables and the predictions are then applied on the x-test values. Then the performance metrics of both models are calculated respectively and in the results we will also show the confusion matrix of both models.

For code validation I will provide the link to the notebook I coded on: [Python code](https://colab.research.google.com/drive/1RM9SSOFp7xATbvoKU8nr5sh-xxiJPl_X?usp=sharing)

## 3.2 survey study

Obviously, many individuals have different opinions when it comes to the applications of machine learning in general, so in order to learn more about others’ thoughts, we created a google forms survey in the timeframe of 1 day gathering numerous responses in the name of the Lebanese International University that is also considered as a social experiment supporting the practical and technical part of the research by also exploring the challenges that this approach faces in today’s society. This survey mainly gathers general thoughts from respondents about the machine learning approach when It comes to using it in healthcare especially in the early prediction of diseases. Validating the hypothesis, we made in the introduction.

**Demographic studies**: First we gathered the audience’s main demographic findings including age group and gender just to simply have the ability to make the analysis based on the audience group that provided the answers. Then we also asked the participants whether they do have medical background or not due to the fact that this survey targets both medical experts and people who are curious to explore the uses of machine learning in healthcare. Then we also questioned whether the participants are diagnosed with medical conditions or not since some of the diagnosed individuals might actually benefit from considering the machine learning approach in disease diagnosis. Since some people aren’t really familiar with machine learning being used In healthcare in general we also gathered the responses that are both familiar and unfamiliar.

**Machine learning focused questions:** Another question was prompted to gather responses that state in which areas do they think that machine learning in healthcare is used in whether it’s early disease diagnosis, drug discovery, or personalized medicine. The next question stated whether the participants would trust using an AI model to assist in disease diagnosis since some people don’t have too much faith in AI models as sometimes they might produce inaccurate results and even have confidentiality issues when sharing their data with the models. the next question just assesses the hypothesis that says “can machine learning assist in early disease diagnosis?” since early disease diagnosis is our main topic of the study and we want to visualize how the participants are satisfied with the hypothesis.

**Privacy and confidentiality:** The next question targets the privacy and confidentiality issues as we said previously and it simply asks whether the participant would be comfortable with sharing their personal data to an AI model or not. Continuing this question, if the participants may have concerns for sharing their data online they are free to write their own concerns.

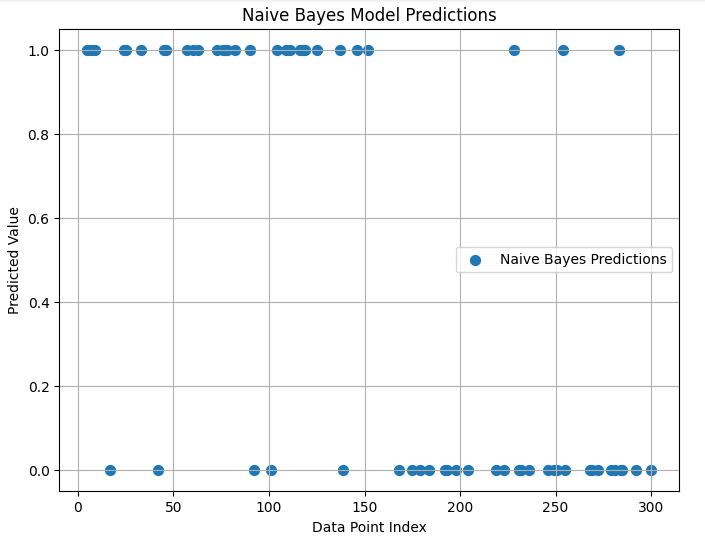
Then if some people think that machine learning in healthcare is effective, how would it benefit disease diagnosis by selecting answers from 4 choices (faster diagnosis, reduced costs, improved accuracy, personalized medicine). The last question sums up the survey in general and lets the participants freely write their concerns for using machine learning in disease diagnosis in general. The link for this survey is found here: [survey](https://forms.gle/ZWmJQAqmT691mSFM7)

# 4.Results

After we predicted the values for both models. The values correspond from 0 to 1 and nothing in between considering that this is a classification problem which means that the predicted value indicates whether individuals are diagnosed with a heart disease or not. To understand the predictions results we visualized it using the matplotlib library. First we will start with the naïve bayes model results and then proceed to the random forest classifier.

## 4.1: Model experiment results:

Here we plotted the naïve bayes predicted values as shown below with the predicted values on the y- axis and the data point index on the x axis with step size 50 and ending at 300 since it is a dataset with 300 entries.



**Figure 4.1:** The naïve bayes predictions

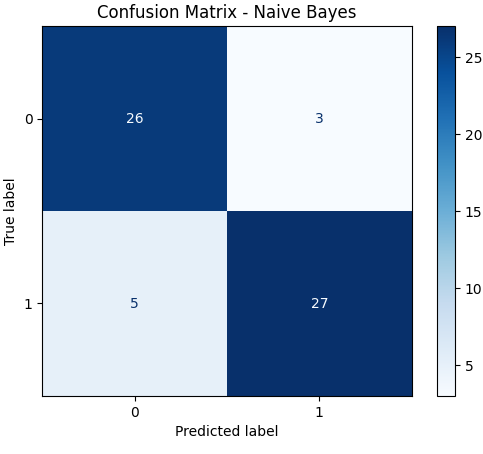
Next we plot the confusion matrix of the naïve bayes model that shows the amount of true positives(TP),true negatives(TN),false positives(FP),false negatives(FN). Those 4 categories correspond to:

**TP**: valid prediction of the presence of a heart disease

**TN**: valid prediction of the absence of a heart disease

**FP**: invalid prediction of the presence of a heart disease

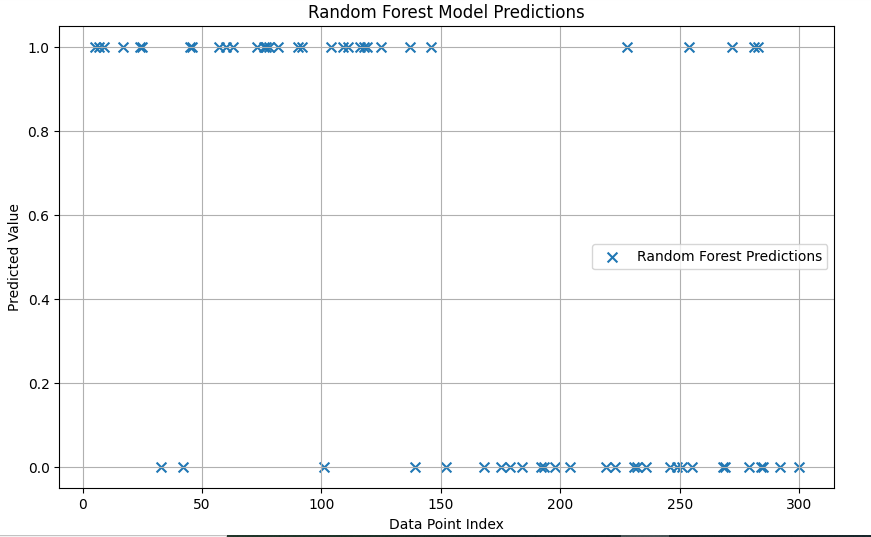
**FN**: invalid prediction of the absence of a heart disease



**Figure 4.2:** the confusion matrix of the naïve bayes model

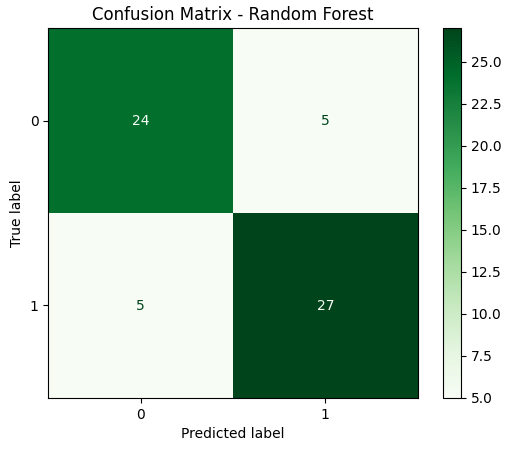
Here we notice that we have 26 true positives ,3 false positives, 5 false negatives, 27 true negatives.

Now we apply the same procedure for the random forest classifier and in the end we plot the accuracy score of both models together to compare both models in the end and see which model was more efficient in predicting results.



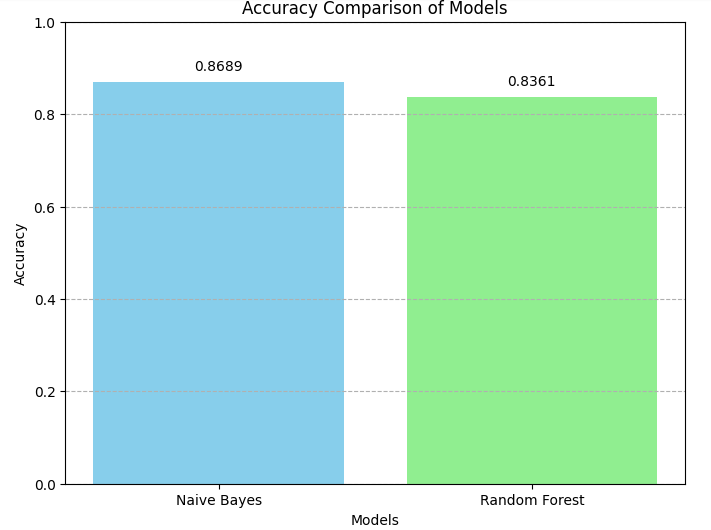
**Figure 4.3:** the random forest predictions

Then we also plot the confusion matrix of the random forest classifier.



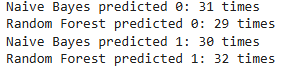
**Figure 4.4** : the confusion matrix of the random forest classifier

Then at the end of this section we plot a bar graph that compares the accuracy scores of both models with light blue representing the naïve bayes model and light green the random forest classifier.



**Figure 4.5** : the accuracy score of both models

And for analysis purposes later on a program was created using the numpy library to count the number of predictions at 0 and 1 respectively for both models that were plotted in **figures 4.1** and **4.3.**

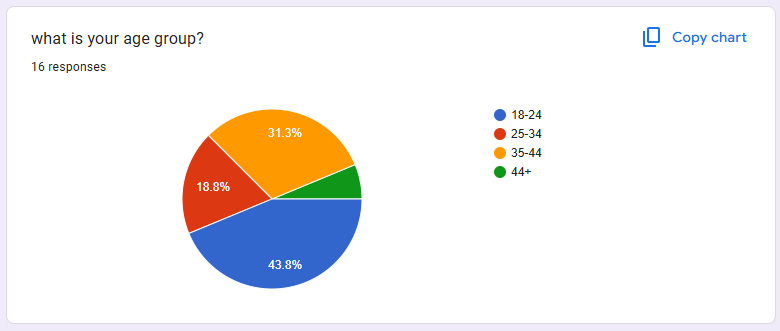


**Figure 4.6:** the output that counted the points at 0 and 1 for both models

## 4.2: survey results:

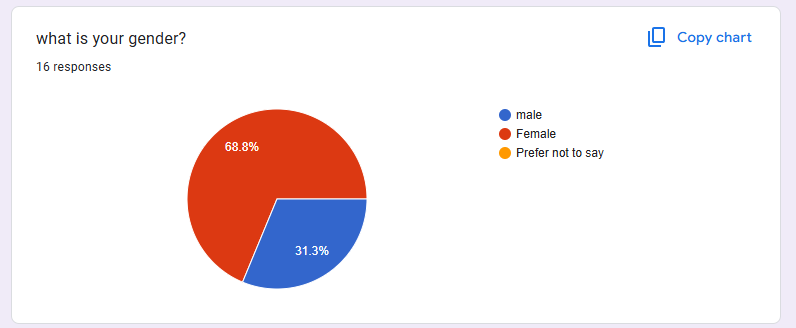
This section is dedicated to showcase the findings of the survey done in the methodology section of this paper.

## 4.2.1 Demographic findings :

****

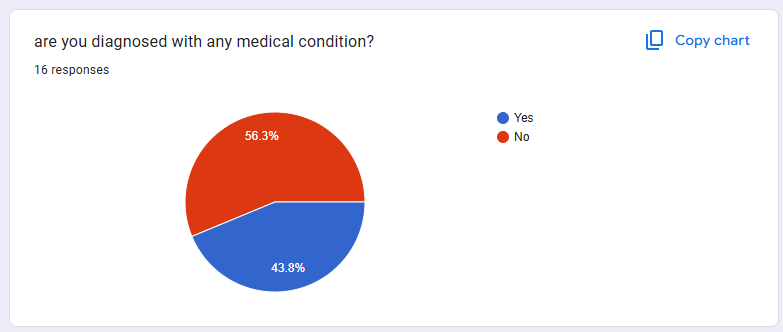
**Figure 4.7:** age groupof respondents.

The above figure showed that there was a total of 16 respondents, 43.8% of them within the age group of 18-24, 31.3% within the age group 35-44, 18.8% within 25-34, with a minority of people aged more than 44. (Soubra J. M.)



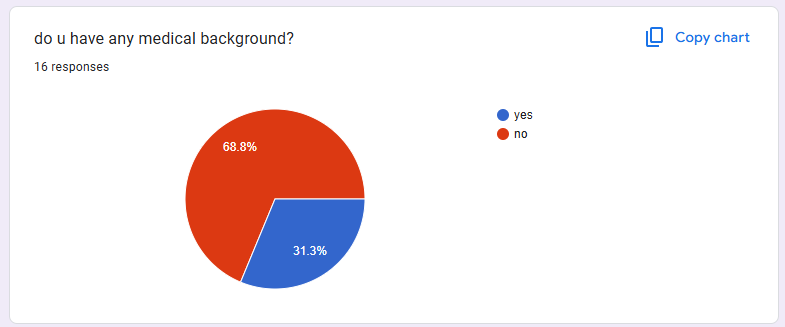
**Figure 4.8:** gender of respondents.

In this survey most respondents were females with a dominant presence of 68.8%. (Soubra J. M.)



**Figure 4.9:** people with a medical condition diagnosis

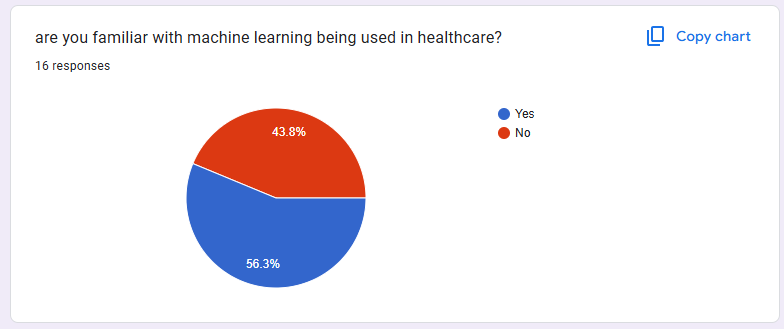
In the above figure 56.3% of the respondents are not diagnosed with medical conditions (Soubra J. M.)



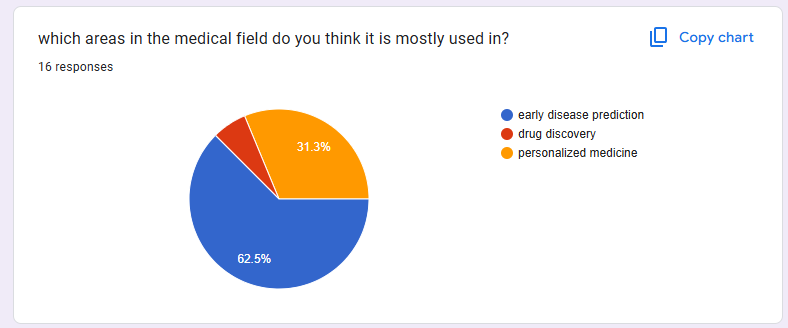
**Figure 4.10:** people with a medical background

Here 68.8% of respondents do not have experience in the medical domain while the remaining participants do not. (Soubra J. M.)

## 4.2.2 Machine learning focused questions:

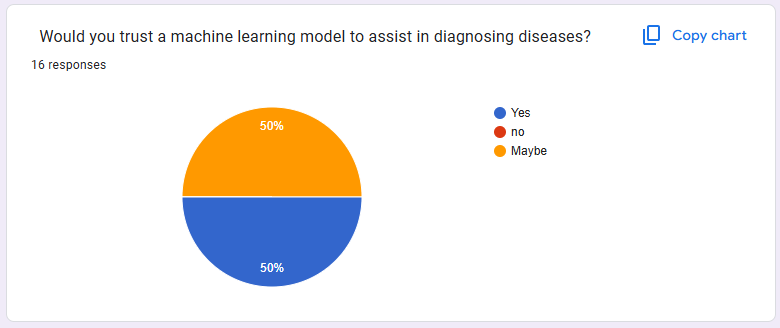


**Figure 4.11 :** people being familiar with machine learning used in healthcare.

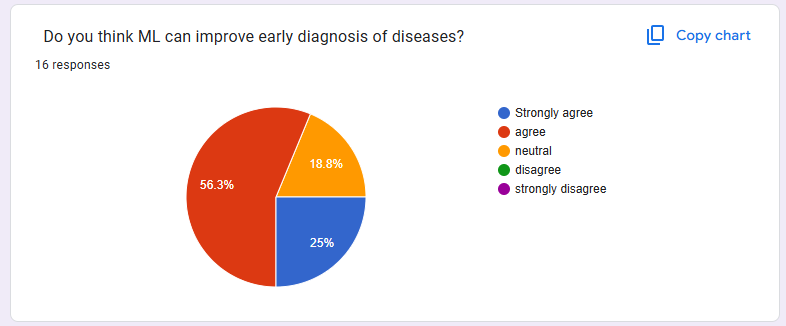
Here 56.3% of people are familiar with machine learning used for health care. (Soubra J. M.) 

**Figure 4.12:** uses of machine learning in healthcare.

Here as we realize the majority of responses were mainly directed towards early disease prediction which is the main topic we are researching with a dominant 62.5% percentage stating the importance of machine learning in early disease prediction. However some individuals claimed that machine learning is used for personalized medicine with a percentage of 31.3%. with a minority choosing drug discovery. (Soubra J. M.)

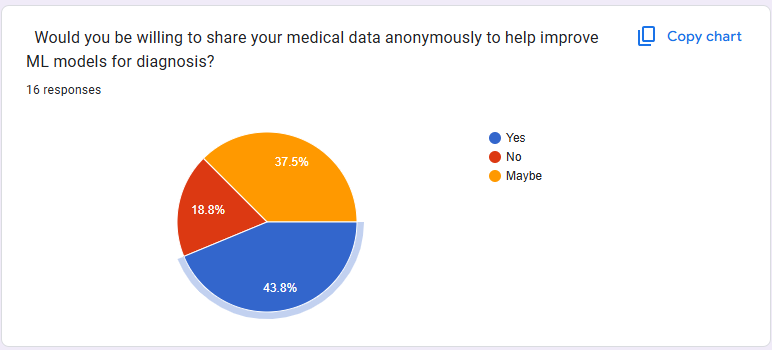


**Figure 4.13:** trusting a machine learning model to assist in diagnosing diseases.

Here we can see that half of the participants do trust machine learning algorithms and models to assist in the disease diagnosis process. However, the other half of the respondents responded with “maybe”. (Soubra J. M.) 

**Figure 4.14:** improving early disease diagnosis.

we notice in **Figure 4.14** that most people agree with the idea that machine learning can be a game changer when it comes to early disease prediction with 25% of respondents strongly agreeing and 56.3% who support that fact with no biases. And 18% of the respondents stated that they neither fully agree nor disagree making the margin of disagreement very minor. (Soubra J. M.)

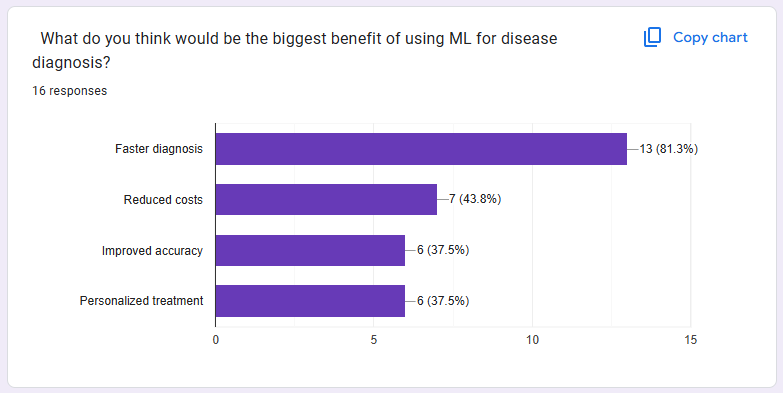


**Figure 4.15:** anonymous data sharing to help ml models provide better predictions

We notice that the issue of anonymity and data confidentiality can be one of the obstacles that this domain faces due to the fact that only 43.8% of the participants stated that they are fully willing to share their anonymous data with those AI models. Whereas the remaining 56.2% of participant displayed a form of mistrust in those ai models with sharing their data. (Soubra J. M.) 

**Figure 4.16:** the concerns for sharing data with ml models.

One of the detailed unique responses that validates the concerns of the respondents is the occurrence of HIPAA violations also known as “The Health Insurance Portability and Accountability Act (HIPAA), a 1996 federal law, is a widely cited and misunderstood privacy statutes.” (Schumaker, August 5, 2021, 5:00 AM). With other responses doubting the security and confidentiality of these models as stated before and the rest feel safe sharing their data online. (Soubra J. M.)



**Figure 4.17:** graph showing thebenefits of machine learning for disease prediction

A dominating percentage of respondents show that ml disease diagnosis is effectively used for faster diagnosis. Also 43.8% state that ml models are cost friendly so therefore they might be used really efficiently in that manner. Whereas similar simultaneous percentages of 37.5% of individuals stated that ml is used effectively for improved diagnoses and accuracy and simultaneously personalized treatment. (Soubra J. M.)

# 5.Discussion

After we plotted the results of the models in section 4.1 and the survey results in section 4.2, this section will focus on the analytic comparison between the naïve bayes and random forest models in terms of predictions and accuracy and it will also focus on the analysis of the survey results provided in the previous section and discussing the limitations we faced during this research.

## 5.1 plot analysis:

This section of the discussion mainly focuses on the technical part of the experiment results. Where we compare the predictions of both models and assess them based on their confusion matrix and their accuracy score.

### **5.1.1 insights and analysis:**

As we can see in **Figure 4.1** (naïve bayes)and **Figure 4.3** (random forest) **both** models performed well predicting classified values that are either 0 (absence of heart disease) or 1 (presence of heart disease).In the naïve bayes model more cases of heart disease presence were predicted since the number of data points at 0 are greater than those at 1 with 31 predictions at 0 and 30 at 1. However, in the case of the random forest it is the total opposite with 29 predictions at 0 and 32 at 1 as stated in **figure 4.6.** After comparing both models prediction-wise, we will also compare them based on their accuracy scores and confusion matrices classification performance metrics. As we can see in **figure 4.2** and **figure 4.4** the amount of true positives in the naïve bayes classifier was greater than that of the random forest meaning that the naive bayes model was more efficient and accurate in heart rate diagnostics that indicate the presence of a heart disease. However, both models equally predicted the absence of a heart disease with 27 true negatives on both ends. Then we compare the accuracy scores of both models in **figure 4.5** and the results are obvious that the naïve bayes classifier had a higher accuracy score standing at 0.86 which is greater than that of the random forest classifier at 0.83. Therefore, in general the naïve bayes model performed better than the random forest model in aspects of accuracy and minimal numbers of false positives.



**Figure 5.1:** the formula for the accuracy score

### **5.1.2 Obstacles and limitations:**

This experiment was performed well however there is always room for more improvement. We did encounter a few limitations to the current experiment one of them is:

1. **Overfitting:** despite both models performing really well, random forest was more subjected to overfitting meaning that the model may be limited to predicting values on new data.
2. **Limited use of performance metrics:** here in this experiment we only used 2 performance metrics to assess the models, despite the good results. But using more classification performance metrics like f1 score, recall score, and precision score in order to maximize the efficiency of the results
3. **Generalization:** the dataset does not represent a broad population in general therefore the model might not function well when subjected to a broader dataset.

### **5.2 survey analysis:**

This part of the discussion focuses on the social part of the experiment results derived by the end.

### **5.2.1: audience and perceptions:**

**Audience analysis:** This survey mostly caught the eyes of younger individuals indicating that our current generation is the most tech-oriented till this day with a dominating percentage of 43.8%. (Soubra J. M.) therefore, showing that younger people nowadays are more invested in machine learning and learning more about technology as seen in **figure 4.7**. However, 68% of the participants lack the medical expertise that also indicates that most of the results derived were from a perspective where people were not deeply specialized in the medical field (**figure 4.10**) (Soubra J. M.).

**Use of ML in healthcare**: early disease diagnosis takes the lead at 62.5% indicating the significant importance and efficiency of using machine learning in EDD which is the main topic of our experiment and research. Also personalized medicine received some recognition at 31.3% indicating that it can also have a good role in choosing the right dose of medical drug to treat each patient accordingly. With the remaining percentage going towards drug discovery signifying that it is also a good role, however it’s use isn’t as recognized as disease diagnosis and personalized treatment. As shown in **figure 4.12.**

**Discussing ML advantages and disadvantages**: Our study also shows that there are 2 halves of people the first half is the one that shows strong optimism in trusting ML models with disease diagnosis. However, the other half is a bit concerning due to the neutral trust in ML models which still presents a major obstacle therefore more studies and developments should be done on those models to improve the confidentiality and integrity of the ml models (**Figure 4.13**). (Soubra J. M.) . Most of the participants believe that the 2 pillars of success in integrating ML models in disease diagnosis are speed, and cost-efficiency. Indicating that in the medical domain speed can make a major difference as every second can be a major deal-breaker when it comes to saving a life and also resulting in saving resources financially speaking to invest in having even more resources to enhance the performance of the medical field. (**figure 4.17**) (Soubra J. M.) Confidentiality and privacy remain remarkable obstacles in trusting machine learning models despite the high proportions (43.8%) of respondents ready to take a risk in assisting the evolution of technical skills in healthcare by sharing their private data with ML models, the other responses showed neutrality meaning that there are still concerns about the secure implementations of those models as shown in **figure 4.15.**

* **5.2.2: limitations:**

Most of the participants displayed the will of contribution to AI and ML, as most participants believe that machine learning is really effective in early disease diagnosis, some of them still have second thoughts about that and that assumption might be due to the minimal experience in understanding successful experiments in health-care focused machine learning. Also the most recognized limitation we encountered here was the privacy issues that received the spotlight despite the feedback stating that a lot of people support the use of machine learning since some of the participant did not show a convincing amount of trust in those AI models in early disease diagnosis and in other fields too.

# Conclusion

The purpose of this study is to use machine learning models more specifically the naïve bayes and the random forest models to predict early heart disease diagnostics. And as a result, both models produced really effective results with 86% and 83% diagnosis accuracies with the performance metrics proving the good results when we plotted both prediction results. The research showed effective results providing an answer for the main statement of this paper. But there were a few limitations in this approach including the quality of the data therefore resulting in a potential case of overfitting which means that this algorithm fits when the training data is applied to predict diagnostics. And the reason for the risk of overfitting is the use of a small dataset. The results found in this experiment are valuable, however the implementation using more varying and complex datasets can make the predictions more efficient. Therefore, concluding that the naive bayes model is one of the best approaches of early disease prediction when small datasets are provided. And also the survey results clearly stated the importance of machine learning in healthcare and both the advantages and disadvantages and shining the light on confidentiality and personal data security issues. The treatment of this issue is recommended to make ml models more efficient in early disease diagnosis. To apply practical application ml models like naïve bayes and random forest classifiers should be used for early disease diagnosis. Some future work recommendations include:

* Shedding the light on handling the privacy and security concerns to increase the usability of the ml disease diagnosis approach.
* The use of feature engineering to make the data cleaner.
* Using a larger data set and utilizing real-world applications.
* Handling data outliers
* Using more machine learning models like SVM, KNN to explore various different results
* Conducting the survey on a larger audience for better results

This research paper contributes to the applications of machine learning in health care as the title of this paper states while focusing on early disease diagnosis using the naïve bayes model that proved to be one of the models that can still achieve a high prediction accuracy rate when the training process is completed through a structured dataset.In addition to that, this study focuses not only on the practical and experimental side, but also on the social side by conducting a survey to gather insights from people who are from different audience groups with some of them being professionals in healthcare. This 2 sided approach makes sure that this study does not only focus on the technical reliability of the findings, but also on the outside world insights that shows the need to secure patient data and take their concerns in consideration and the importance of integrating machine learning models in the healthcare realm. In summary, this paper defines the importance of the use of machine learning models to enhance patient outcomes through the process of early disease diagnosis. Despite both classifiers providing really good results, their applications in larger and more realistic datasets is encouraged and recommended for maximum efficiency. Future work should embark on addressing privacy issues, comparison of models, using larger datasets, and reducing the probability of potential overfitting. Also data scientists teaming up with health care professionals should be more encouraged to dive deeper into this domain therefore increasing the implementation in machine learning models in clinics especially through EDD (early disease diagnosis).

# References

Chan, F. T., Jiang, B., & Tang, N. K. (n.d.). The development of intelligent decision support tools to aid the. *Int. J. Production Economics 65 (2000) 73}84*, 73.

Makary , M. A., M. A. Makary and M. Daniel, & Daniel, M. (2016). the third leading cause of death in the US. *Medical error*, BMJ, vol. 353, doi: 10.1136/bmj.i2139.

] A. Bordes, X. G. (2012). Joint learning of words and meaning for open-text semantic parsing . *Proceedings of the AISTATS*.

Addis., T. (1956). Towards an” expert” diagnostic system. . *ICL Technical Journal* , 79-105.

ADITI SITE, J. N. (2021). Systematic Review on Machine-Learning Algorithms Used in Wearable-Based eHealth Data Analysis. *IEEE Access* .

al, A. A. (2020). ) Artifcial intelligence and machine learning to fight COVID-19. *American Physiological Society, Bethesda*.

Allet L, K. R. (2010). Wearable systems for monitoring mobility-related activities in chronic disease: a systematic review. *Sensors*.

Alloghani M., A.-J. D. (2020). A systematic review on supervised and unsupervised machine learning algorithms for data science.

Amitha Mathew, P. A. (2021). Deep Learning Techniques: An Overview. *Department of Computer Science and Engineering, School of Engineering, Avinashilingam*.

AN Ramesh, C. K. (2004). Artificial intelligence in medicine. *The University of Hull Academic Surgical Unit, Castle Hill Hospital, Cottingham, UK*, 334.

Ardebilib, A. T. (2020). Real-time RT-PCR in COVID-19 detection: issues affecting the results. *EXPERT REVIEW OF MOLECULAR DIAGNOSTICS*.

Artificial Intelligence. (Press, Cambridge, 2014. Accessed: Feb. 15, 2021. [Online]. Available). *Cambridge International Dictionary of English. Cambridge University press*.

Barwais FA, C. T. (2013). Physical activity, sedentary behavior and total wellness changes among sedentary adults: a 4-week randomized controlled trial . *Health Qual Life Outcomes*.

Bilu, Y. a. (2002). The advantage of functional prediction based on clustering of yeast genes and its correlation with non-sequence based classification. *J. Comput Biol* , 193–210.

Coates A., L. H. (2011). An analysis of single-layer networks in unsupervised feature learning. *J. Mach. Learn.*

Collin Burns, P. I. (2023). Weak-to-strong generalization: Eliciting strong capabilities with weak supervision. . *arXiv preprint arXiv* .

Cook DJ, T. J. (2013). recovery in the elderly after major surgery: assessment of mobility recovery using wireless technology . *Ann Thorac Surg*.

D. Ciresan, U. M. (2012). Multi-column deep neural networks for image Classification . *in: Proceedings of the CVPR*.

D.C. Ciresan, U. M. (2012). Transfer learning for Latin and Chinese Characters with Deep Neural Networks . *in: Proceedings of the IJCNN*.

Dash, S. S. (2019). Big data in healthcare Management,analysis and future prospects. *Journal of Big Data*, 54.

Durrani , E. B., & Jain, E. D. (n.d.). Should I Be Scared Of First AI Software Engineer Or Not ? *DepartmentofComputerEngineering, ShriRam College Engineering & Management Banmore Madhya Pradesh*, 1.

eHealth, W.-b. a. (2019). EHealth adoption in primary healthcare in the EU is on The Rise. *European Comission*.

Eisen, M. (1998). Cluster analysis and display of genome-wide expression patterns. *Proc. Natl Acad. Sci. USA*, 95.

Ellis K., G. S. (2014). Identifying active travel behaviors in challenging environments using GPS, accelerometers, and machine learning algorithms. *Front. Public Health.*

Fourneret, E. (2020). The Hybridization of the Human with Brain. *Cambridge Quarterly of Healthcare Ethics*, 1.

Fraser Philp, A. F.‑N. (2021). Implications for the design of a Diagnostic Decision Support System (DDSS) to reduce time and cost to diagnosis in paediatric shoulder instability .

Freak-Poli RL, W. R. (2011). Participant characteristics associated with greater reductions in waist circumferenceduring a four-month, pedometer-based, workplace health program. *BMC Public Health*.

G. David, C. L. (2013). Economic Measurement of Medical Errors Using a Hospital Claims Database. *Value Health*, vol. 16,no. 2, pp. 305–310.

Gaidaa Maher Dogheim, A. H. (2023). Patient Care through AI-driven Remote Monitoring: Analyzing the Role of Predictive Models and Intelligent Alerts in Preventive Medicine. *Journal of Contemporary Healthcare Analytics*, 95.

Gasch, A. a. (2002). Exploring the conditional coregulation of yeast gene expression through fuzzy k-means clustering . *Genome Biol*, 1-22.

Ghahramani, Z. (September 16, 2004). Unsupervised Learning. *Gatsby Computational Neuroscience Unit,University College London, UK*.

Goodacre, R. e. (1998). Rapid identification of urinary tract infection bacteria using hyperspectral whole organism fingerprinting and artificial neural networks. *Microbiology*.

Hardika, B., Valencia, C., Adzani, R. M., Barus, I. R., & Fami, A. (2024). Students' Perceptions of the Future of Devin AI: A Case Study at IPB University. *Insearch information research journal*, 65.

Harvineet Singh, R. S. (2021). Fairness violations and mitigation under . *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*.

Harvineet Singh, V. M. (2022). Generalizability challenges of mortality risk prediction models: A retrospective analysis on a multi-center database. *PLOS Digital Health*.

J. Cortada, D. G. (jan. 2021). The value of analytics in healthcare. *From insights to outcomes*.

Jiménez, D. C. (jul-dec 2023). AI-Based Solutions for Enhancing Health Data Integration: Leveraging Machine Learning to Combine Electronic Health Records, Genomic Data, and Wearable Sensors. *Hong Kong Journal of AI and Medicine*, 17.

Jordi Laguarta, F. H. (2020). COVID-19 Artificial Intelligence Diagnosis Only Using Cough Recordings . 1.

Joseph Futoma, M. S.-V. (2020). the myth generalisability in clinical research and machine learning in health care. *The Lancet Digital Health* , e489–e492.

Kama Ramudu, D. V. (2023). CHALLENGES AND LIMITATIONS OF MEDICAL DISEASE IDENTIFICATION USING AI. *Contemporary Applications of Data Fusion for Advanced Healthcare Informatics* , 344.

Kaouk J.H., G. J. (2019). Robot-assisted surgery for benign distal ureteral strictures: step-by-step technique using the SP® surgical system. *BJU Int.* , 733–739.

Kaplan, N. e. (2004). A functional hierarchical organization of the protein sequence space . *BMC Bioinformatics,* .

Kaunitz, J. D. (2015). The Discovery of PCR: ProCuRement of Divine Power. *West Los Angeles VAMC and UCLA School of Medicine*.

Ku. Chhaya A. Khanzode, D. R. (January-April 2020). ADVANTAGES AND DISADVANTAGES OF MACHINE LEARNING: A LITERATURE REVIEW.

Lee E.J., K. Y. (2017). Artificial intelligence in stroke imaging. *Deep into the brain* , 19.

Leijnen S., v. V. (2020). The neural network zoo. *proceedings*.

Madeira, S. a. (2004). Biclustering algorithms for biological data analysis: a survey . *IEEE Trans. Comput. Biol. Bioinformatics,*.

Mahesh, B. (2018). Machine Learning Algorithms - A Review. *International Journal of Science and Research (IJSR)*.

Marco Sebastiani, C. V. (n.d.). Personalized Medicine and Machine Learning: A Roadmap for the Future. *Journal of clinical medicine*. doi: https://doi.org/10.3390/jcm11144110

Markus Bertl, P. R. (2020). A survey on AI and Decision Support Systems in Psychiatry- Uncovering a Dilemma.

Md Manjurul Ahsan, S. A. (15 march 2022). Machine-Learning-Based Disease Diagnosis: A Comprehensive Review. doi: https://doi.org/10.3390/healthcare10030541

Miotto R., W. F. (2018). Deep learning for healthcare: Review, opportunities and challenges. *Brief. Bioinform.* .

Mohamed Said Ibrahim, D. S. (2023). Machine Learning and Predictive Analytics: Advancing Disease Prevention in Healthcare. *Journal of Contemporary Healthcare Analytics*, 55.

Mohamed Said Ibrahim, S. S. (2023 ). Machine Learning and Predictive Analytics: Advancing Disease Prevention in Healthcare. *Journal of Contemporary Healthcare Analytics*, 57.

Mostafa Langarizadeh, F. M. (Nov 2016). Applying Naive Bayesian Networks to Disease Prediction: a Systematic Review. *Acta Inform Medica*. Retrieved from https://pmc.ncbi.nlm.nih.gov/articles/PMC5203736/

Office-based Physician Electronic Health Record Adoption. (2019). *Office of the National Coordinator for Health Information Technology*.

Patel S, P. H. (2012). A review of wearable sensors and systems with application in rehabilitation. *J Neuroeng Rehabil*.

Pushpa Singh, N. S. (23 April 2021). Diagnosing of disease using machine learning. 89-111. doi:https://doi.org/10.1016/B978-0-12-821229-5.00003-3

Qureshi, K., Din, S., Jeon, G., & Piccialli, F. (2020). An accurate and dynamic predictive model for a smart M-Health system using Machine Learning . *Inf. Sci.*

R., B. (2009). Activity monitoring in chronic obstructive pulmonary disease. *J Cardiopulm Rehabil*.

R., G. (May 28 2019 ). Introduction to machine learning algorithms: linear regression.

Rajpurkar P, I. J. (2017). Radiologist-level pneumonia detection on chest X-rays with deep learning. *cheXNet*, 3.

Rao S.R., D. C. (2011). Electronic health records in small physician practices: availability, use, and perceived benefits. *J. Am. Med. Inform. Assoc.*

Rizk, S. H. (2022). Ethical and Regulatory Challenges of Emerging Health Technologies. *applied ethics in the digital world* , 111.

Sandhya N. dhage, C. K. (2016). A review on Machine Learning Techniques. *In International Journal on Recent and Innovation Trends in Computing and communication* , Volume 4, issue 3.

Saria., A. S. (2020 ). From development to deployment: dataset shift, causality, and shift-stable models in health AI. *Biostatistics*, 345–352.

Sauter, V. L. (2017). Decision Support Systems for Business Intelligence. *John Wiley & Sons*.

Schumaker, E. (August 5, 2021, 5:00 AM). What is a HIPAA violation? Retrieved from https://www.beneschlaw.com/a/web/29158/What%20is%20a%20HIPAA%20violation\_%20-%20ABC%20News.pdf

Shailaja, K., Seetharamulu, B., & Jabbar, M. (march 29-31 2018). Machine learning in healthcare: A review. *In Proceedings of the 2018 Second Conference in Electronics Communication and Aerospace Technology (ICECA)*, 910-914.

Siddiqui M.K., M.-M. R. (2020). A review of epileptic seizure detection using machine learning classifiers. *Brain inform*.

Soubra, J. M. (n.d.). applications of machine learning in healthcare. Retrieved from https://forms.gle/eAmrennPMsgt7oGc7

Soubra, J. (November 23 2024). Applications of machine learning in healthcare. Retrieved from https://forms.gle/8FMM47KoGJjbsEA17

Steyvers, M., & Kumar, A. (2024). Three Challenges for AI-Assisted. *APS assosciation for psychological science*, 723.

Subbaswamy., S. S. (2019). Tutorial: safe and reliable machine learning . *arXiv preprint arXiv*.

Sutton R.S., B. A. (2018). Reinforcement Learning: An Introduction. *MIT Press*.

T. Mikolov, I. S. (2013). Distributed representations of words and phrases and their compositionality. *in: Proceedings of the NIPS*.

Tavazoie, S. e. (1999). Systematic determination of genetic network architecture . *Nat. Genet.*, 281-285.

Tian L., Z. D. (2020). Radiomics-based machine-learning method for prediction of distant metastasis from soft-tissue sarcomas. *Clin. Radiol*, 76.

Tudor-Locke C, L. L. (2009). Why do pedometers work? A reflection upon the factors related to successfully increasing physical activity. . *Sports Med.*

Turing, A. (1950). Turing A. Computing machinery and intelligence. Mind.

V. Saini et al. (July 2017). Drivers of poor medical care. *The Lancet, vol. 390, no. 10090, pp. 178–190,*, vol. 390, no. 10090, pp. 178–190,.

Vaes AW, C. A. (2013). . Effect of ‘activity monitor-based’ counseling on physical activity and health-related outcomes in patients with chronic diseases: a systematic view and meta-analysis. *Ann Med*, 397.

Wati D.A.R., A. D. (2017). Design of face detection and recognition system for smart home security application. *nd International conferences on Information Technology, Information Systems and Electrical Engineering (ICITISEE)*.

Wicks P, V. T. (2014). Subjects no more: what happens when trial participants realize they hold the power? . *Br Med J.*

Chan, F. T., Jiang, B., & Tang, N. K. (n.d.). The development of intelligent decision support tools to aid the. *Int. J. Production Economics 65 (2000) 73}84*, 73.

Makary , M. A., M. A. Makary and M. Daniel, & Daniel, M. (2016). the third leading cause of death in the US. *Medical error*, BMJ, vol. 353, doi: 10.1136/bmj.i2139.

] A. Bordes, X. G. (2012). Joint learning of words and meaning for open-text semantic parsing . *Proceedings of the AISTATS*.

Addis., T. (1956). Towards an” expert” diagnostic system. . *ICL Technical Journal* , 79-105.

ADITI SITE, J. N. (2021). Systematic Review on Machine-Learning Algorithms Used in Wearable-Based eHealth Data Analysis. *IEEE Access* .

al, A. A. (2020). ) Artifcial intelligence and machine learning to fight COVID-19. *American Physiological Society, Bethesda*.

Allet L, K. R. (2010). Wearable systems for monitoring mobility-related activities in chronic disease: a systematic review. *Sensors*.

Alloghani M., A.-J. D. (2020). A systematic review on supervised and unsupervised machine learning algorithms for data science.

Amitha Mathew, P. A. (2021). Deep Learning Techniques: An Overview. *Department of Computer Science and Engineering, School of Engineering, Avinashilingam*.

AN Ramesh, C. K. (2004). Artificial intelligence in medicine. *The University of Hull Academic Surgical Unit, Castle Hill Hospital, Cottingham, UK*, 334.

Ardebilib, A. T. (2020). Real-time RT-PCR in COVID-19 detection: issues affecting the results. *EXPERT REVIEW OF MOLECULAR DIAGNOSTICS*.

Artificial Intelligence. (Press, Cambridge, 2014. Accessed: Feb. 15, 2021. [Online]. Available). *Cambridge International Dictionary of English. Cambridge University press*.

Barwais FA, C. T. (2013). Physical activity, sedentary behavior and total wellness changes among sedentary adults: a 4-week randomized controlled trial . *Health Qual Life Outcomes*.

Bilu, Y. a. (2002). The advantage of functional prediction based on clustering of yeast genes and its correlation with non-sequence based classification. *J. Comput Biol* , 193–210.

Coates A., L. H. (2011). An analysis of single-layer networks in unsupervised feature learning. *J. Mach. Learn.*

Collin Burns, P. I. (2023). Weak-to-strong generalization: Eliciting strong capabilities with weak supervision. . *arXiv preprint arXiv* .

Cook DJ, T. J. (2013). recovery in the elderly after major surgery: assessment of mobility recovery using wireless technology . *Ann Thorac Surg*.

D. Ciresan, U. M. (2012). Multi-column deep neural networks for image Classification . *in: Proceedings of the CVPR*.

D.C. Ciresan, U. M. (2012). Transfer learning for Latin and Chinese Characters with Deep Neural Networks . *in: Proceedings of the IJCNN*.

Dash, S. S. (2019). Big data in healthcare Management,analysis and future prospects. *Journal of Big Data*, 54.

Durrani , E. B., & Jain, E. D. (n.d.). Should I Be Scared Of First AI Software Engineer Or Not ? *DepartmentofComputerEngineering, ShriRam College Engineering & Management Banmore Madhya Pradesh*, 1.

eHealth, W.-b. a. (2019). EHealth adoption in primary healthcare in the EU is on The Rise. *European Comission*.

Eisen, M. (1998). Cluster analysis and display of genome-wide expression patterns. *Proc. Natl Acad. Sci. USA*, 95.

Ellis K., G. S. (2014). Identifying active travel behaviors in challenging environments using GPS, accelerometers, and machine learning algorithms. *Front. Public Health.*

Fourneret, E. (2020). The Hybridization of the Human with Brain. *Cambridge Quarterly of Healthcare Ethics*, 1.

Fraser Philp, A. F.‑N. (2021). Implications for the design of a Diagnostic Decision Support System (DDSS) to reduce time and cost to diagnosis in paediatric shoulder instability .

Freak-Poli RL, W. R. (2011). Participant characteristics associated with greater reductions in waist circumferenceduring a four-month, pedometer-based, workplace health program. *BMC Public Health*.

G. David, C. L. (2013). Economic Measurement of Medical Errors Using a Hospital Claims Database. *Value Health*, vol. 16,no. 2, pp. 305–310.

Gaidaa Maher Dogheim, A. H. (2023). Patient Care through AI-driven Remote Monitoring: Analyzing the Role of Predictive Models and Intelligent Alerts in Preventive Medicine. *Journal of Contemporary Healthcare Analytics*, 95.

Gasch, A. a. (2002). Exploring the conditional coregulation of yeast gene expression through fuzzy k-means clustering . *Genome Biol*, 1-22.

Ghahramani, Z. (September 16, 2004). Unsupervised Learning. *Gatsby Computational Neuroscience Unit,University College London, UK*.

Goodacre, R. e. (1998). Rapid identification of urinary tract infection bacteria using hyperspectral whole organism fingerprinting and artificial neural networks. *Microbiology*.

Hardika, B., Valencia, C., Adzani, R. M., Barus, I. R., & Fami, A. (2024). Students' Perceptions of the Future of Devin AI: A Case Study at IPB University. *Insearch information research journal*, 65.

Harvineet Singh, R. S. (2021). Fairness violations and mitigation under . *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*.

Harvineet Singh, V. M. (2022). Generalizability challenges of mortality risk prediction models: A retrospective analysis on a multi-center database. *PLOS Digital Health*.

J. Cortada, D. G. (jan. 2021). The value of analytics in healthcare. *From insights to outcomes*.

Jiménez, D. C. (jul-dec 2023). AI-Based Solutions for Enhancing Health Data Integration: Leveraging Machine Learning to Combine Electronic Health Records, Genomic Data, and Wearable Sensors. *Hong Kong Journal of AI and Medicine*, 17.

Jordi Laguarta, F. H. (2020). COVID-19 Artificial Intelligence Diagnosis Only Using Cough Recordings . 1.

Joseph Futoma, M. S.-V. (2020). the myth generalisability in clinical research and machine learning in health care. *The Lancet Digital Health* , e489–e492.

Kama Ramudu, D. V. (2023). CHALLENGES AND LIMITATIONS OF MEDICAL DISEASE IDENTIFICATION USING AI. *Contemporary Applications of Data Fusion for Advanced Healthcare Informatics* , 344.

Kaouk J.H., G. J. (2019). Robot-assisted surgery for benign distal ureteral strictures: step-by-step technique using the SP® surgical system. *BJU Int.* , 733–739.

Kaplan, N. e. (2004). A functional hierarchical organization of the protein sequence space . *BMC Bioinformatics,* .

Kaunitz, J. D. (2015). The Discovery of PCR: ProCuRement of Divine Power. *West Los Angeles VAMC and UCLA School of Medicine*.

Ku. Chhaya A. Khanzode, D. R. (January-April 2020). ADVANTAGES AND DISADVANTAGES OF MACHINE LEARNING: A LITERATURE REVIEW.

Lee E.J., K. Y. (2017). Artificial intelligence in stroke imaging. *Deep into the brain* , 19.

Leijnen S., v. V. (2020). The neural network zoo. *proceedings*.

Madeira, S. a. (2004). Biclustering algorithms for biological data analysis: a survey . *IEEE Trans. Comput. Biol. Bioinformatics,*.

Mahesh, B. (2018). Machine Learning Algorithms - A Review. *International Journal of Science and Research (IJSR)*.

Marco Sebastiani, C. V. (n.d.). Personalized Medicine and Machine Learning: A Roadmap for the Future. *Journal of clinical medicine*. doi: https://doi.org/10.3390/jcm11144110

Markus Bertl, P. R. (2020). A survey on AI and Decision Support Systems in Psychiatry- Uncovering a Dilemma.

Md Manjurul Ahsan, S. A. (15 march 2022). Machine-Learning-Based Disease Diagnosis: A Comprehensive Review. doi: https://doi.org/10.3390/healthcare10030541

Miotto R., W. F. (2018). Deep learning for healthcare: Review, opportunities and challenges. *Brief. Bioinform.* .

Mohamed Said Ibrahim, D. S. (2023). Machine Learning and Predictive Analytics: Advancing Disease Prevention in Healthcare. *Journal of Contemporary Healthcare Analytics*, 55.

Mohamed Said Ibrahim, S. S. (2023 ). Machine Learning and Predictive Analytics: Advancing Disease Prevention in Healthcare. *Journal of Contemporary Healthcare Analytics*, 57.

Mostafa Langarizadeh, F. M. (Nov 2016). Applying Naive Bayesian Networks to Disease Prediction: a Systematic Review. *Acta Inform Medica*. Retrieved from https://pmc.ncbi.nlm.nih.gov/articles/PMC5203736/

Office-based Physician Electronic Health Record Adoption. (2019). *Office of the National Coordinator for Health Information Technology*.

Patel S, P. H. (2012). A review of wearable sensors and systems with application in rehabilitation. *J Neuroeng Rehabil*.

Pushpa Singh, N. S. (23 April 2021). Diagnosing of disease using machine learning. 89-111. doi:https://doi.org/10.1016/B978-0-12-821229-5.00003-3

Qureshi, K., Din, S., Jeon, G., & Piccialli, F. (2020). An accurate and dynamic predictive model for a smart M-Health system using Machine Learning . *Inf. Sci.*

R., B. (2009). Activity monitoring in chronic obstructive pulmonary disease. *J Cardiopulm Rehabil*.

R., G. (May 28 2019 ). Introduction to machine learning algorithms: linear regression.

Rajpurkar P, I. J. (2017). Radiologist-level pneumonia detection on chest X-rays with deep learning. *cheXNet*, 3.

Rao S.R., D. C. (2011). Electronic health records in small physician practices: availability, use, and perceived benefits. *J. Am. Med. Inform. Assoc.*

Rizk, S. H. (2022). Ethical and Regulatory Challenges of Emerging Health Technologies. *applied ethics in the digital world* , 111.

Sandhya N. dhage, C. K. (2016). A review on Machine Learning Techniques. *In International Journal on Recent and Innovation Trends in Computing and communication* , Volume 4, issue 3.

Saria., A. S. (2020 ). From development to deployment: dataset shift, causality, and shift-stable models in health AI. *Biostatistics*, 345–352.

Sauter, V. L. (2017). Decision Support Systems for Business Intelligence. *John Wiley & Sons*.

Schumaker, E. (August 5, 2021, 5:00 AM). What is a HIPAA violation? Retrieved from https://www.beneschlaw.com/a/web/29158/What%20is%20a%20HIPAA%20violation\_%20-%20ABC%20News.pdf

Shailaja, K., Seetharamulu, B., & Jabbar, M. (march 29-31 2018). Machine learning in healthcare: A review. *In Proceedings of the 2018 Second Conference in Electronics Communication and Aerospace Technology (ICECA)*, 910-914.

Siddiqui M.K., M.-M. R. (2020). A review of epileptic seizure detection using machine learning classifiers. *Brain inform*.

Soubra, J. M. (n.d.). applications of machine learning in healthcare. Retrieved from https://forms.gle/eAmrennPMsgt7oGc7

Soubra, J. (November 23 2024). Applications of machine learning in healthcare. Retrieved from https://forms.gle/8FMM47KoGJjbsEA17

Steyvers, M., & Kumar, A. (2024). Three Challenges for AI-Assisted. *APS assosciation for psychological science*, 723.

Subbaswamy., S. S. (2019). Tutorial: safe and reliable machine learning . *arXiv preprint arXiv*.

Sutton R.S., B. A. (2018). Reinforcement Learning: An Introduction. *MIT Press*.

T. Mikolov, I. S. (2013). Distributed representations of words and phrases and their compositionality. *in: Proceedings of the NIPS*.

Tavazoie, S. e. (1999). Systematic determination of genetic network architecture . *Nat. Genet.*, 281-285.

Tian L., Z. D. (2020). Radiomics-based machine-learning method for prediction of distant metastasis from soft-tissue sarcomas. *Clin. Radiol*, 76.

Tudor-Locke C, L. L. (2009). Why do pedometers work? A reflection upon the factors related to successfully increasing physical activity. . *Sports Med.*

Turing, A. (1950). Turing A. Computing machinery and intelligence. Mind.

V. Saini et al. (July 2017). Drivers of poor medical care. *The Lancet, vol. 390, no. 10090, pp. 178–190,*, vol. 390, no. 10090, pp. 178–190,.

Vaes AW, C. A. (2013). . Effect of ‘activity monitor-based’ counseling on physical activity and health-related outcomes in patients with chronic diseases: a systematic view and meta-analysis. *Ann Med*, 397.

Wati D.A.R., A. D. (2017). Design of face detection and recognition system for smart home security application. *nd International conferences on Information Technology, Information Systems and Electrical Engineering (ICITISEE)*.

Wicks P, V. T. (2014). Subjects no more: what happens when trial participants realize they hold the power? . *Br Med J.*

## 

## 