National College of Ireland

Classification Models for Improving Identification of Heart Diseases in Healthcare Industry : Eastern Europe

MSc Research Project
MSc. in Data Analytics

Omkar Terdal Ramesh
Student ID: x16104439

School of Computing
National College of Ireland

Supervisor: Dr. Catherine Mulwa

National College of Ireland
Project Submission Sheet
School of Computing

National College of Ireland

Student Name:	Omkar Terdal Ramesh
Student ID:	x16104439
Programme:	MSc. in Data Analytics
Year:	2020
Module:	MSc Research Project
Supervisor:	Dr. Catherine Mulwa
Submission Due Date:	17/08/2020
Project Title:	Classification Models for Improving Identification of Heart Diseases in Healthcare Industry : Eastern Europe
Word Count:	XXX
Page Count:	17

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	
Date:	19th August 2020

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Classification Models for Improving Identification of Heart Diseases in Healthcare Industry : Eastern Europe

Omkar Terdal Ramesh
x16104439

Abstract

The aim of this project is to differentiate between types of heart diseases from the same dataset using classification modelling.This project proposes to use the latest machine learning methods to accurately predict the type of cardiovascular disease in a patient with symptoms and previous medical history.Numerous classification models and machine learning approaches are proposed to be used for the classification of cardiovascular diseases.

1 Introduction

Heart is one of the most important organs in the human body. Heart diseases or also known as cardiovascular diseases is the one of the major proportions of death causes around the globe. Which makes heart diseases high priority for diagnosing and preventive measures. With the application of machine learning methods to accelerate classification and preventive measure for causes of cardiovascular diseases would benefit the whole world. Automation of such procedures for identifying the particular diseases need to be very precise and accurate as it is a very delicate organ and any wrong measures may lead to the loss of life. This project proposes to automate the process of classifying different heart diseases based on patient symptoms and medical history while using machine learning and data mining methods to achieve the optimum accuracy. This current paper is organized as follows. Section 2 contains previous work related to the subject. Section 3 contains the methodolgy proposed to be used for the flow of the project. Section 4 includes the types of models to be used and the results which are hoped to achieve with this project. Section 5 provides the results and comparison of the models. Section 6 contains conclusion and future works of the project.

1.1 Motivation and Project Background

Machine learning is the process where models are developed learn from data without actually being programmed to. This has led the methodologies to improve so much that machine learning is being used in every known field to make significant changes. Machine learning methods provide the most accurate results based on computation without rigorous use of manpower. The advancements it brought to the medicine industry led to early

detection of diseases and accurate diagnosis methods to prevent wrong treatments and help medicine rely more on the upcoming technologies to be implemented for the betterment of people's lifestyle and the businesses involved. Heart diseases are one among the most highly rated death causes globally and machine learning seeks to provide early detection in heart diseases and symptoms related to them. This will revolutionize the way the heart diseases will be classified and treated in the future.

In recent years, there have been numerous advanced researches taking place in the field of medicine and some of them mostly on the subject of heart diseases and how to take preventive measures to avoid them. A hybrid approach was used used to classify heart diseases based on a dataset from Cleveland. By making use of Artificial Neural networks(ANN), decision trees and Support vector Machine(SVM) they were able to improve the accuracy of predictions by 80 percent by using well known classifiers. But this study showed a slight preference towards patients already diagnosed with cardiovascular diseases and stated that patients getting checked for the first time should have a subset of features which should massively reduce the time for diagnosis of symptoms.(Iftikhar et al.; 2017) And this can only be obtained with high level of accuracy.

Continuous health monitoring systems is one of the many applications enabled by Internet of Things(IoT), where the patients were insisted to wear devices to monitor their status continuously, The data was then to be stored in a Hbase due to the large volumes of incoming data of the wearable sensor. These sensors were connected to the cloud for continuous income of data and monitoring issues but the sensors were unable to pick up any new symptoms or read any data if the device was removed. Though the research was conducted continuously there were issues relating to the efficiency and accuracy of such devices.(Kumar and Gandhi; 2017)

A knowledge based system was used to predict different types of diseases using hybrid intelligent methods on real world data sets. But each individual task had to be separated to make sure the optimum accuracy was reached in predicting each type of disease. This reduced the time optimization for correlating each of the diseases and detecting them alongside each other. By using fuzzy-ruled based techniques they were able to achieve enough accuracy on each type of disease and the prior symptoms which caused it.(Nilashi et al.; 2017)

From the above literature, there were certain drawbacks noticed in terms of wearable devices able to predict unknown outcomes and also unable to clearly classify disease based on historical patient data. With this in mind, this project proposes to address the real world data in such a way that optimum accuracy is always obtained on the automated system even without any external factors.

1.2 Research Question

RQ "Can predictive modelling can improve classification of cardiovascular diseases or other abnormalities in the heart based on patient history and previous known symptoms ?"

This research has provided a methodological way to diagnose patients based on their symptoms and history for chronological cardiovascular diseases. Which can in turn be

implemented into other types of detection to improve diagnosis time frame and preventive measures.

1.3 Project Objectives

The objective of this project to improve classification time frame of the Heart diseases in Eastern Europe with the latest advancements in classification models and machine learning methods. This will also help in identifying early symptoms of patients so future illnesses can be prevented. This project will also try to overcome any repetitive data to get the most accurate results. The main project objectives are as stated in the Figure 1 below.

Chapters	Objectives
1	Introduction
1.1	Motivation and Project Background
1.2	Research Question
1.3	Project Objectives
2	Critical Review of Heart Disease detection
2.1	Introduction
2.2	Literature review
2.3	Conclusion
3	Methodology
3.1	Introduction
3.2	Data Collection
3.3	Data Preprocessing
3.4	Project design Flow
4	Implementation ,Evaluation and Results.
4.1	Introduction
4.2	Implementation ,Evaluation and Results of ANN.
4.3	Implementation, Evaluation and Results of RF and SVM.
5	Comparison of Developed Models
6	Conclusion and Future work.

Figure 1: Project Objectives

2 A Critical Review of Heart Disease Detection and Identified gaps (2014-2019)

2.1 Introduction

In this section we will critically evaluate and identify the types of studies conducted on detection of heart diseases. The section will be divided into multiple subsections to identify the different types of technology used for the detection of heart diseases. Subsection 2.2 reviews the artificial intelligence and Machine learning studies. Subsection 2.3 will review the different neural network approaches applied in the field.

2.2 A Literature Review on Heart Disease Detection using AI and Machine learning techniques

There are numerous studies conducted in accordance with cardiovascular diseases using Machine learning Algorithms, Artificial Intelligence and so on. The European Society of Cardiology states that 3.6 million people are diagnosed with heart diseases every year, Most of which die within the span of 1-2 years(Haq et al.; 2018). Heart disease patients have a lot of factors including blood pressure, diabetes, smoking and so on which can help in more effective diagnosis (Murthy and Meenakshi; 2014). Annual health reports from W.H.O shows that nearly 17.8 million people died due to heart attacks in 2008 worldwide. This shows that heart attacks are the number one cause of deaths globally and the most number of deaths than any other disease (Ravish et al.; 2014).

In the modern society, each individual is conscious about their personal safety and well being. The number of increasing studies and diagnostics found also implies that an average person has many ways to contract a disease. A large amount of data which is being generated by the health industry is not being effectively used. Approached such as Machine learning algorithms and Artificial intelligence will help us in effectively using the data to yield more useful results and also help with diagnosis of heart diseases. Using these technologies will help in providing highly accurate prediction and classification of heart diseases (Zhang and Han; 2017).

Over the past few years multiple studies have been conduction on the Cleveland data set to accurately predict the diagnosis and decide which method or study could benefit the healthcare industry. Numerous studies have shown that just using distinct methods will not be satisfactory enough to help in the effective prediction of diagnosis, Some studies also included that using not only distinct methods but to also converge the techniques using the similarities to form Hybrid methods (Rathnayakc and Ganegoda; 2018). These Hybrid methods should help in identifying the drawbacks of the distinct methods, while these Hybrid methods can be used to enhance the results obtained through the distinct methods. But the study shows that enhancing the models by layering not only improved the overall results but also gave ideas that models can be layered on each other for different purposes and can be used to predict different types of outcomes based on the specific industry. This not only helps the health industry but can also show significant improvements through the development of future models.

A large number of risks factors are involved in the prediction of heart diseases, though most of them are not used while the models are executed. This is to reduce the attributes which might affect the accuracy. So to reduce the dimensionality, the are some attribute selection methods which can used such as Chi-squared (Palaniappan and Awang; 2008). As observed in the studies above that most of them used the Cleveland dataset for their models and some of them used data from W.H.O. most of the data would not have similar attributes which all the studies can be compared to. According to the studies we need to make sure the data we collect needs to have all the relatable attributes so that the studies can be compared to each other so that future studies will be able to benefit from these studies (Pouriyeh et al.; 2017).

Numerous studies were conducted in accordance with cardiovascular diseases using data mining technologies which include neural networks, linear regression, ANN, Machine learning algorithms and so on. These studies were inclined to improve the accuracy that previously recorded results from different approaches that were unsatisfactory and can be further improved upon (Kim and Kang; 2017). Hadoop, a Big data processing system when merged with machine learning methods has been the preferred platform to conduct such studies as they can handle large datasets and can process results without too many instructions.That's where Apache spark came in as the second generation processing system for big data (Gebara et al.; 2015).

The increasing number of diagnostics discovered due to the intervention of technology into the field of medical science has proven the fact there are many ways for a person to be infected by a disease. The intervention of technology has helped us to identify the problem beforehand or detect the symptoms earlier. Continuous monitoring of one's health is a growing trend in all age groups as the growing levels of awareness leads to the prevention of life threatening diseases such as cancer or heart diseases that can greatly improve the life span of individuals (Nair et al.; 2017).

Data mining uses different techniques for different problems, these techniques helps in identifying the underlying hidden patterns in data. These techniques are supervised learning, unsupervised learning and semi-supervised learning. It can be described as the discovery of pattern recognition and useful information from large amounts of data (Han et al.; 2011). The major cause of deaths related to cardiovascular diseases is the Ishemic heart disease. Early detection of its symptoms may lead to prolonged life span and reduce mortality rates (Tantimongcolwat et al.; 2008). Electrocardiograpghy (ECG) is used to monitor heart malfunctions which usually arises from improper blood flow to the heart. This allows the practitioner to view the activity inside the heart through electrodes attached to the patient. Magnetcardiography (MCG) is a technique which involves monitoring the magnetic field radiated by the cardiac tissues. This was replaced with ECG due to contact free probes Kanzaki et al. (2003). With the usage of MCG coronary disease can be detected in patients with acute chest pains (Park et al.; 2005).

In (Chitra and Seenivasagam; 2013), the study shows the development of a computer aided heart disease prediction system which was intended to help increase the time frame of diagnosis for heart diseases. Their results showed that data mining can play a major role in the classification of heart diseases.Due to its nature, feature reduction techniques

maybe applied to improve accuracy and search time for classification. (Bhatla and Jyoti; 2012) says that application of data mining techniques have been in close relation to diagnosis of heart diseases in recent times. The work of (Srinivas et al.; 2010) states of data mining exploration aspects which went beyond classification scope, clustering mostly, association rule mining and time series analysis. The purpose of this project will be to integrate 3 different types of techniques to evaluate the best accuracy deliverable in order to classify heart diseases.

Being overweight and Obesity is on the rise worldwide, due to the source of food and lack of health consciousness. This can also lead to heart diseases. The physical state of individuals is strongly influenced in the early stages of their life (Fergus et al.; 2017). Encouraging children to be more active and healthy is of great interest to researchers and policy companies who compare risk levels based on the upbringing (Oyebode and Mindell; 2013)

2.3 Conclusion

From the above literature we see that there are numerous ways for individuals to come up with cardiovascular diseases and researches which have helped in pursuing the quicker diagnosis of such diseases. Each research was solely intended for a particular group and not as a whole. These researches targeted specific symptoms to diagnose. From this we learn that there is still room for development in this field and with the proposed project we will contribute a little to the advancements in the research field.

3 Methodology

3.1 Introduction

We have used SDLC (Software Development Life Cycle) methodology for the find the highest achievable accuracy for identification of heart diseases. The purpose of this project is to identify existing issues and implement machine learning algorithms and Artificial intelligence methods for the detection of heart diseases. This implemented methodology can be followed through with the steps shown below in Figure2. These steps are further described in the subsections.

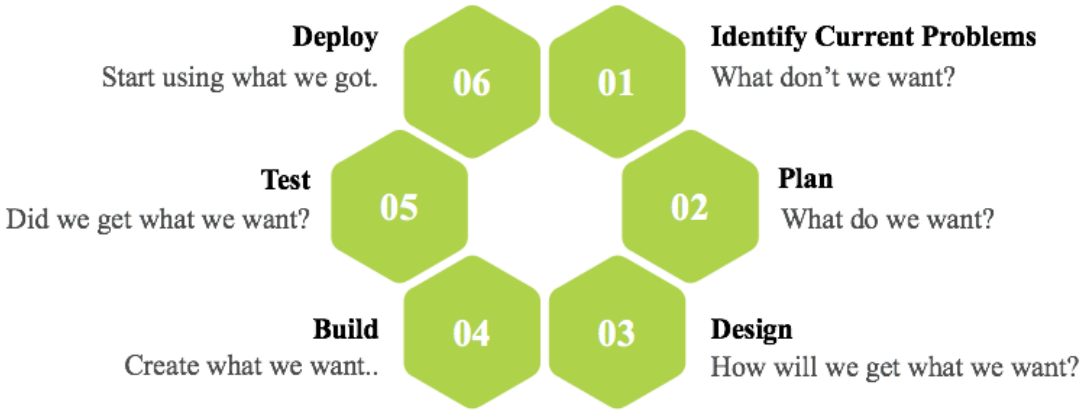


Figure 2: Identification of heart Diseases Methodology

3.2 Data collection

The Identification of heart diseases dataset is collected from a database from India. The Dataset consists of over 4000 rows of data with over 10 columns of attributes. The data was segregated to 3000 rows of training data and the rest for test data. They were segregated in a .csv file.

3.3 Data Preprocessing

Data preprocessing is a step we take to train the data models to yield more accuracy in the results. During this we need to go through the data to make sure we have no missing values, irregularities. We filtered all the rows and columns to remove unfilled data and segregated the data for training and testing purposes. WE have taken care of such issues in our proposed project.

3.4 Project Design Process Flow

The design of the proposed project has a logic built business model which is used to extensively train data for the identification of heart heart disease. These steps involve Planning, Implementation and Analysis. The data will be trained with one artificial intelligence neural network method and two machine learning algorithms to compare the

accuracy of each method. The whole project is deployed on the Python. The results will also be plotted on Python.Project design process flown as shown in 3

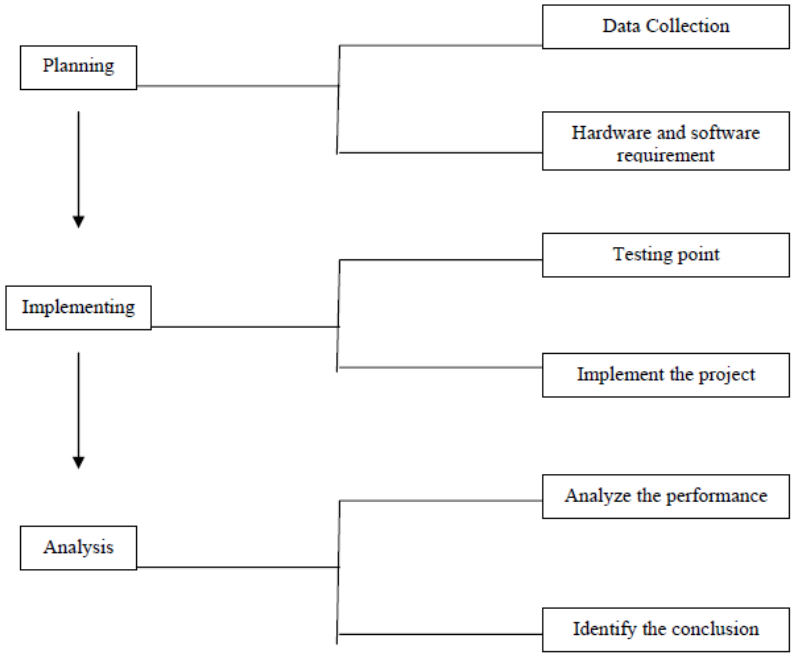


Figure 3: Project Design Process Flow

3.4.1 Planning

Planning is a stage where all the information and requirements such as hardware and software are obtained and installed appropriately. The planning phase has two main elements namely Data collection and meeting the requirements.

3.4.2 Implementing

In the Implementaton phase we develop a business intelligent model, to classify different attributes based on the business model structure using a suitable machine learning technique.The model was evaluated by a scientific approach to measure accuracy. We are using Artificial neural network, Random Forest and SVM models for our proposed project.

3.4.3 Analysis

In the final Analysis phase, We test our classification model on the heart disease dataset and also measure the performance on our dataset. To carefully evaluate the performance of the classification model and make it comparable to other approaches, we used accuracy to measure the effectiveness of the classifiers. The primary source of a performance measurements is a coincidence matrix(classification matrix or contingency table) The Figure below shows a coincidence matrix or a two-class classification problem. The equations of the most commonly used metrics that can be calculated from the coincidence matrix are also displayed.

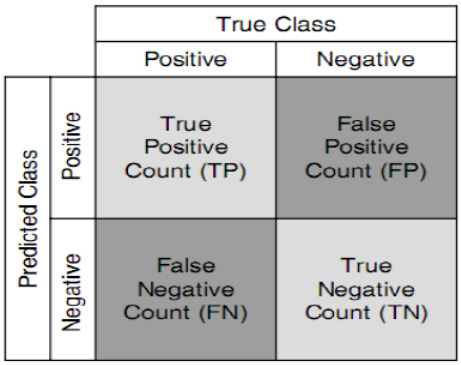


Figure 4: Confusion Matrix

$$\text{True Positive Rate} = \frac{TP}{TP + FN}$$

$$\text{True Negative Rate} = \frac{TN}{TN + FP}$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

Figure 5: Confusion Matrix Formulae

4 Implementation, Evaluation and Results of Heart Disease Prediction Models

4.1 Introduction

This Section discusses the implementation, evaluation and outcomes of different models used to identify heart diseases. The different machine learning models such as Random Forest and SVM, and also using Artificial Neural networks. These models will have pre–trained model weights, and can perform attribute extraction with the help of network layers. The proposed project can be classified as a multi-class classification problem.

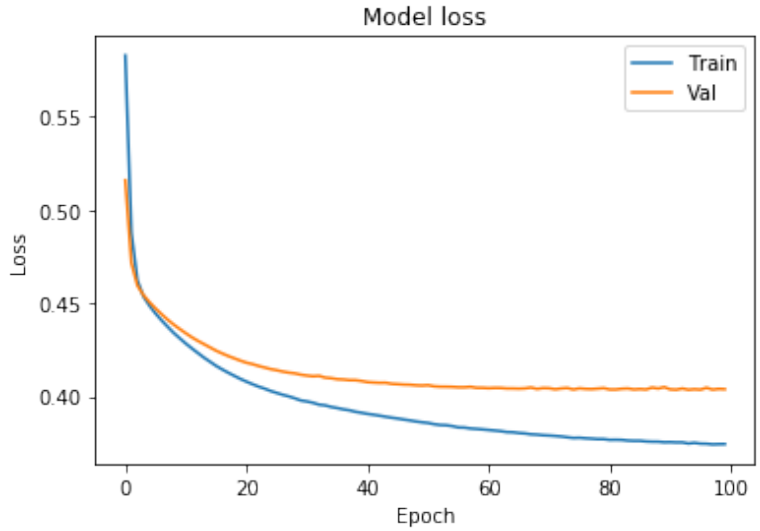
4.2 Implementation, Evaluation and Results of ANN

Artificial Neural network (ANN) is a computing system designed to simulate the way a human brain analyzes and processes information. ANN models are used where problems seem impossible to solve by human or statistical standard. ANN has self learning capabilities and hence will produce better results as more data is allotted to them. We use

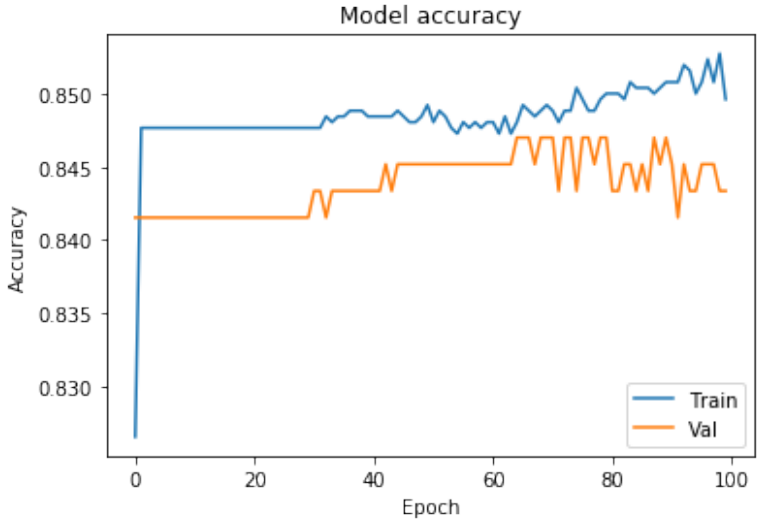
Python language to execute these proposed models. The ANN model can be implemented with the help of keras and tensorflow libraries.

4.2.1 Evaluation and Results of ANN

Total of 4241 rows of data were preprocessed and segregated into training and test data. We then run jupyter notebook to establish connection to python libraries and execute the model. To evaluate the performance the proposed model uses metrics such as accuracy, loss, recall, precision and f1-score. The model loss and model accuracy graphs are as shown below in the Figure 6a and Figure 6b.



(a)



(b) ANN Model Loss Graph And Model Accuracy Graph

The ANN models shows that accuracy achieved by this model is 85.61. Though the loss analysed is around 40 percent. The precision percent was observed to be 86 percent. The F1-score to be around 91 percent. The results for the ANN model yielded the best accuracy compared to the other models. We observe the significant reduction in loss after

every epoch. The accuracy of the model can be increased by more number of epochs or by obtaining a larger dataset with similar attributes. The ANN model learns in detail.

4.3 Evaluation and Results of Random Forest and SVM

Before we deploy the machine learning models we need to preprocess and train the data so we can feed it to the ML models. With ML models the data is made sure that it has all the attributes from the source and each one is included for the model to process. In the figure below we see how the attributes are segregated within the models.

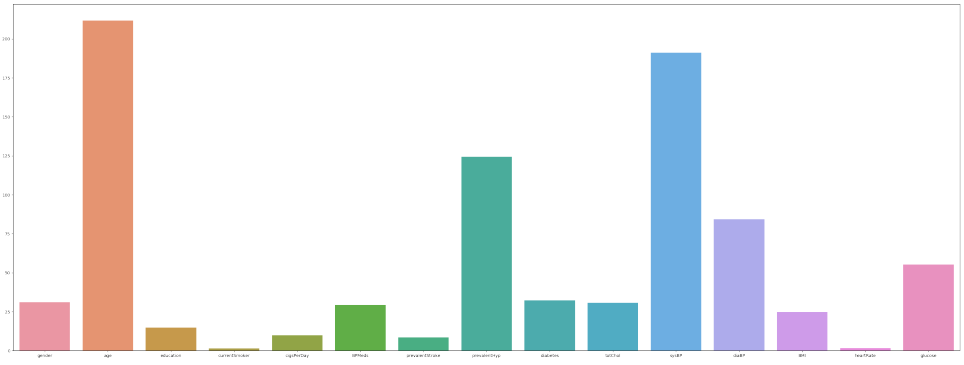


Figure 7: Segregated attribtes

We go ahead and segregate the gender attribute to calculate the number of instances encountered in each gender while also making sure the histogram shows the plot with gender affected with heart disease. The following 8 will display both of them.

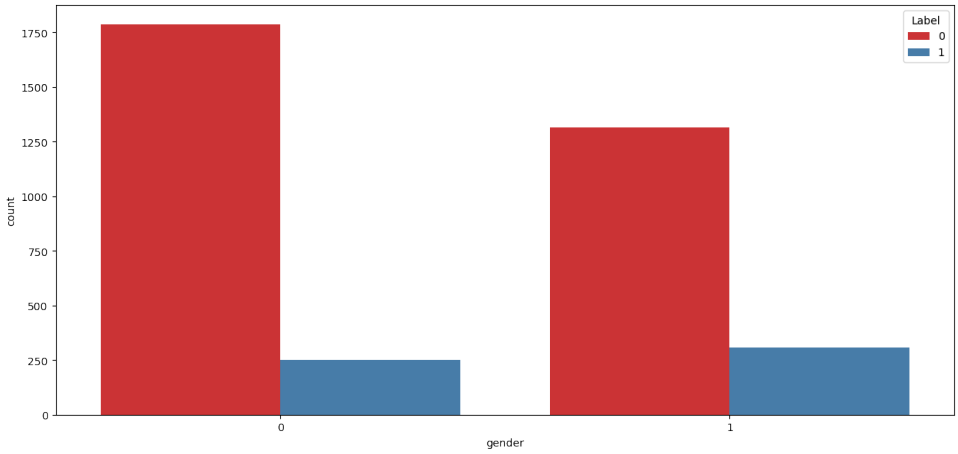


Figure 8: Gender Segregation

We further more segregate the data to know the age groups and the amount of infected people in these age groups. By doing so we will be easily able to feed the generic data to ML models to compute the accuracy. The age based distribution of data can be found in the following 9

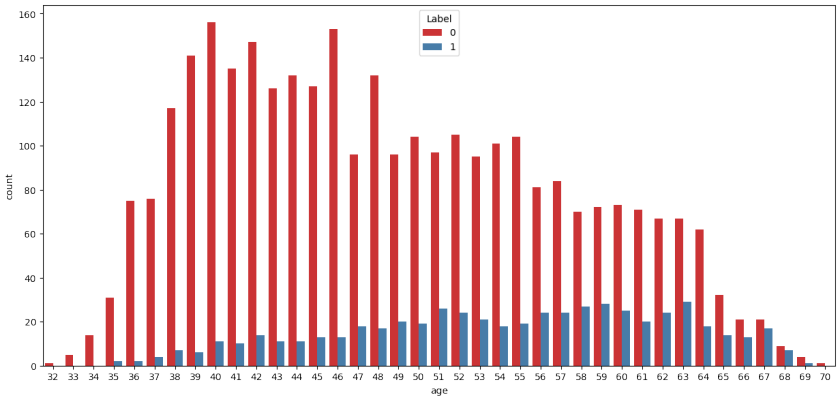


Figure 9: Age Distribution

We then proceed to train and evaluate the results of the Random Forest Machine Learning algorithm. By evaluating the results of the model we will be able to compare results to the outcome of ANN. The Random Forest method yielded 85.15 percent accuracy. While the precision is also 85 percent. The F1-score rose to 92 percent.The following 10 shows the results of the model.

```
In [39]: print(classification_report(y_test,predictions_rf))
```

	precision	recall	f1-score	support
0	0.85	0.99	0.92	931
1	0.64	0.05	0.10	167
accuracy			0.85	1098
macro avg	0.75	0.52	0.51	1098
weighted avg	0.82	0.85	0.79	1098

Figure 10: Radom Forest Model Results

We proceed to train the SVM model next and evaluate the results of the models. The SVM model yielded 84.79 percent accuracy. Though the accuracy is closer to Random forest, It is quite lower compared to the ANN model. The precision of SVM model is also 85 percent and the f1-score is 92 percent. The Following 11 displays the results of the SVM Model.

```
In [48]: print(classification_report(y_test,predictions_svm))
```

	precision	recall	f1-score	support
0	0.85	1.00	0.92	931
1	0.00	0.00	0.00	167
accuracy			0.85	1098
macro avg	0.42	0.50	0.46	1098
weighted avg	0.72	0.85	0.78	1098

Figure 11: SVM Model Results

5 Comparison of Developed Models

In this project, we have performed a comparative analysis of an Artificial Neural network and two other Machine learning algorithms mainly Random Forest and SVM. All of the objectives from Chapter 1, subsection 3 were completed. These models were evaluated based on the test dataset and the performace was improvised. The ANN yielded the most accuracy of all the three models at 85.61 percent. The f1-score for all three models remained the same at 0.92. Only the ANN model has the recall rate of 1.00. The following Figure12 shows the comparison of precision between all three models. The results of these predictive models has solved the research question from Chapter 1, subsection 1.2.

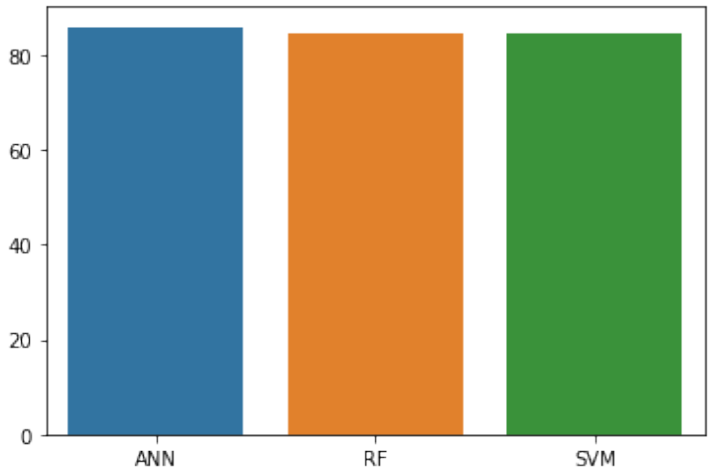


Figure 12: Comparison of Deployed Models

6 Conclusion and Future work

To identify heart diseases and its multiple factors, we have used deep learning Artificial neural network with many predefined structures. The dataset has been tested with ANN and different machine learning algorithms namely SVM and Random Forest. By doing so we have met all the objective requirements for the project from Chapter 1, subsection 1.3. Though the accuracy of each model was averaged at 80 percent. By delving in deeper we found out that we required more factors or attributes to further enhance the accuracy of these models. The results achieved with these implementations has solved the research question stated in Chapter 1, subsection 1.2. Due to the number of attributes involved the behaviour of all the models are similar. Performing this project we have observed that with a larger dataset and more attributes we can further enhance the accuracy by increasing the training data and number of epochs. In Future work, we will try to minimise the loss by adjusting the parameters based on needs which also helps in increasing the accuracy of the models.

7 Acknowledgement

I would like to thank Dr.Catherine Mulwa for all the guidance and support provided throughout these tough times with supervision sessions for doing the project. I would

also like to thank my Mother, Father and Brother for all the support and trust in me.

References

Bhatla, N. and Jyoti, K. (2012). An analysis of heart disease prediction using different data mining techniques, *International Journal of Engineering* **1**(8): 1–4.

Chitra, R. and Seenivasagam, V. (2013). Review of heart disease prediction system using data mining and hybrid intelligent techniques, *Journal on Soft Computing (ICTACT)* **3**(4): 605–609.

Fergus, P., Hussain, A. J., Hearty, J., Fairclough, S., Boddy, L., Mackintosh, K., Stratton, G., Ridgers, N., Al-Jumeily, D., Aljaaf, A. J. et al. (2017). A machine learning approach to measure and monitor physical activity in children, *Neurocomputing* **228**: 220–230.

Gebara, F. H., Hofstee, H. P. and Nowka, K. J. (2015). Second-generation big data systems, *Computer* **48**(1): 36–41.

Han, J., Pei, J. and Kamber, M. (2011). *Data mining: concepts and techniques*, Elsevier.

Haq, A. U., Li, J. P., Memon, M. H., Nazir, S. and Sun, R. (2018). A hybrid intelligent system framework for the prediction of heart disease using machine learning algorithms, *Mobile Information Systems* **2018**.

Iftikhar, S., Fatima, K., Rehman, A., Almazyad, A. S. and Saba, T. (2017). An evolution based hybrid approach for heart diseases classification and associated risk factors identification, *Biomedical Research* **28**(8).

Kanzaki, H., Nakatani, S., Kandori, A., Tsukada, K. and Miyatake, K. (2003). A new screening method to diagnose coronary artery disease using multichannel magnetocardiogram and simple exercise, *Basic research in cardiology* **98**(2): 124–132.

Kim, J. K. and Kang, S. (2017). Neural network-based coronary heart disease risk prediction using feature correlation analysis, *Journal of healthcare engineering* **2017**.

Kumar, P. M. and Gandhi, U. D. (2017). A novel three-tier internet of things architecture with machine learning algorithm for early detection of heart diseases, *Computers & Electrical Engineering* .

Murthy, H. N. and Meenakshi, M. (2014). Dimensionality reduction using neuro-genetic approach for early prediction of coronary heart disease, *International Conference on Circuits, Communication, Control and Computing*, IEEE, pp. 329–332.

Nair, L. R., Shetty, S. D. and Shetty, S. D. (2017). Applying spark based machine learning model on streaming big data for health status prediction, *Computers & Electrical Engineering* .

Nilashi, M., bin Ibrahim, O., Ahmadi, H. and Shahmoradi, L. (2017). An analytical method for diseases prediction using machine learning techniques, *Computers & Chemical Engineering* .

Oyebode, O. and Mindell, J. (2013). Use of data from the health survey for england in obesity policy making and monitoring, *Obesity Reviews* **14**(6): 463–476.

Palaniappan, S. and Awang, R. (2008). Intelligent heart disease prediction system using data mining techniques, *2008 IEEE/ACS international conference on computer systems and applications*, IEEE, pp. 108–115.

Park, J.-W., Hill, P. M., Chung, N., Hugenholtz, P. G. and Jung, F. (2005). Magnetocardiography predicts coronary artery disease in patients with acute chest pain, *Annals of noninvasive electrocardiology* **10**(3): 312–323.

Pouriyeh, S., Vahid, S., Sannino, G., De Pietro, G., Arabnia, H. and Gutierrez, J. (2017). A comprehensive investigation and comparison of machine learning techniques in the domain of heart disease, *2017 IEEE Symposium on Computers and Communications (ISCC)*, IEEE, pp. 204–207.

Rathnayakc, B. S. S. and Ganegoda, G. U. (2018). Heart diseases prediction with data mining and neural network techniques, *2018 3rd International Conference for Convergence in Technology (I2CT)*, IEEE, pp. 1–6.

Ravish, D., Shanthi, K., Shenoy, N. R. and Nisargh, S. (2014). Heart function monitoring, prediction and prevention of heart attacks: Using artificial neural networks, *2014 International Conference on Contemporary Computing and Informatics (IC3I)*, IEEE, pp. 1–6.

Srinivas, K., Rao, G. R. and Govardhan, A. (2010). Analysis of coronary heart disease and prediction of heart attack in coal mining regions using data mining techniques, *Computer Science and Education (ICCSE), 2010 5th International Conference on*, IEEE, pp. 1344–1349.

Tantimongcolwat, T., Naenna, T., Isarankura-Na-Ayudhya, C., Embrechts, M. J. and Prachayasittikul, V. (2008). Identification of ischemic heart disease via machine learning analysis on magnetocardiograms, *Computers in biology and medicine* **38**(7): 817–825.

Zhang, W. and Han, J. (2017). Towards heart sound classification without segmentation using convolutional neural network, *2017 Computing in Cardiology (CinC)*, IEEE, pp. 1–4.