**Heart Disease Prediction Project**

**PHASE II REPORT**

***Submitted by***

Supervisor:- Ms. Tanu Dhiman (E8141)

Co-Supervisor :- Ms. Neha Dutta (E12830)

**NAME OF THE CANDIDATE(S)**

**NAME - UID**

ASHI SHARMA - 20BCS7931

REETA DEVI - 20BCS2109

ADITYA SHARMA - 20BCS9250

***in partial fulfillment for the award of the degree of***

**BACHELOR OF ENGINEERING**

**IN**

**COMPUTER SCIENCE & ENGINEERING**

****

**Abstract**

The paper focuses on the construction of an artificial intelligence-based heart disease detection system using machine learning algorithms. We show how machine learning can help predict whether a person will develop heart disease. In this paper, a python-based application is developed for healthcare research as it is more reliable and helps track and establish different types of [health monitoring](https://www.sciencedirect.com/topics/mathematics/health-monitoring) applications. We present data processing that entails working with [categorical variables](https://www.sciencedirect.com/topics/mathematics/categorical-variable) and conversion of categorical columns. We describe the main phases of application developments: collecting databases, performing [logistic regression](https://www.sciencedirect.com/topics/nursing-and-health-professions/logistic-regression-analysis), and evaluating the dataset’s attributes. A random forest classifier algorithm is developed to identify heart diseases with higher accuracy. We then discuss the random forest classifier algorithm, including the experiments and the results, which provide better accuracies for research diagnoses. We conclude the paper with objectives, limitations and research contributions.

1. **Introduction**

Heart diseases are often used in exchange for cardiovascular diseases. These kinds of diseases mainly refer to the conditions of blocked or narrowed blood vessels, resulting in a stroke, [chest pain](https://www.sciencedirect.com/topics/nursing-and-health-professions/thorax-pain) or [angina](https://www.sciencedirect.com/topics/nursing-and-health-professions/angina-pectoris), and heart attack. Other kinds of heart conditions, such as those affecting the rhythm, valve, or muscle of the heart, are other types of heart diseases. On the other hand, machine learning is crucial for determining whether anyone has suffered from heart disease. In either case, if these are predicted ahead of time, doctors would have a much easier time gaining crucial information for treating and diagnosing patients. Heart disease is mainly an incorrect symptom of [coronary artery disease](https://www.sciencedirect.com/topics/nursing-and-health-professions/coronary-artery-disease). It is also known as a cardiac disease; therefore, it is not with cardiovascular disease, which is any blood vessel disease.

Python is a programming language with a high level of object-oriented abstraction with a spirited, energetic collection of building options and quick development cycles.  Analysis, it is regarded as one of the safest programming languages with numerous applications in the medical field. Furthermore, it is regarded as a well-liked and well-accepted programming language with applications traversing over AI-based software developments and several other web applications. The python framework is used easily for creating a desktop or web-based application. With the application of python programming in the health care sectors, especially for detecting heart diseases, clinicians and institutions can provide better and improvised outcomes for the patients through scalable and dynamic applications. However, the coding packages and libraries used in this project are Pandas, Matplotlib, IPython, Numpy, Python, SciPy, and many others.

1. **Literature review**

The project comprises of detecting the presence of heart diseases using Python. The dataset comprised several factors, such as Chol, treetops, sex, age, and others. Several other import libraries, such as matplotlib, Numpy, Pandas, warnings, and many others, were used for the project. [Correlation matrix](https://www.sciencedirect.com/topics/mathematics/correlation-matrix), histogram, [support vector classifier](https://www.sciencedirect.com/topics/nursing-and-health-professions/support-vector-machine), K Neighbors Classifier, Random Forest Classifier, and Decision Tree Classifier were used for assessing the outcomes of the specified dataset using a python programming language. Additionally, Python is also considered an open-source language that encourages developing innovative solutions for the health care sectors and supplies better outcomes for the patients, resulting in enhanced care delivery. However, the language also complies with the HIPAA checklist for assuring the safety of medical information. The major causes of heart disease are diabetes, obesity, unhealthy diet, overweight, excessive alcohol use, and physical inactivity. Therefore, heart disease includes [arrhythmia](https://www.sciencedirect.com/topics/nursing-and-health-professions/heart-arrhythmia) that is considered as [atherosclerosis](https://www.sciencedirect.com/topics/nursing-and-health-professions/atherosclerosis) is the hardening of the arteries caused by a heart rhythm abnormality. During a heart attack, some people experience these symptoms. Additionally, pain that spreads to the arm, dizziness or [light headedness](https://www.sciencedirect.com/topics/nursing-and-health-professions/dizziness), throat, snoring, and sweating can occur. Heart attacks, strokes, and [coronary heart disease](https://www.sciencedirect.com/topics/nursing-and-health-professions/ischemic-heart-disease), also known as heart failure and [coronary artery disease](https://www.sciencedirect.com/topics/nursing-and-health-professions/coronary-artery-disease), are much more common in people over 65 than in younger people.

1. **Problem Statement**

Currently, the health care sector is generating information from several facilities and patients. By applying the best usage of this data, doctors can easily anticipate superior methods for treatment and enhance the complete delivery system of the health care sectors . One of the most important uses is that the python framework can help make sense and encourage computational facilities in extracting valuable insights from the information over the health care sectors. Moreover, Python is considered to be one of the most renowned programming languages all around the globe. 32% of the UK individuals considered this programming language a secured language for developing healthcare applications . Patients can experience symptoms, such as chest pain, shortness of breath, and fatigue when plaque grows large enough to obstruct blood flow.

Additionally, the health care projects made using the Python language must deal with HIPPA (Health Insurance Portability and Accountability Act) requirements for dealing with healthcare records. Python supports computer security, as it has built-in tools that provide software-defined security. However, Python is currently used in the health care field for data science and machine learning applications that improve patient outcomes. The algorithms of machine learning encourage healthcare analytics to use Python, as developers can easily establish tracking and [health monitoring](https://www.sciencedirect.com/topics/mathematics/health-monitoring) applications. Thus, in this case, also, python programming is used for detecting heart disease.

1. **Objectives**

The research aims are to detect heart disease using the python programming language.

The objectives of the study are as follows:

To critically analyze the ways python language is used to detect heart disease.

To critically investigate the previous activities and apply a suitable methodological approach for superscribing the identified problem.

To critically apply data interpretation strategies in python language for health problem detection.

**5.References**

[[2]](https://www.sciencedirect.com/science/article/pii/S2772442522000016#bb2)Mathur P.**Overview of machine learning in healthcare** Machine Learning Applications using Python, A Press, Berkeley, CA (2019), pp. 1-1 [CrossRef](https://doi.org/10.1007/978-1-4842-3787-8_1)[Google Scholar](https://scholar.google.com/scholar_lookup?title=Overview%20of%20machine%20learning%20in%20healthcare&publication_year=2019&author=P.%20Mathur) [[3]](https://www.sciencedirect.com/science/article/pii/S2772442522000016#bb3) Guleria P., Sood M.**Intelligent learning analytics in healthcare sector using machine learning**Machine Learning with Health Care Perspective, Springer,Cham (2020), pp. 39-55 [CrossRef](https://doi.org/10.1007/978-3-030-40850-3_3)[Google Scholar](https://scholar.google.com/scholar_lookup?title=Intelligent%20learning%20analytics%20in%20healthcare%20sector%20using%20machine%20learning&publication_year=2020&author=P.%20Guleria&author=M.%20Sood) [[4]](https://www.sciencedirect.com/science/article/pii/S2772442522000016#bb4) Kumar V.V. **Healthcare Analytics Made Simple: Techniques in Healthcare Computing using Machine Learning and Python** Packt Publishing Ltd. (2018) Accessed on 5th March, 2021 [Google Scholar](https://scholar.google.com/scholar_lookup?title=Healthcare%20Analytics%20Made%20Simple%3A%20Techniques%20in%20Healthcare%20Computing%20using%20Machine%20Learning%20and%20Python&publication_year=2018&author=V.V.%20Kumar) [[5]](https://www.sciencedirect.com/science/article/pii/S2772442522000016#bb5) BelItSoft **Python in healthcare** BelItSoft (2017) Available at: <https://belitsoft.com/custom-application-development-services/healthcare-software-development/python-healthcare>, [Accessed on 5th March, 2021] [Google Scholar](https://scholar.google.com/scholar_lookup?title=Python%20in%20healthcare&publication_year=2017&author=BelItSoft) [[6]](https://www.sciencedirect.com/science/article/pii/S2772442522000016#bb6) Nithya B., Ilango V. **Predictive analytics in health care using machine learning tools and technique** 2017 International Conference on Intelligent Computing and Control Systems (ICICCS), IEEE (2017), pp. 492-49 [ViewRecordinScopus](https://www.scopus.com/inward/record.url?eid=2-s2.0-85047615879&partnerID=10&rel=R3.0.0)[GoogleScholar](https://scholar.google.com/scholar_lookup?title=Predictive%20analytics%20in%20health%20care%20using%20machine%20learning%20tools%20and%20techniques&publication_year=2017&author=B.%20Nithya&author=V.%20Ilango)[[7]](https://www.sciencedirect.com/science/article/pii/S2772442522000016#bb7)McPadden J., Durant T.J., Bunch D.R., Coppi A., Price N., Rodgerson K., Torre Jr. C.J., Byron W., Hsiao A.L., Krumholz H.M., Schulz W.L.**Health care and precision medicine research: analysis of a scalable data science platform** J. Med. Internet Res., 21 (4) (2019), pp. e13043-e13045[CrossRef](https://doi.org/10.2196/13043)[View Record in Scopus](https://www.scopus.com/inward/record.url?eid=2-s2.0-85064537463&partnerID=10&rel=R3.0.0)[Google Scholar](https://scholar.google.com/scholar_lookup?title=Health%20care%20and%20precision%20medicine%20research%3A%20analysis%20of%20a%20scalable%20data%20science%20platform&publication_year=2019&author=J.%20McPadden&author=T.J.%20Durant&author=D.R.%20Bunch&author=A.%20Coppi&author=N.%20Price&author=K.%20Rodgerson&author=C.J.%20Torre%C2%A0Jr.&author=W.%20Byron&author=A.L.%20Hsiao&author=H.M.%20Krumholz&author=W.L.%20Schulz) [[8]](https://www.sciencedirect.com/science/article/pii/S2772442522000016#bb8) Panesar A. **Machine Learning and AI for Healthcare (1–73)**

Apress, Coventry, UK (2019)Available at <https://iedu.us/wp-content/uploads/edd/2020/01/Arjun_Panesar_Machine_Learning_and_AI_for_Health-iedu.us_.pdf>, [Accessed on 5th March, 2021][Google Scholar](https://scholar.google.com/scholar?q=Machine%20Learning%20and%20AI%20for%20Healthcare) [[9]](https://www.sciencedirect.com/science/article/pii/S2772442522000016#bb9)Davenport T., Kalakota R.

**The potential for artificial intelligence in healthcare** Future Healthcare J., 6 (2) (2019), p. 94 [CrossRef](https://doi.org/10.7861/futurehosp.6-2-94)[View Record in Scopus](https://www.scopus.com/inward/record.url?eid=2-s2.0-85073819217&partnerID=10&rel=R3.0.0)[Google Scholar](https://scholar.google.com/scholar_lookup?title=The%20potential%20for%20artificial%20intelligence%20in%20healthcare&publication_year=2019&author=T.%20Davenport&author=R.%20Kalakota) [[10]](https://www.sciencedirect.com/science/article/pii/S2772442522000016#bb10) Vidhya Analytics

**Commonly used machine learning algorithms (with python and R codes)** (2021)

Analytics Vidhya. Available at: <https://www.analyticsvidhya.com/blog/2017/09/common-machine-learning-algorithms/>, [Accessed on 5th March, 2021][Google Scholar](https://scholar.google.com/scholar?q=Commonly%20used%20machine%20learning%20algorithms) [[11]](https://www.sciencedirect.com/science/article/pii/S2772442522000016#bb11)

Mayfield H., Smith C., Gallagher M., Hockings M. **Use of freely available datasets and machine learning methods in predicting deforestation** Environ. Model. Softw., 87 (2017), pp. 17-28 [Article](https://www.sciencedirect.com/science/article/pii/S1364815216308428)[Download PDF](https://www.sciencedirect.com/science/article/pii/S1364815216308428/pdfft?md5=794302daa3d420b3ec93c0d5e47c4379&pid=1-s2.0-S1364815216308428-main.pdf)[View Record in Scopus](https://www.scopus.com/inward/record.url?eid=2-s2.0-84994000503&partnerID=10&rel=R3.0.0)[Google Scholar](https://scholar.google.com/scholar_lookup?title=Use%20of%20freely%20available%20datasets%20and%20machine%20learning%20methods%20in%20predicting%20deforestation&publication_year=2017&author=H.%20Mayfield&author=C.%20Smith&author=M.%20Gallagher&author=M.%20Hockings) [[12]](https://www.sciencedirect.com/science/article/pii/S2772442522000016#bb12)Barot P. **Why use python in healthcare applications?** BoTree Technologies. (2020)Available at: <https://www.botreetechnologies.com/blog/python-in-healthcare-application/>, [Accessed on 5th March, 2021] [Google Scholar](https://scholar.google.com/scholar?q=Why%20use%20python%20in%20healthcare%20applications) [[13]](https://www.sciencedirect.com/science/article/pii/S2772442522000016#bb13) Bhanot K. **Predicting presence of heart diseases using machine learning** Towards Data Sci. (2019) Available at: <https://towardsdatascience.com/predicting-presence-of-heart-diseases-using-machine-learning-36f00f3edb2c>, [Accessed on 5th March, 2021] [Google Scholar](https://scholar.google.com/scholar_lookup?title=Predicting%20presence%20of%20heart%20diseases%20using%20machine%20learning&publication_year=2019&author=K.%20Bhanot) [[14]](https://www.sciencedirect.com/science/article/pii/S2772442522000016#bb14) Zaidi J. **Project: Predicting heart disease with classification machine learning algorithms** Towards Data Sci. (2020, [Accessed on 5th March, 2021][Google Scholar](https://scholar.google.com/scholar_lookup?title=Project%3A%20Predicting%20heart%20disease%20with%20classification%20machine%20learning%20algorithms&publication_year=2020&author=J.%20Zaidi)[[15]](https://www.sciencedirect.com/science/article/pii/S2772442522000016#bb15) Ozgur C., Colliau T., Rogers G., Hughes Z. **Matlab vs. Python vs. R** J. Data Sci., 15 (3) (2017), pp. 355-371 [View Record in Scopus](https://www.scopus.com/inward/record.url?eid=2-s2.0-85065682302&partnerID=10&rel=R3.0.0)[Google Scholar](https://scholar.google.com/scholar_lookup?title=Matlab%20vs.%20Python%20vs.%20R&publication_year=2017&author=C.%20Ozgur&author=T.%20Colliau&author=G.%20Rogers&author=Z.%20Hughes) Srinath K.R. **Python–the fastest growing programming language** Int. Res. J. Eng. Technol., 4 (12) (2017), pp. 354-357 [View Record in Scopus](https://www.scopus.com/inward/record.url?eid=2-s2.0-85074435485&partnerID=10&rel=R3.0.0)[Google Scholar](https://scholar.google.com/scholar?q=Pythonthe%20fastest%20growing%20programming%20language) Copeland B. **Advancing opportunities in health care with python-based machine learning** Chief HealthCare Executive (2019), [Accessed on 5th March, 2021] [Google Scholar](https://scholar.google.com/scholar_lookup?title=Advancing%20opportunities%20in%20health%20care%20with%20python-based%20machine%20learning&publication_year=2019&author=B.%20Copeland)

[[16]](https://www.sciencedirect.com/science/article/pii/S2772442522000016#bb16)

[[17]](https://www.sciencedirect.com/science/article/pii/S2772442522000016#bb17)

Copeland B.

**Advancing opportunities in health care with python-based machine learning**

Chief HealthCare Executive (2019)

Available at: <https://www.chiefhealthcareexecutive.com/view/advancing-opportunities-in-healthcare-with-pythonbased-machine-learning>, [Accessed on 5th March, 2021]

[Google Scholar](https://scholar.google.com/scholar_lookup?title=Advancing%20opportunities%20in%20health%20care%20with%20python-based%20machine%20learning&publication_year=2019&author=B.%20Copeland)