Machine Learning for Heart Disease Prediction

Albayrak, Melike Gülşen*

1 Important Information About Midterm

WRITE YOUR GITHUB REPO LINK ON LINE 35 IN THIS FILE!

Project Proposal submisson will be done by uploading a zip file to the ekampus system along with the Github repo link. If you do not upload a zip file to the system and do not provide a Github repo link, you will be deemed not to have entered the midterm and final exams.

You must upload your project folder (`YourStudentID.zip` file) to *ekampus.ankara.edu.tr* until 16 April 2023, 23:59.

Read the README.md file in the project folder for more information.

2 Introduction

Heart disease is a leading cause of morbidity and mortality worldwide. Early detection and intervention can improve patient outcomes, but traditional risk prediction models have limited accuracy. The purpose of this project is to investigate the use of machine learning techniques for predicting the presence of heart disease in patients based on their demographic and clinical characteristics, using the Heart disease dataset from the UCI Machine Learning repository which is a website that i found from my searchings about the most helpful dataset providers to find suitable datasets for researchs. After my investigations for finding a good dataset, I chose the Heart Disease Dataset that has 303 observations with 14 variables including categorical (sex, chest pain type (cp), fasting blood sugar $>$ 120 mg/dl (fbs), resting electrocardiographic results (restecg), exercise induced angina (exang), slope of the peak exercise ST segment (slope), number of major vessels colored by fluoroscopy (ca), thalassemia (thal), target), discrete numeric (maximum heart rate achieved (thalach)) and continous numeric (age, resting blood pressure (trestbps, in mm Hg), serum cholesterol (chol, in mg/dl), ST depression induced by exercise relative to rest (oldpeak)) variables.

After chosing the dataset, i pursued on looking for articles related to heart diseases and i saw that prediction of heart disease's existence or non-existence is a vital situation and

*21080562, Github Repo

machine learning techniques are now has a great impact for these prediction studies. And also in my further research, i noticed that some articles show that during this prediction of heart diseases ,with the help of machine learning, demographical, medical or clinical background knowledges about patients are also important for the process. So at last, i designed my question as "Can machine learning techniques accurately predict the presence of heart disease in patients based on their demographic and clinical characteristics?". Then, i just focused more specifically on the articles related to my research question for the project.

2.1 Literature Review

Heart disease is a major public health problem that can lead to serious health consequences if not diagnosed and treated early. Machine learning techniques have shown promise in accurately predicting the presence of heart disease based on demographic and clinical characteristics. This literature review summarizes the findings of four studies that have examined the application of machine learning algorithms for predicting heart disease. All of these studies discusses the importance of demographic and clinical characteristics such as age, sex, blood pressure, and cholesterol levels in predicting heart disease. The authors evaluate various studies that have used machine learning algorithms including decision trees, artificial neural networks, support vector machines, and random forests etc. for heart disease prediction and compare their performance (Kutiame et al., 2022), (Aljanabi et al., 2018), (Singh & Kumar, 2020). And one of the studies especially highlights the importance of selecting appropriate features and preprocessing techniques to improve the accuracy of the model (Reddy et al., 2019).

Overall, these studies suggest that ML techniques can be used for heart disease prediction with demographical and clinical informations and have the potential to improve patient outcomes. However, more research is needed to identify the most effective ML algorithms for heart disease prediction and to address the challenges associated with using ML in clinical settings.

References section is created automatically by Rmarkdown. There is no need to change the references section in the draft file.

You shouldn't delete the last 3 lines. Those lines are required for References section.

3 References

Aljanabi, M., Qutqut, M., & Hijjawi, M. (2018). Machine learning classification techniques for heart disease prediction: A review. *International Journal of Engineering and Technology*, *7*, 5373–5379. https://doi.org/10.14419/ijet.v7i4.28646

Kutiame, S., Millham, R., Adekoya, A. F., Tettey, M., Weyori, B. A., & Appiahene, P. (2022). Application of machine learning algorithms in coronary heart disease: A systematic literature review and meta-analysis. *International Journal of Advanced Computer Science and Applications*, *13*(6).

Reddy, N. S. C., Nee, S. S., Min, L. Z., & Ying, C. X. (2019). Classification and feature selection approaches by machine learning techniques: Heart disease prediction. *International Journal of Innovative Computing*, *9*(1).

Singh, A., & Kumar, R. (2020). Heart disease prediction using machine learning algorithms. *2020 International Conference on Electrical and Electronics Engineering (ICE3)*, 452–457.