Examining The Interplay of Lifestyle Diseases and Mental Health: Devising a Possible Machine Learning Based System for Diagnosis

Hrishav Banerjee[1], Divya Singh[2]

[1]Programmer Analyst Trainee, Cognizant Technology Solutions India Pvt. Ltd., Gurgaon.
[2]Software Engineer, Bosch Global Software Technologies.

Abstract

In this fast-paced world, we frequently neglect our health, adopting unhealthy poor habits that have a long-term negative impact on our health yet typically go unnoticed. These conditions are known as lifestyle diseases because they result from poor lifestyle choices. This results in chronic non-communicable diseases (NCDs), which account for over 70% of global fatalities each year. Early identification of these conditions is essential to preventing such grave consequences, as is changing your lifestyle choices, and eating nutritious food with a balanced diet rather than bingeing on junk food.

We rarely pay attention to our mental health in our daily lives, but in this pressured lifestyle, our mental health suffers the most. It is a well-known reality that people are more concerned with staying physically well than with improving their mental health. Mental disabilities are very common among this generation and distresses 8% of the world's population. People with mental disorders often fall prey to social injustice and prejudice. Poverty, inequality, body shaming and social prejudices are societal factors that increase the risk of mental illness and affect the course of life for such affected people. The relationship between lifestyle diseases and mental health is very intricate and diverse. Lifestyle diseases such as obesity and diabetes often arise from sedentary lifestyle and stress. This establishes the relationship between lifestyle diseases and diverse mental health conditions. The COVID-19 pandemic has had a significant influence on mental health around the world. With the social isolation, economic uncertainty, and dread, there has been a huge increase in mental health disorders. As people deal with the pandemic's hardships, anxiety, sadness, and stress have become more widespread. It is apparent that tackling the pandemic's mental health consequences is a vital component of worldwide recovery efforts.

Keywords: lifestyle diseases, NCD, Mental Health, Diabetes, CVD, asthma, osteoporosis, machine learning, diagnosis, detection.

1 Introduction

1.1 General introduction

Lifestyle diseases, which can also be referred as non-communicable diseases (NCDs), are severe health problems caused by poor choice of lifestyle and other unhealthy habits. Obesity, diabetes, cardiovascular disease, hypertension, and some types of cancer are all examples of lifestyle diseases. Lifestyle diseases and mental health problems usually go together, for example, a person suffering from obesity can also show symptoms of depression due to societal factors as well as hormonal changes in body caused due to obesity. leptin and insulin are the two hormones which influence our eating habits and appetite and this hormone is shown to be on a higher level and cause a condition called "stress eating" which ultimately causes obesity. Individuals with poor mental health may resort to harmful lifestyle choices such as stress eating, sedentary activity, or substance misuse as coping mechanisms. In contrast, unhealthy habits can contribute to mental health problems by increasing stress, worry, and sadness because of the physical health concerns they produce. Students who were physically inactive (those who did not engage in any vigorous or moderate-intensity activity) grew from 6 to 28 percent in prevalence during the pandemic, whereas those who engaged in adequate activity (at least 60 minutes per day) significantly declined from 60 to 17 percent [8].

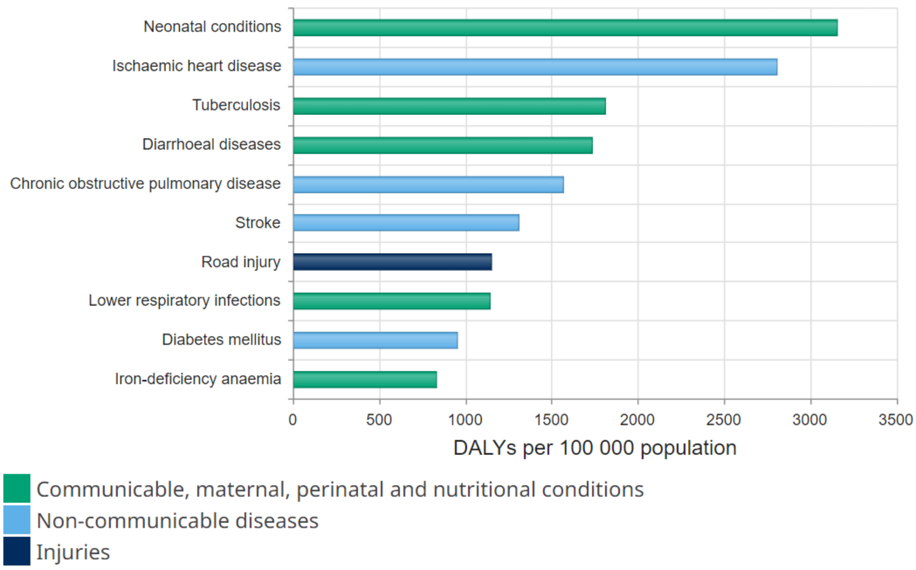

Fig. 1. Contribution by different non-communicable diseases to disability-adjusted life-years worldwide in 2005(Data adapted from WHO) [1]

NCD(s) takes the life of approximately 42 million people per year, or roughly double the population of New York City, and account for 74% of all global mortality. Every year, about 15 million individuals (almost twice the population of New Jersey) die from NCDs among the age groups of 35 and 75. The majority of these "premature" fatalities occur in low- and middle-income nations. Cancer (9.3 million), respiratory diseases (4.1 million), and diabetes (1.5 million) are the leading causes of NCD-related deaths, accounting for 17.9 million annual fatalities [2]. These lifestyle diseases are responsible for more than 82% of all premature deaths caused by NCD(s).

1.2 Main Contributions of this research paper

The fundamental goal of this study is to investigate and clarify the complex interaction between lifestyle disorders and mental health problems. Our research intends to accomplish the following major goals:

- **Investigate Prevalence and Co-occurrence:** To examine the prevalence rates of lifestyle diseases and mental health problems in each population, emphasizing the extent of their co-occurrence as well as any demographic or socioeconomic tendencies.
- **Evaluate health outcomes:** Assess the influence of the co-occurrence of lifestyle diseases and mental health disorders on overall health outcomes, such as quality of life, treatment responsiveness, and healthcare utilization.

- **Identify Underlying Mechanisms:** Investigate the biological, psychological, and social mechanisms that mediate the link between lifestyle diseases and mental health concerns, shedding light on the pathways by which they interact.

1.3 Major lifestyle diseases

Diabetes. Type 2 diabetes is a prevalent and deadly long-term disease in which body cells are not able to utilize glucose to produce energy. Diabetes affects one out of every twelve people in India, which is the highest among all other countries. The Western Pacific area has the most diabetes patients, with the greatest prevalence of the disease (37.5%). Meanwhile, the Middle East and North Africa area has the highest adult diabetes prevalence (10.9%) [3]. Millions of individuals throughout the world have been afflicted with the lethal chronic disease known as diabetes. Of all diabetes forms, Type 2 Diabetes Mellitus (T2DM) affects 90% of the population. Due to a lack of knowledge and an underfunded healthcare system, millions of T2DM patients go misdiagnosed [10].

Region	Type 1 diabetes in children (0-14 yr)		Diabetes in adults (20-79 yr)				Hyperglycemia in pregnancy (20-49 yr)	
	2013		2013		2035		2013	
	Number in thousands	Newly diagnosed in thousands	Number in millions	Comparative prevalence	Number in millions	Comparative prevalence	Cases in live births in millions	Comparative prevalence
Africa	39.1	6.4	19.8	5.7%	41.5	6.0%	4.6	14.4%
Europe	129.4	20.0	56.3	6.8%	68.9	7.1%	1.7	12.6%
Middle East and North Africa	64.0	10.7	34.6	10.9%	67.9	11.3%	3.4	17.5%
North America and Caribbean	108.6	16.7	36.8	9.6%	50.4	9.9%	0.9	10.4%
South and Central America	45.6	7.3	24.1	8.2%	38.5	8.2%	0.9	11.4%
South East Asia	77.9	12.5	72.1	8.7%	123.0	9.4%	6.3	25.0%
Western Pacific	32.5	5.3	138.2	8.1%	201.8	8.4%	3.7	11.9%
World	497.1	78.9	381.8	8.3%	592.0	8.8%	21.4	14.8%

Fig. 2. Number of infants (0-14 years) with type 1 diabetes, adults (20-79 years), and pregnant women (20-49 years) with hyperglycemia (type 2 or gestational diabetes). (Data from the International Diabetes Federation Diabetes Atlas, 6th edition, 2013) [4]

Asthma. Asthma is a chronic lifestyle illness in which airways narrow and create breathing issues that must be treated with an inhaler. Every third child in Delhi has a lung problem. In 2016, India reported 35 million cases of asthma. Air pollution is responsible for 30% of all premature deaths in the country. Asthma was diagnosed in 8.4% of Canadians under the age of 12 in the 2003 Canadian Community Health Survey, with teenagers (> 12%) having the highest prevalence [5]. Two of the most prevalent chronic respiratory health issues are obstructive airway illnesses, such as asthma and Chronic Obstructive Pulmonary Disease (COPD). Making a diagnosis for either of these disorders requires medical competence. As a result, this method takes a lot of time from healthcare professionals, and there is operator and intra-operator variability in the diagnostic quality [11].

Cardiovascular Diseases. Heart diseases are perhaps the most common among lifestyle diseases and cause the greatest number of mortalities every year. Every year, more than 2.7 million people (about the population of Mississippi) die in India from heart disease, with 52% of them being under the age of 70. Heart disease was the most common cause of death in 2015, followed by malignancy (595,930), with cardio-vascular diseases (CVD) remaining one of the two leading causes of death in the United States since 1975, accounting for 633,842 deaths, or one in every four fatalities [6]. There are multiple conditions under CVD(s) such as Arrhythmia, Cardiomyopathy and Coronary Artery diseases. Heart disorders, often known as cardiovascular diseases, are the leading causes of death in the world today. Heart disease impairs blood vessel function and can result in coronary artery infections, which weaken the patient's body. As a result, there is a need for a dependable, accurate, and feasible system that can diagnose heart illness in a timely manner so that the cardiac patient can receive effective therapy before it progresses to a major problem, ultimately culminating in a heart attack [12].

Osteoporosis. It is a bone-weakening disorder that causes severe problems in motor control mainly in the knees and ankle. androgen hormones in men and estrogen hormones in women. There are approx. 10 million cases are recorded each which proves the severity of such a condition, though it usually occurs among an older bracket of ages but can also affect the younger population. According to the International Osteoporosis Foundation, one in every three women aged 50 and older, and one in every five males, will suffer from osteoporotic fractures over their lifetime [7]. In this disorder, the bone density decreases which results in lower bone strength and deformed bone structure.

2 Scope and Limitations

It is critical to recognize some constraints that may limit the scope and depth of our investigation:

- **Diversity of Lifestyle Diseases:** The phrase "lifestyle diseases" refers to a wide range of ailments, each with its own set of features. While we make every effort to address various frequent lifestyle disorders, the breadth of this document may not cover all potential diseases in this area.
- **Data Availability:** The quality and availability of data on lifestyle diseases and mental health disorders may vary by place and time, thereby limiting the scope of our analysis.
- **Generalizability:** Because the association between lifestyle disorders and mental health varies across people and geographic places, the findings of this study may not be universally relevant. As a result, when generalizing the findings to different contexts, care should be taken.

3 Proposed Methodology

The methodology used to design this system is purely based on machine learning models (ML). The development of ML models requires a huge volume of data which is to be fed into the system during the training phase. This data was collected through google forms, but such amount of data was insufficient to train the model, hence, we collected data form Kaggle to train the model accurately. The idea is to train the model with data and when any user enters his/her symptoms, the model should be capable enough to map the input to a disease which is most likely to occur, hence, supporting the idea of early detection of diseases. The flow of the system in described in bullet points below:

- **Data Collection:** Data is collected using google forms, and due to volume restrictions, we collected data from Kaggle.
- **Model Selection:** We wanted a model which can classify data between 2 distinct sets or between true and false, so we selected logistic regression for diabetes, for brain tumor analysis, we selected CNN model which can classify and analyze data based on image classification.
- **Model Training:** The model is trained using data collected from google forms and Kaggle, this training is used for supervised learning of machine learning model, in this case, the model recognizes data points and will be able to analyze future inputs based on training.

- **Website Development:** A website has been developed using HTML, CSS and JS and it has ML as its backbone. An API (Application Programming Interface) will be developed which will act as a medium of communication between the website and the ML model, fetching data from the website and transferring it to the ML model and vice-versa.
- **Logistic Regression Model:** For Diabetes, we have trained a model based on logistic regression algorithm which is used for classification, it gives output which is of Boolean type i.e., true/false, 0 or 1 etc. This helps us predict if the user has a particular disease or not, based on the inputs received.
- **CNN Model:** CNN model is a very efficient model when it comes to analyzing data from images. So, we used this model to analyze brain tumor images which we collected from the internet. This brain tumor model is not deployed on the website yet as we are still working on it. This model will help analyze images entered by user and predict if the results are positive or negative.
- Once this system is developed and deployed properly, it will help a lot of patients to get instant insights and reports from their symptoms and will give wider access to healthcare to everyone, which is our goal.

4 Role of Machine Learning

4.1 General Introduction

Machine learning (ML) models are critical in furthering our understanding of the complex link between lifestyle diseases and mental health issues. These models may evaluate large and complex datasets, including genetic, clinical, behavioral, and environmental variables, to find hidden patterns and connections that would be difficult to detect using traditional statistical methods. Predictive indicators, risk factors, and early warning signals for both lifestyle diseases and mental health concerns can be identified using ML algorithms, allowing for early intervention and individualized treatment options. Furthermore, by investigating intricate interactions and feedback loops, ML models can help understand the fundamental mechanisms connecting these two domains.

Training a machine learning model is carried out using preprocessed data. To accomplish this, it could be necessary to use supervised learning, which teaches the model to predict the existence of a particular disease based on the data acquired. use fictitious data to assess the trained model's predictions' accuracy. This may call for evaluating the model's performance using metrics like precision, recall, and accuracy after running it on a different dataset. Mental illness can affect the cognition, emotion, and behavior among the people. For children, their ability to learn could be interfered by mental disorders. Besides that, mental illness can cause inconvenience to the adults, especially in their families, workplaces, and in the society. There are many types of mental disorders commonly known as schizophrenia, depression, bipolar disorder, and anxiety [9].

4.2 ML Models Used to Detect Lifestyle Diseases

Logistic Regression. Logistic regression is a popular approach in the field of Supervised Learning. Its main goal is to predict categorical dependent variables using a set of independent variables. When we need to forecast outcomes that fall into discrete or categorical categories, we use logistic regression. It provides probabilistic values within the range of 0 and 1. These results may be either True or False, 0 or 1, or Yes or No. While logistic regression and linear regression have certain similarities, their applications differ. Logistic regression is designed to handle classification jobs, whereas linear regression is designed to handle regression difficulties. The equation of the straight line:

$$x = b_0 + b_1y_1 + b_2y_2 + \ldots + b_ny_n \quad (1)$$

In Logistic Regression, x can only be between 0 and 1, therefore divide the previous equation by (1-x):

$$x/1-x; 0 \text{ for } x=0 \text{ and infinity for } x=1 \quad (2)$$

$$\text{Log}[x/1-x] = b_0 + b_1y_1 + b_2y_2 + \ldots + b_ny_n \quad (3)$$

Convolutional Neural Network (CNN). Convolutional Neural Networks) are a part of deep neural networks that are mostly used for visual data analysis. While neural networks are frequently associated with matrix multiplications, it is vital to emphasize that ConvNets deviate from this typical method and instead employ a unique technique known as convolution. Convolution explains how the form of one function changes when it is joined with another, resulting in the development of a third function. A neural network consists of three layers: an input layer, hidden layers (where most of the computation occurs), and an output layer. Convolutional Neural Networks (CNNs) are inspired by the structure of the human brain. Artificial neurons, or nodes within CNNs, take inputs, execute computations on them, and produce an output, like how a brain neuron disseminates information throughout the body. photos are used as input in CNNs, with the input layer receiving arrays of pixel values from these photos. Convolutional Neural Networks (CNNs) are frequently made up of numerous hidden layers that extract information from images using mathematical processes.

Machine Learning Model Outcome for Diabetes. We created a Logistic Regression model to predict the occurrence of diabetes through an ML model trained on patient symptoms collected via Google Forms and using dummy data from Kaggle. There are two graphs based on our model, with respect to age and BMI. We have observed these patterns through training data fetched from Kaggle.

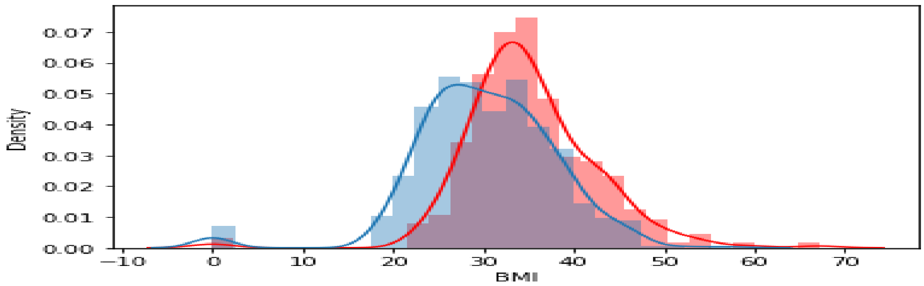

Fig. 3. BMI vs. Diabetes Density graph

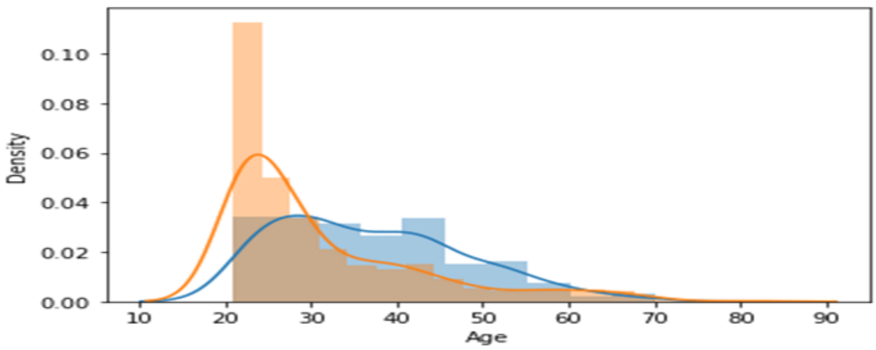

Fig. 4. Age vs. Diabetes Density graph

4.3 Implementation Details of ML based Healthcare System. A machine learning model is being used on the backend to generate similar results on the website. The ML model and the website communicate via an API (Application Programming Inter-face). The user fills out the questionnaire and clicks the submit button; the results are then loaded into the ML model in the backend. The ML model (logistic regression) is trained using a tagged dataset taken from Google Forms as well as datasets from Kaggle because we do not yet have access to such large datasets. When a user inputs his or her own symptoms into the questionnaire, the information is loaded into the ML model, which compares the type of symptoms to the training data and classifies the symptoms into several illness categories (currently, testing is limited to diabetes data). The API acts as a link between the website and the backend ML model, retrieving input data from the website and transferring it to the ML model, as well as retrieving results from the ML model and displaying them on the website.

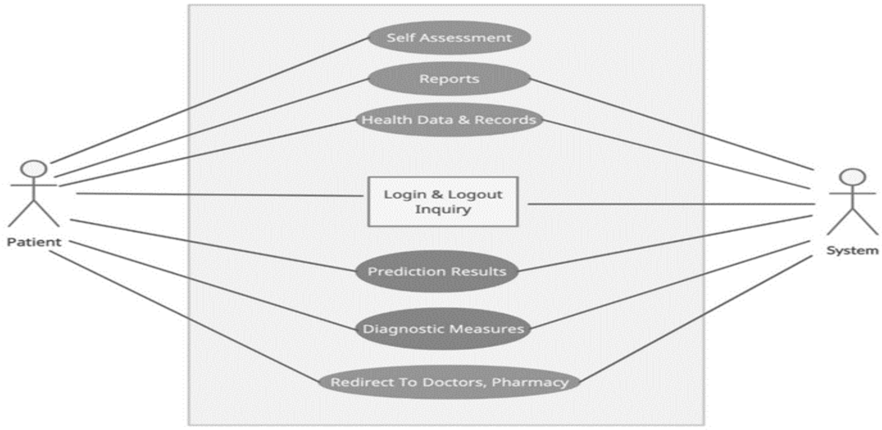

Fig. 5. Flow Chart of Working of Website

5 Results and Conclusion

In machine learning, performance metrics are measures used to evaluate the quality and effectiveness of a model. Common metrics include accuracy, precision, recall, F1 score, confusion matrix etc. These metrics help assess how well a model is making predictions and can guide the model's improvement during training and evaluation. The choice of metrics depends on the specific goals and characteristics of the problem being solved. Our models are also evaluated against metrics to ensure their accuracy in predicting data. The Performance metric used in this project are as follows:

- **Confusion Matrix:** In machine learning, a confusion matrix is a table that describes the performance of a classification model. It presents a summary of the predicted versus actual class labels, providing insights into the model's accuracy, precision, recall and other performance metrices. The matrix typically has four entries: true positive (correctly predicted positive instances), false positive (incorrectly predicted as positive), true negative (correctly predicted negative instances) and false negative (incorrectly predicted as negative).
- **Precision:** It is a performance metric that measures the accuracy of positive predictions. It is calculated as the number of true positives divided by the sum of true positives and false positives. A high precision indicates a low rate of false positives, emphasizing the reliability of positive predictions.
- **Recall:** Recall, also known as sensitivity or true positive rate, is a performance metric that measures the ability of a model to capture all relevant instances of a positive class. It is calculated as the number of true positives divided by the sum of true positives and false negatives.

Our model's accuracy so far (Still under development):
Brain tumor accuracy: 97%, diabetes accuracy: 86%, heart disease prediction: 87%.

6 Results and Conclusion

The results of this extensive study shed important light on the complex link between mental health and lifestyle disorders. It has been clear through the rigorous examination of a wide variety of studies and data sources that lifestyle diseases including obesity, diabetes, cardiovascular disorders, and substance addiction have a significant and reciprocal influence on mental health. The study has shown the complexity of this relationship by highlighting the function of both physiological and psychological factors. Diet, exercise, and sleep patterns were found to be important mediators connecting these two domains of lifestyle choices and habits. This study also highlights the value of early intervention and holistic healthcare strategies in addressing these connected issues.

References

1. https://pubmed.ncbi.nlm.nih.gov/17804063/, Martin Prince, Vikram Patel, Shekhar Saxena, Mario Maj, Joanna Maselko, Michael R Phillips, Atif Rahman, No health without mental health.
2. https://www.who.int/data/gho/data/themes/topics/topic-details/GHO/ncd-mortality,World Health Organization Global Health Observatory, Noncommunicable diseases: Mortality.
3. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4478580/.
4. Statistics Canada. Asthma. Health Reports. 2005;16(2). http://publications.gc.ca/Collection-R/Statcan/82-003-XIE/0020482-003-XIE.pdf.
5. Olvera Lopez E, Ballard BD, Jan A. Cardiovascular Disease. [Updated 2023 Aug 7]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2023 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK535419/.
6. Sözen T, Özışık L, Başaran NÇ. An overview and management of osteoporosis. Eur J Rheumatol. 2017 Mar;4(1):46-56. doi: 10.5152/eurjrheum.2016.048. Epub 2016 Dec 30. PMID: 28293453; PMCID: PMC5335887.
7. https://datacatalog.worldbank.org/search/dataset/0037712/World-Development-Indicators.
8. Xiang, M., Liu, Y., Yamamoto, S. et al. Association of Changes of lifestyle behaviors before and during the COVID-19 pandemic with mental health: a longitudinal study in children and adolescents. Int J Behav Nutr Phys Act 19, 92 (2022). https://doi.org/10.1186/s12966-022-01327-8.
9. Mental Health Prediction Using Machine Learning: Taxonomy, Applications, and Challenges, Jetli Chung1and Jason Teo, https://doi.org/10.1155/2022/9970363.
10. Ganie, S.M., Malik, M.B. & Arif, T. Performance analysis and prediction of type 2 diabetes mellitus based on lifestyle data using machine learning approaches. J Diabetes Metab Disord 21, 339–352 (2022). https://doi.org/10.1007/s40200-022-00981-w.
11. Shuting Xu, Ravinesh C Deo, Jeffrey Soar, Prabal Datta Barua, Oliver Faust, Nusrat Homaira, Adam Jaffe, Arm Luthful Kabir, U. Rajendra Acharya, Automated detection of airflow obstructive diseases: A systematic review of the last decade (2013-2022), https://doi.org/10.1016/j.cmpb.2023.107746.
12. Riyaz, L., Butt, M.A., Zaman, M., Ayob, O. (2022). Heart Disease Prediction Using Machine Learning Techniques: A Quantitative Review. In: Khanna, A., Gupta, D., Bhattacharyya, S., Hassanien, A.E., Anand, S., Jaiswal, A. (eds) International Conference on Innovative Computing and Communications. Advances in Intelligent Systems and Computing, vol 1394. Springer, Singapore. https://doi.org/10.1007/978-981-16-3071-2_8.