

Abstract

Heart is the main organ that pumps blood to the body and for proper functioning of the body. Heart disease is a fatal human disease increasing globally in both developed and undeveloped countries and consequently, causes death. Normally, in this disease, the heart fails to supply a sufficient amount of blood to other parts of the body in order to accomplish their normal functionalities. It associates many risk factors in heart disease and a need of the time to get accurate, reliable, and sensible approaches to make an early diagnosis to achieve prompt management of the disease. Data mining is a commonly used technique for processing enormous data in the healthcare domain. Researchers apply several data mining and machine learning techniques to analyze huge complex medical data, helping healthcare professionals to predict heart disease. The proposed method is to build a machine learning model capable of classifying whether the person has heart disease or not. Different algorithms are compared and the best model is used for predicting the outcome.

TABLE OF CONTENTS

CHAPTER No.	TITLE	PAGE No
	ABSTRACT	i
	LIST OF FIGURES	iii
1	INTRODUCTION	1
	1.1 Outline of the Project	1
	1.2 Existing system	2
	1.3 Proposed System	3
2	LITERATURE SURVEY	4
	2.1 General	4
	2.2 Review of Literature Survey	4
3	METHODOLOGY	8
	3.1 MACHINE LEARNING	8
	3.2 Feasibility study	10

	3.3 Construction of a Predictive Model	11
	3.4 Project Requirements	12
	3.5 System Architecture	14
	3.6 Module Documents	17
	3.7 Data Validation/ Cleaning/Preparing Process	18
	3.8 Exploration data analysis of visualization	19
	3.9 Prediction result by accuracy	20
	3.10 Used Python Packages	23
4	RESULTS AND DISCUSSION	25
5	CONCLUSION AND FUTURE WORK	28
	5.1 CONCLUSION	28
	5.2 Future work	28
	APPENDIX	29
	A. SCREENSHOTS of the application	29
	B. Reference	26

CHAPTER 1

INTRODUCTION

1.1 OUTLINE OF THE PROJECT

Data science is an interdisciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from structured and unstructured data and apply knowledge and actionable insights from data across a broad range of application domains.

The term "data science" has been traced back to 1974, when Peter Naur proposed it as an alternative name for computer science. In 1996, the International Federation of Classification Societies became the first conference to specifically feature data science as a topic. However, the definition was still in flux.

The term "data science" was first coined in 2008 by D.J. Patil, and Jeff Hammerbacher, the pioneer leads of data and analytics efforts at LinkedIn and Facebook. In less than a decade, it has become one of the hottest and most trending professions in the market.

Data science is the field of study that combines domain expertise, programming skills, and knowledge of mathematics and statistics to extract meaningful insights from data.

Data science can be defined as a blend of mathematics, business acumen, tools, algorithms and machine learning techniques, all of which help us in finding out the hidden insights or patterns from raw data which can be of major use in the formation of big business decisions.

Data Scientist:

Data scientists examine which questions need answering and where to find the related data. They have business acumen and analytical skills as well as the ability to

mine, clean, and present data. Businesses use data scientists to source, manage, and analyze large amounts of unstructured data.

Required Skills for a Data Scientist:

- **Programming:** Python, SQL, Scala, Java, R, MATLAB.
- **Machine Learning:** Natural Language Processing, Classification, Clustering.
- **Data Visualization:** Tableau, SAS, D3.js, Python, Java, R libraries.
- **Big data platforms:** MongoDB, Oracle, Microsoft Azure, Cloudera.

1.2 Existing System:

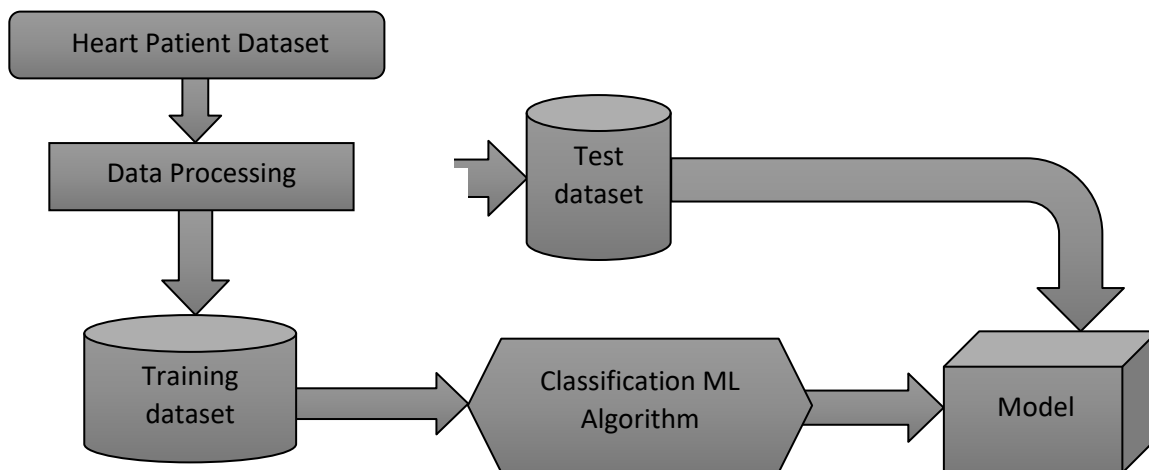
The Acoustic analysis has the potential to be a relatively low cost and non-invasive diagnostic tool for point-of-care assessment and remote monitoring of LVAD recipients. Prior work on acoustic analysis in LVAD recipients has focused on analysis of LVAD sounds to detect pump thrombosis. The analysis of heart sounds in LVAD recipients has been limited by the dominant LVAD sounds present within their precordial sounds. We have developed a novel signal processing pipeline to mitigate LVAD-generated sounds within precordial sounds recorded from LVAD recipients, potentially rendering automated heart sound analysis in this population feasible. Their analysis revealed that pump activity in LVAD recipients induces heart sounds with diverse acoustic signatures, such as S1 and S2 sounds with varying frequency ranges and relative amplitudes, and variabilities in S2 components based on changes in aortic valve mobility across different pump speeds.

Disadvantages:

1. Analyzing the sound graphs and getting the sound it's a complex process.
2. Better classification can be done by proper comparison of algorithms.

1.3 Proposed System:

The proposed method is to build a machine learning model for classification of heart disease. The process carries from data collection where the past data related to heart disease are collected. Data mining is a commonly used technique for processing enormous data in the healthcare domain. Heart disease, if found before proper treatment, can save lives. Machine learning is now applied and mostly used in health care where it reduces the manual effort and better model makes error less which leads to saving lives. The data analysis is done on the dataset proper variable identification done that is both the dependent variables and independent variables are found. Then proper machine learning algorithms are applied on the dataset where the pattern of data is learnt. After applying different algorithms a better algorithm is used for the prediction of outcome.



Architecture of Proposed model

Advantages:

1. The data used here will not be as complex as sound graphs and analysis of data will be easy.
2. Performance metrics of different algorithms are compared, and the better prediction is done.

CHAPTER 2

LITERATURE SURVEY

2.1 General

A literature review is a body of text that aims to review the critical points of current knowledge on and/or methodological approaches to a particular topic. It is a secondary source and discusses published information in a particular subject area and sometimes information in a particular subject area within a certain time period. Its ultimate goal is to bring the reader up to date with current literature on a topic and forms the basis for another goal, such as future research that may be needed in the area and precedes a research proposal and may be just a simple summary of sources. Usually, it has an organizational pattern and combines both summary and synthesis.

A summary is a recap of important information about the source, but a synthesis is a reorganization, reshuffling of information. It might give a new interpretation of old material or combine new with old interpretations or it might trace the intellectual progression of the field, including major debates. Depending on the situation, the literature review may evaluate the sources and advise the reader on the most pertinent or relevant of them.

2.2 Review of Literature Survey

Title: REVIEW OF HEART DISEASE PREDICTION SYSTEM USING DATA MINING AND HYBRID INTELLIGENT TECHNIQUES

Author: R. Chitra¹ and V. Seenivasagam²

Year: 2013

The Healthcare industry generally clinical diagnosis is done mostly by doctor's expertise and experience. The Computer Aided Decision Support System plays a major role in the medical field. With the growing research on heart disease predicting systems,

it has become important to categorize the research outcomes and provide readers with an overview of the existing heart disease prediction techniques in each category. Neural Networks are one of many data mining analytical tools that can be utilized to make predictions for medical data. From the study it is observed that Hybrid Intelligent Algorithm improves the accuracy of the heart disease prediction system. The commonly used techniques for Heart Disease Prediction and their complexities are summarized in this paper.

Title:EFFECTIVE PREDICTION OF CARDIOVASCULAR DISEASE USING CLUSTEROF MACHINE LEARNING ALGORITHMS

Author: G. Jignesh Chowdary¹, Suganya. G ² , Premalatha. M³

Year:2020

Cardiovascular diseases are one of the diseases that account for the loss of millions of lives each year. Lack of early prediction is the primary reason for the loss of lives, and this encourages researchers to develop intelligent systems for better prediction. The heart is a muscular organ that is largely made of cardiac muscle. The heart pumps oxygenated blood throughout the body by rhythmic contraction and dilation. Cardiovascular diseases deal with the problems related to heart and blood vessels. Cardiovascular diseases are the life-threatening diseases of the present-day world. According to the World Health Organisation(WHO), 17 million people die each year due to Cardiovascular disease, which accounts for 31% of the deaths worldwide[1]., and it is among the top 10 global causes of deaths around the world [2]. It is challenging to identify cardiovascular disease because it depends on many factors. Here are some significant factors that cause cardiovascular disease; they are namely high blood pressure, excessive lipids in blood, stress, obesity, metabolic syndrome, diabetes, and many more. Sometimes family history may also be a factor for the disease. Several tests are performed before the diagnosis of Cardiovascular disease; they are the electrocardiogram(ECG), blood pressure, cholesterol, blood sugar, and auscultation.

Title:A Survey on Prediction Techniques of Heart Disease using Machine Learning

Author: Mangesh Limbitote, Dnyaneshwari Mahajan, Kedar Damkondwar, Pushkar Patil

Year:2020

Heart is one of the most important part of the body. It helps to purify and circulate blood to all parts of the body. Most number of deaths in the world are due to Heart Diseases. Some symptoms like chest pain, faster heartbeat, discomfort in breathing are recorded. This data is analysed on regular basis. In this review, an overview of the heart disease and its current procedures is firstly introduced. Furthermore, an in-depth analysis of the most relevant machine learning techniques available on the literature for heart disease prediction is briefly elaborated. Heart disease is the kind of disease which can cause the death. Each year too many people are dying due to heart disease. Heart disease can occur due to the weakening of heart muscle. Also, heart failure can be described as the failure of the heart to pump the blood. Heart disease is also called coronary artery disease (CAD). CAD can occur due to insufficient blood supply to arteries.

Title: Heart Disease Prediction Using Effective Machine Learning Techniques

Author: Avinash Golande, Pavan Kumar T

Year:2019

In today's era deaths due to heart disease has become a major issue approximately one person dies per minute due to heart disease. This is considering both male and female categories and this ratio may vary according to the region. Also this ratio is considered for the people of age group 25-69. This does not indicate that the

people with other age groups will not be affected by heart diseases. This problem may start in the early age group also and predicting the cause and disease is a major challenge nowadays. Here in this paper, we have discussed various algorithms and tools used for prediction of heart diseases.

Title: Heart Disease Diagnosis and Prediction Using Machine Learning and Data Mining Techniques

Author: 1Animesh Hazra, 2Subrata Kumar Mandal, 3Amit Gupta, 4Arkomita Mukherjee and 5Asmita Mukherjee

Year: 2017

A popular saying goes that we are living in an “information age”. Terabytes of data are produced every day. Data mining is the process which turns a collection of data into knowledge. The healthcare industry generates a huge amount of data daily. However, most of it is not effectively used. Efficient tools to extract knowledge from these databases for clinical detection of diseases or other purposes are not much prevalent. The aim of this paper is to summarize some of the current research on predicting heart diseases using data mining techniques, analyze the various combinations of mining algorithms used and conclude which technique(s) are effective and efficient. Also, some future directions on prediction systems have been addressed.

CHAPTER 3

METHODOLOGY

3.1 MACHINE LEARNING

Machine learning is to predict the future from past data. Machine learning (ML) is a type of artificial intelligence (AI) that provides computers with the ability to learn without being explicitly programmed. Machine learning focuses on the development of Computer Programs that can change when exposed to new data and the basics of Machine Learning, implementation of a simple machine learning algorithm using python. Process of training and prediction involves the use of specialized algorithms. It feeds the training data to an algorithm, and the algorithm uses this training data to give predictions on a new test data. Machine learning can be roughly separated into three categories. There is supervised learning, unsupervised learning and reinforcement learning. Supervised learning programs are both given the input data and the corresponding labeling to learn data has to be labeled by a human being beforehand. Unsupervised learning has no labels. It provided the learning algorithm. This algorithm has to figure out the clustering of the input data. Finally, Reinforcement learning dynamically interacts with its environment and it receives positive or negative feedback to improve its performance.

Data scientists use many different kinds of machine learning algorithms to discover patterns in python that lead to actionable insights. At a high level, these different algorithms can be classified into two groups based on the way they “learn” about data to make predictions: supervised and unsupervised learning. Classification is the process of predicting the class of given data points. Classes are sometimes called targets/ labels or categories. Classification predictive modeling is the task of approximating a mapping function from input variables(X) to discrete output variables(y). In machine learning and statistics, classification is a supervised learning