



IJIRCCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 9, Issue 3, March 2021

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 7.488

 9940 572 462

 6381 907 438

 ijircce@gmail.com

 www.ijircce.com

Heart Disease Identification Method Using Machine Learning Techniques

Ayyapan S, Hari Haran V K, Mano K, Mr. B.Jaikumar,M.E,

Department of Computer Science and Engineering, Paavai College of Engineering, Namakkal, Tamil Nadu, India
Assistant Professor, Department of Computer Science and Engineering, Paavai College of Engineering, Namakkal,
Tamil Nadu, India

ABSTRACT: According to WHO (World Health Organization), Heart diseases are the reason for 12 million deaths every year. In most of the countries, half of the deaths are due to cardiovascular diseases. The early diagnosis of cardiovascular sicknesses can help in settling on choices on the way of life changes in high hazard patients and thusly diminish the difficulties. In this paper, machine learning techniques are used for the detection of heart disease. We also applied sampling techniques for handling unbalanced datasets. Various machine learning methods are used to predict the overall risk. The Framingham heart disease dataset is publically available on the Kaggle. This dataset is used in our experiments. The end goal is to predict whether the patient has a 10-year risk of future coronary heart disease (CHD). The datas.

KEYWORDS: Heart disease, CHD, Machine learning, sampling..

I. INTRODUCTION

There is a huge amount of data available in the Information Industry. This data is of no use until it is converted into useful information. It is necessary to analyze this huge amount of data and extract useful information from it. Extraction of information is not the only process we need to perform data mining also involves other processes such as Data Cleaning, Data Integration, Data Transformation, Data Mining, Pattern Evaluation and Data Presentation. Once all these processes are over, It would be able to use this information in many applications such as Fraud Detection, Market Analysis, Production Control, Science Exploration, etc.

1.1 DATA MINING

Data Mining is defined as extracting information from huge sets of data. Data mining is the computational process of discovering patterns in large data sets involving methods at the intersection of artificial intelligence, machine learning, statistics, and database systems. It is an interdisciplinary subfield of computer science. The overall goal of the data mining process is to extract information from a data set and

transform it into an understandable structure for further use. Aside from the raw analysis step, it involves database and data management aspects, data pre-processing, model and inference considerations, interestingness metrics, complexity considerations, post-processing of discovered structures, visualization, and online updating. Data mining is the analysis step of the "knowledge discovery in databases" process, or KDD.

1.1.1 OBJECTIVE

The main objective of the study is to identify factors related to the onset and course of coronary **heart disease** and stroke. The overall objective of our work is to predict more accurately the presence of heart disease.

1.2 LITERATURE SURVEY

1.2.1 TITLE : HEART DISEASE IDENTIFICATION BYSELECTING ALGORITHMS

AUTHOR : Jian Ping Li & Asif khan

YEAR : 2020

Heart disease is one of the complex diseases and globally many people suffered from this disease. On time and efficient identification of heart disease plays a key role in healthcare, particularly in the field of cardiology. The article, It

proposed an efficient and accurate system to diagnosis heart disease and the system is based on machine learning techniques. The system is developed based on classification algorithms includes Support vector machine, Logistic regression, Artificial neural network, K-nearest neighbor, Naïve bays, and Decision tree while standard features selection algorithms have been used such as Relief, Minimal redundancy maximal relevance, Least absolute shrinkage selection for and Local learning for removing irrelevant and redundant features. The features selection algorithms are used for features selection to increase the classification accuracy and reduce the execution time of classification system. The performance measuring metrics are used for assessment of the performances of the classifiers.

1.2.2 TITLE : DATA MINING FOR THE ENTERPRISE

AUTHOR : Charly Kleissner

YEAR : 2019

The emergence of comprehensive data warehouses which integrate operational data with customer, supplier, and market data have resulted in an explosion of information. Competition requires timely and sophisticated analysis on an integrated view of that data. However, there has been a growing gap between more powerful data warehousing systems and the users' ability to effectively analyze and act on the information they contain. Data mining tools and services are providing the leap necessary to close this gap. A comprehensive architectural overview proposes data mining integration solutions for data warehouses, application servers, thick clients, and thin clients. This article concludes with an analysis of current trends relevant to enterprise usage of data mining tools and methodologies. We anticipate an increase in standardization efforts of data mining application programming interfaces. We have only scratched the surface in our description of different data mining algorithms and methodologies. Even though there is some consolidation of core data mining methodologies going on, variations of existing algorithms as well as new ones are plentiful. This makes it challenging to compare the various methods with respect to accuracy and performance.

1.2.3 TITLE : DATA MINING CONCEPTS AND TECHNIQUES

AUTHOR : Shivam Agarwal

YEAR: 2018

Data mining is a field of intersection of computer science and statistics used to discover patterns in the information bank. The main aim of the data mining process is to extract the useful information from the dossier of data and mould it into an understandable structure for future use. There are different process and techniques used to carry out data mining successfully. The enormous usage of computers has provided a huge amount of data for one's disposal. Because of the spiralling amount of data, experts have been facing challenges in extracting useful and meaningful information from it. This has led to data mining. Data mining is a non trivial process of extraction of information which is hidden, previously unknown and is potentially useful, from large databases. Data mining can also be explained as finding the correlations in a large relational database based on the different depth of angles we analyze it. It is a powerful tool with high potential that helps the organizations or companies to increase their sales and gain more profit from the information about the dealings of their customers. Data mining provides us with the useful information that queries and reports are not able to provide us efficiently. The information that is extracted by the data mining technique is not explicitly available in the database, whereas database application only projects the information that is available in the info bank with a restricted manipulation capacity. So data mining is best described as knowledge unearthing in databases.

1.2.4 TITLE: EMPIRICAL STUDY ON APPLICATIONS OF DATA MINING TECHNIQUES IN HEALTH CARE

AUTHOR: Harlan Karun and Sri Krishnan

YEAR : 2018

The healthcare environment is generally perceived as being 'information rich' yet 'knowledge poor'. There is a wealth of data available within the healthcare systems. However, there is a lack of effective analysis tools to discover hidden relationships and trends in data. Knowledge discovery and data mining have found numerous applications in business and scientific domain. Valuable knowledge can be discovered from application of data mining techniques in healthcare system. In this study, we briefly examine the potential use of classification based data mining techniques such as Rule based, decision tree and Artificial Neural Network to massive volume of healthcare data. In particular we consider a case study using classification techniques on a medical data set of diabetic patients. Data mining is the non trivial extraction of implicit previously unknown and potentially useful information about data". Data mining technology provides a user- oriented approach to novel and hidden patterns in the data. The discovered knowledge can be used by the healthcare administrators to improve the quality of service. The discovered knowledge can also be used

by the medical practitioners to reduce the number of adverse drug effect, to suggest less expensive therapeutically equivalent alternatives. Following are some of the important areas of interests where data mining techniques can be of tremendous use in health care management .

II. SYSTEM ANALYSIS

2.1 EXISTING SYSTEM

Data mining is the process of discovering patterns in large data sets involving methods at the intersection of machine learning, statistics, and database systems. Data mining is an interdisciplinary subfield of computer science with an overall goal to extract information (with intelligent methods) from a data set and transform the information into a comprehensible structure for further use. Data mining is the analysis step of the "knowledge discovery in databases" process, or KDD. Data mining tools have been developed for effective study of medical information, in order to assist clinicians in making better diagnosis for treatment purposes. Today medical services have come a long way to treat patients with various diseases. Among the most ethal one is the heart disease problem which cannot be seen with a naked eye and comes instantly when its limitations are reached. Heart disease is one of the disease due to that death will occurred mostly, and according to the world health organization the percentage is more for that.

2.1.1 DISADVANTAGE

- Less sensitivity
- Less accuracy
- Less Performance

2.2 PROPOSED SYSTEM

In this method different input attributes has been used in order to overcome the issue of prediction of heart disease. It proposed an efficient and accurate system to diagnosis heart disease and the system is based on machine learning techniques. The proposed system uses the data mining clustering techniques to predict the heart disease. The proposed feature selection algorithm (FCMIM) is feasible with classifier support vector machine for designing a high-level intelligent system to identify heart disease. The suggested diagnosis system (FCMIM-SVM) achieved good accuracy as compared to previously proposed methods. So for that here used improved K-means Algorithm Instead of just Simple K-Means for the Accuracy of the Clustering Centred .

2.2.1 ADVANTAGE

- High accuracy.
- Intelligent and effective heart disease prediction.
- Improved decision making for prevention or the treatment of heart diseases in its early stages.

III. MODULE DESCRIPTION

- Dataset Collection
- Data Pre-processing
- Data Clustering
- Classification

Data Set Collection

Collect the Heart dataset consists of heart information a data set (or dataset) is a collection of data. Most commonly a data set corresponds to the contents of a single database table, or a single statistical data matrix, where every column of the table represents a particular variable, and each row corresponds to a given member of the data set in question Show in 6.1.

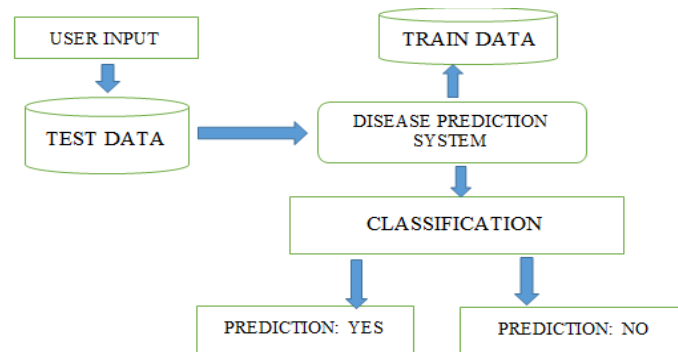


Fig.6.1 Data Set Collection

Data-Preprocessing

Data filtering, is the process of detecting and (or removing) missing records from a record set, table, or database and refers to identifying incomplete, incorrect, inaccurate or irrelevant parts of the data and then replacing, modifying, or deleting the dirty or coarse data.

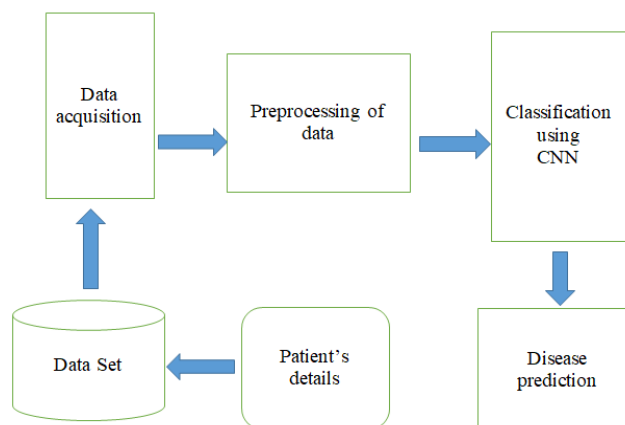
Data Clustering

Cluster analysis or clustering is the task of grouping a set of objects in such a way that objects in the same group (called a cluster) are more similar (in some sense or another) to each other than to those in other groups (clusters). The k-Means Clustering method starts with k initial clusters as specified. At each iteration, the records are assigned to the cluster with the closest centred, or centre. After each iteration, the distance from each record to the center of the cluster is calculated.

Classification

Data classification is a diverse process that involves various methods and criteria for Product dataset within a database or repository SVM classification is the process of organizing data into categories for its most effective and efficient use and predict result. This is generally done through a database or business intelligence software that provides the ability to scan, identify and separate data.

IV. SYSTEM ARCHITECTURE



SYSTEM TESTING

The purpose of testing is to discover errors. Testing is the process of trying to discover every conceivable fault or weakness in a work product. It provides a way to check the functionality of components, sub-assemblies, assemblies and/or a finished product It is the process of exercising software with the intent of ensuring that the Software system meets its requirements and user expectations and does not fail in an unacceptable manner. There are various types of test. Each test type addresses a specific testing requirement.

Unit testing

Unit testing involves the design of test cases that validate that the internal program logic is functioning properly, and that program inputs produce valid outputs. All decision branches and internal code flow should be validated. It is the testing of individual software units of the application. It is done after the completion of an individual unit before integration. This is a structural testing, that relies on knowledge of its construction and is invasive. Unit tests perform basic tests at component level and test a specific business process, application, and/or system configuration. Unit tests ensure that each unique path of a business process performs accurately to the documented specifications and contains clearly defined inputs and expected results.

Unit testing is usually conducted as part of a combined code and unit test phase of the software lifecycle, although it is not uncommon for coding and unit testing to be conducted as two distinct phases.

Test strategy and approach

Field testing will be performed manually and functional tests will be written in detail.

Test objectives

- * All field entries must work properly.
- * Pages must be activated from the identified link.
- * The entry screen, messages and responses must not be delayed.

Features to be tested

- * Verify that the entries are of the correct format
- * No duplicate entries should be allowed
- * All links should take the user to the correct page

Integration testing

Integration tests are designed to test integrated software components to determine if they actually run as one program. Testing is event driven and is more concerned with the basic outcome of screens or fields. Integration tests demonstrate that although the components were individually satisfactory, as shown by successful unit testing, the combination of components is correct and consistent. Integration testing is specifically aimed at exposing the problems that arise from the combination of components.

Software integration testing is the incremental integration testing of two or more integrated software components on a single platform to produce failures caused by interface defects.

The task of the integration test is to check that components or software applications, e.g. components in a software system or – one step up – software applications at the company level – interact without error.

Functional test

Functional tests provide systematic demonstrations that functions tested are available as specified by the business and technical requirements, system documentation, and user manuals.

Functional testing is centred on the following items:

Valid Input : identified classes of valid input must be accepted.

Invalid Input : identified classes of invalid input must be rejected.

Functions : identified functions must be exercised.

Output : identified classes of application outputs must be exercised.

Systems/Procedures: interfacing systems or procedures must be invoked.

Organization and preparation of functional tests is focused on requirements, key functions, or special test cases. In addition, systematic coverage pertaining to identify Business process flows; data fields, predefined processes, and successive processes must be considered for testing. Before functional testing is complete, additional tests are identified and the effective value of current tests is determined.

System Test

System testing ensures that the entire integrated software system meets requirements. It tests a configuration to ensure known and predictable results. An example of system testing is the configuration oriented system integration test. System testing is based on process descriptions and flows, emphasizing pre-driven process links and integration

points.

White Box Testing

White Box Testing is a testing in which in which the software tester has knowledge of the inner workings, structure and language of the software, or at least its purpose. It is used to test areas that cannot be reached from a black box level.

Black Box Testing

Black Box Testing is testing the software without any knowledge of the inner workings, structure or language of the module being tested. Black box tests, as most other kinds of tests, must be written from a definitive source document, such as specification or requirements document, such as specification or requirements document. It is a testing in which the software under test is treated, as a black box .you cannot “see” into it. The test provides inputs and responds to outputs without considering how the software works.

Acceptance Testing

User Acceptance Testing is a critical phase of any project and requires significant participation by the end user. It also ensures that the system meets the functional requirements.

V. CONCLUSION

This paper aimed to analyse the application of data mining in medical domain and some of the algorithms used to predict diseases. It is observed that results may vary for different disease diagnosis based on the tools and techniques used. Data mining provides good results in disease diagnosis when appropriate tools and techniques applied. Hence data mining is the promising field for healthcare predictions.

REFERENCES

- [1] Ahadi .R .L, Haapala .H, and Vihavainen .A,[2015], “Exploring machine learning methods to automatically identify students in need of assistance,” in Proc. 11th Annu. Int. Conf. Int. Comput. Educ. Res.,pp. 121–130.
- [2] Allen .L .A, Stevenson ,L .W, Grady .K .L, Goldstein .N .E, Matlock .D .D, Arnold R .M,Cook .N .R, Felker .G .M, Francis .G .S, Hauptman .P .J, Havranek .E .P, Krumholz .H .M, Mancini .D, Riegel .B, and Spertus .J .A [2012] , “Decision making in advanced heart failure: A scientific statement from the American heart association,” *Circulation*, vol. 125, no. 15, pp. 1928-1952.
- [3] Al-Shayea .Q .K [2001], “Artificial neural networks in medical diagnosis,” *Int. J. Comput. Sci. Issues*, vol. 8, no. 2, pp. 150-154.
- [4] Ansarullah .S .I and Kumar .P,[2019] “A systematic literature review on cardiovascular disorder identification using knowledge mining and machine learning method,” *Int. J. Recent Technol. Eng.*, vol. 7, no. 6S, pp. 1009-1015.
- [5] Bhavani thuraisingham, [2019], “A Primer For Understanding And Applying Data Mining”,[7].
- [6] Bui .A .L,Horwich .T .B, and Fonarow .G .C, [2011]“Epidemiology and risk prole of heart failure,” *Nature Rev. Cardiol.*
- [7] Charly Kleissner [2019], “Data Mining For Enterprise”,[2].
- [8] Detrano .R, Janosi .A,Steinbrunn .W, Psterer .M,Schmid .J .J, Sandhu .S, Guppy .K .H, Lee .S, and Froelicher .V,[1989], “International application of a new probability algorithm for the diagnosis of coronary artery disease,” *Amer. J. Cardiol.*, vol. 64, no. 5, pp. 304-310.
- [9]EzenCan .A,BoyerE,Kellogg .S,andBooth.S,[2015]“Unsupervisedmoeling for understanding MOOC discussion forums: A learning analytics approach,” in Proc. 5th Int. Conf. Learn. Anal. Knowl.
- [10] Gennari .J .H,Langley .P, and Fisher .D,[Sep.1989], “Models of incremental concept formation,” *Artif. Intell.*, vol. 40, nos. 13, pp. 116-1.
- [11] Ghwanmeh.S, Mohammad .A, and Al-Ibrahim .A[2011],“Innovative artificial neural networks-based decision support system for heart diseases diagnosis,” *J. Intel. Learn. Syst. Appl.*, vol. 5, no. 3, 2013, Art. no. 353-96.
- [12] Harlan Karun and Sri Krishnan , [2018] “Empirical Study On Application Of Data Mining Techniques in healthcare”,[4].
- [13] Haslina Mohd1 AndSharifah MasturaSyedMohamad,[2019-2020] “Acceptance Model Of Electronic Medical Record”, “Application Of Data Mining TechniquesTo Health CareData”,[5],[6].
- [14] Heidenreich .P .A, Trogon .J .G, Khavjou .O .A, Butler .J Dracup .K, Ezekowitz .M .D, Finkelstein .E .A, Hong .Y, Johnston .S .C,Khera .A, Lloyd-Jones .D .M, Nelson S. N, Nichol .G,Orenstein .D, Wilson P.W .F, and Woo .Y .J

- [2011], "Forecasting The future of cardiovascular disease in the unite states: A policy statement from the American heart association," *Circulation*, vol. 123, no. 8, pp. 933944.
- [15] Hung .J .L,Wang M .C, Wang .S,Abdelrasoul .M,Li .Y,[2017] and He,"Identifying at-risk students for early interventions—A time-series clustering approach," *IEEE Trans. Emerg. Topics Comput.*, vol. 5, no. 1, pp. 45–55.
- [16] Hung.J .L and Zhang .K,[2008] "Revealing online learning behavior andactivity patterns and making predictions with data mining techniques inonline teaching," *J. Online Learn. Teach.*, vol. 4, no. 4, pp. 426–436.
- [17] Ko .K .C and Leu .F .Y,[2018], "Analyzing attributes of successful learnersby using machine learning in an undergraduate computer course," in*Proc. 32nd IEEE Int. Conf. Adv. Inf. Netw. Appl. (AINA-2018)*,Krakow,Poland, pp. 801–806.
- [18] Jian Ping Li and Asif Khan[2020], " Heart Disease Identification by selecting Algorithms",[1].
- [19] Kotsiantis .S and Kanellopoulos .D,[2006] "Association rules mining: A recentoverview," *Int. Trans. Comput. Sci. Eng.*, vol. 32, no. 1, pp. 71–82.
- [20] Li .Y, Li .T, and Liu .H,[Dec 2017], "Recent advances in feature selection and its applications," *Knowl. Inf. Syst.*, vol. 53, no. 3, pp. 551577.
- [21] Lopez-Sendon .J [2011], "the heart failure epidemic," *Medicographia*, vol. 33, no. 4, pp. 363369.
- [22] Nazir .S, Shahzad .S, Mahfooz.S, and Nazir .M, [2018],"Fuzzy logic based decision support system for component security evaluation," *Int. Arab J. Inf. Technol.*, vol. 15, no. 2, pp. 224231.
- [23] Quille .K and Bergin .S,[2016] "Programming: Further factors that influence success," in *Psychology of Programming Interest Group (PPIG)*.Cambridge, U.K.: Univ. Cambridge.
- [24] ShivamAgarwal[2018],"Data Mining Concepts And Techniques",[3].
- [25] Tsanas .A,Little .M .A, McSharry P. E, and Ramig .L .O,[2011], "Nonlinear speech analysis algorithms mapped to a standard metric achieve clinically useful quantification of average Parkinson's disease symptom severity," *J. Roy. Soc. Interface*, vol. 8, no. 59, pp. 842855.
- [26] Weiguo Fan, Linda Wallace ,[2020], "Tapping The Power Of Text Mining",[8].
- [27] Weiguo Fan, Linda Wallace [2020],*The Next Generation Of Clinical Decision (Support : LinkingEvidence Practice)*,[9].



INNO  SPACE
SJIF Scientific Journal Impact Factor

Impact Factor:
7.488

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

 9940 572 462  6381 907 438  ijircce@gmail.com



www.ijircce.com

Scan to save the contact details