JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

ADVANCEMENTS IN HEART DISEASE FORECASTING: QUINE MCCLUSKEY BINARY CLASSIFIER (QMBC) UNVEILED

[1]Goli Mahender, [2]Gunnala Shruthi, [3]Jella Prathyusha,[4] B.Samirana Acharya
[1,2,3] Scholar, [4]Assistant Professor, Department of CSE Email:
Guru Nanak Institutions Technical Campus, Hyderabad

ABSTRACT:

Cardiovascular sickness is the essential justification for mortality around the world, liable for around 33% of all passings. To help clinical experts in rapidly distinguishing and diagnosing patients, various AI and information mining strategies are used to foresee the illness. Numerous analysts have created different models to help the proficiency of these forecasts. Include determination and extraction strategies are used to eliminate superfluous highlights from the dataset, in this way decreasing calculation time and expanding the proficiency of the models. In this review, we present another gathering Quine McCluskey Binary Classifier (QMBC) procedure for recognizing patients determined to have a few types of coronary illness and the people who are not analyzed. The QMBC model uses a gathering of seven models, including strategic relapse, choice tree, irregular woods, Kclosest neighbor, credulous Bayes, support vector machine, and multi-facet perceptron, and performs incredibly well on double class datasets.

1 INTRODUCTION

The term ‘‘Heart Disease’’ (HD) is used to refer to a variety of pathological disorders that have an impact on the heart and blood vessels. It encompasses a variety of heart-related conditions, including but not limited to vascular diseases and disturbances in heart rhythm. As per the World Health Organization (WHO), it is the deadliest and most devastating disease, taking over 18 million in lives a year. To diagnose it, healthcare professionals rely on a patient’s medical history and various tests, such as blood pressure, blood sugar, and cholesterol tests. Additionally, modern medical procedures like electrocardiograms, exercise stress tests, X-rays, echocardiography, coronary angiography, radionuclide tests, MRI scans, and CT scans can aid in the identification of cardiac conditions. Heart failure is the result of chronic issues that damage or weaken the heart muscles, leading to reduced ejection fraction. It is a condition that can affect both adults and children and cause severe damage to other vital organs in the body

The primary risk factors associated with heart failure are age, ethnicity, family history, hereditary factors, lifestyle choices, and pre-existing cardiovascular disease (CVD) or genetics. While it affects both men and women equally, women are more likely to develop heart failure later in life [4]. To diagnose diseases at an early stage, ML is becoming an increasingly important tool. It aims to identify patterns hidden in observations and draw conclusions that are consistent with new information. Researchers have investigated the grouping of various techniques to create hybrid models that can outperform standalone models. Typically, these models have two phases. A subset of characteristics is chosen in phase-1 using Feature Selection (FS) and Feature Extraction (FE) techniques.

1.1 OBJECTIVE

As per the World Health Organization (WHO), it is the deadliest and most devastating disease, taking over 18 million in lives a year. To diagnose it, healthcare professionals rely on a patient’s medical history and various tests, such as blood pressure, blood sugar, and cholesterol tests. Additionally, modern medical procedures like electrocardiograms, exercise stress tests, X-rays, echocardiography, coronary angiography, radionuclide tests, MRI scans, and CT scans can aid in the identification of cardiac conditions. Heart failure is the result of chronic issues that damage or weaken the heart muscles, leading to reduced ejection fraction. It is a condition that can affect both adults and children and cause severe damage to other vital organs in the body.

1.2 SCOPE OF THE PROJECT:

This flexibility allows us to leverage the strengths of different models and exploit their complementary nature, ultimately improving the overall performance. By adopting ensemble learning, we aim to maximize the predictive power of our models

and provide more robust and reliable predictions for the problem under investigation. This choice is supported by previous studies and empirical evidence that demonstrate the efficiency of ensemble learning in a variety of domains and tasks. Overall, the decision to employ ensemble learning is driven by our pursuit of improved performance, enhanced generalization, and the desire to extract the full potential from the available data. Through this approach, we expect to achieve more accurate and reliable predictions, thereby contributing to advancements in the field of the health section

.1.3 PROBLEM STATEMENT

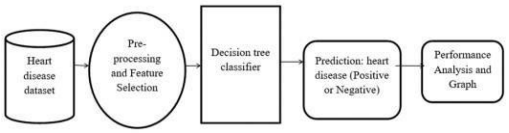
The problem addressed in this study revolves around the urgent need for efficient identification and diagnosis of cardiovascular diseases, which remain the leading cause of mortality globally. Various AI and data mining techniques have been employed to predict these diseases, aiming to support healthcare professionals in swift decision-making. However, the abundance of features in datasets poses a challenge, requiring feature selection and extraction methods to enhance computational efficiency and model effectiveness. In response, this review introduces the Quine McCluskey Binary Classifier (QMBC) approach, designed to differentiate between patients diagnosed with different types of heart diseases and those without any diagnosis. The QMBC model integrates seven distinct models and exhibits exceptional performance on binary class datasets, offering promising prospects for improving disease detection and patient management in clinicalsettings.

1.4 EXISTING SYSTEM:

Cardiovascular disease is the primary reason for mortality worldwide, responsible for around a third of all deaths. To assist medical professionals in quickly identifying and diagnosing patients, numerous machine learning and data mining techniques are utilized to predict the disease. Many researchers have developed various models to boost the efficiency of these predictions. Feature selection and extraction techniques are utilized to remove unnecessary features from the dataset, thereby reducing computation time and increasing the efficiency of the models. Additionally, modern medical procedures like electrocardiograms, exercise stress tests, X-rays, echocardiography, coronary angiography, radionuclide tests, MRI scans, and CT scans can aid in the identification of cardiac conditions.

1.4.1 Existing System Disadvantages: One main difficulty lies in analyzing such data without compromising patients' privacy and personal data, which is a primary concern in healthcare applications.Distributed data without disclosing sensitive information about the data subjects.

1.5 SYSTEM ARCHITECTURE:



1.5.1 EXPLANATION:

In this project data owner has a register all details and then login. Data owner can be an upload a document. Data owner can have a send request to the data user. Data user can search a query with uploaded document. The file has also a download it will show an encryption format. Data user also a send a request to the cloud server. Cloud server can a login. It will accept a key approve. Cloud server can also see all the data information's. Cloud server can also see all the user information. Cloud server can see all the stored information. Cloud server can approve a key request from the user. Then data owner has get the request data owner can send a secret key to the user. Then user can also download a file. If the user has given wrong keys it gets warning the user has a blockpermanently. The file it gets an attacks.

1.6 PROPOSED SYSTEM

In this paper, we propose a secure verifiable semantic.Searching scheme that treats matching between queries and documents as an optimal matching task. We treat the document words as "suppliers," the query words as "consumers," and the semantic information as "product," and design the minimum word transportation cost (MWTC) as the similarity metric between queries and documents.

1.6.1 PROPOSED SYSTEM ADVANTAGES:

Providing more security,Reducing storage cost.

For secure semantic optimal matching on theciphertext,

2 PAPER DESCRIPTION

2.1 GENERAL:

We assume that the data owner is trusted, and the data users are authorized by the data owner. The communication channels between the owner and users are secure on existing security protocols such as SSL, TLS. With regard to the cloud server, our scheme resists a more challenging security model which is beyond the "semi-honest server" used in

other secure semantic searching schemes. In our model, the dishonest cloud server attempts to return wrong/forged search results and learn sensitive information, but would not maliciously delete or tamper with the outsourced documents. Therefore, our secure semantic scheme should guarantee the verifiability, and confidentiality under such a security model.

2.2 METHODOLOGIES

- Data Collection
- Dataset
- Data Preparation
- Model Selection
- Saving the Trained Model **Data collection**

This is the first real step towards the real development of a machine learning model, collecting data. This is a critical step that will cascade in how good the model will be, the more and better data thatwe get, the better our model will perform.
There are several techniques to collect the data, like web scraping, manual interventions and etc.

Dataset

The dataset consists of 303 individual data. There are
14 columns in the dataset, which are described below.

1. ***Age***: displays the age of the individual.
2. ***Sex***: displays the gender of the individual using the following format:
   1 = male
3. 0 = female
4. ***Chest-pain type(*cp*)***: displays the type of chest-pain experienced by the individual using the following format: 1 = typical angina
   2 = atypical angina
   3 = non — anginal pain 4 = asymptotic
5. ***Resting Blood Pressure(*trestbps*)***: displays the resting blood pressure value of an individual in mmHg (unit)
6. ***Serum Cholestrol(*chol*)***: displays the serum cholesterol in mg/dl (unit)
7. ***Fasting Blood Sugar(*fbs*)***: compares the fasting blood sugar value of an individual with 120mg/dl.
   If fasting blood sugar > 120mg/dl then: 1(true)

Else: 0 (false

Data Preparation

Wrangle data and prepare it for training. Clean that which may require it (remove duplicates, correct errors, deal with missing values, normalization, and data type conversions, etc.) Randomize data, which erases the effects of the particular order in which we collected and/or otherwise prepared our data Visualize data to help detect relevant relationships between variables or class imbalances (bias alert!), or perform other exploratory analysis Split into training and evaluation sets.

Model Selection

We used Decision Tree Classifier machine learning algorithm,

We got a accuracy of 96.7% on test set so we implemented this algorithm.

Saving the Trained Model

When you have completed training and testing your model and are ready to use it in a production environment, the first step is to save it as either a .h5 or .p file. You can use libraries like TensorFlow or Pickle to do this, but make sure they are installed in your environment first. Once you have confirmed this, you can import the module and save the model as a .h5 or .p file.

2.3 TECHNIQUE USED OR ALGORITHM USED ANOVA:

ANOVA is a statistical technique that measures the significance of variations among categories or groups of data. The F-test score is used to determine the degree of variance in the target variable that can be attributed to the variance in a particular feature. The following describes how the ANOVA Algorithm 3 work. In order to exclude attributes that do not strongly correlate to the target variable, a threshold may need to be set on the F-test result. A fresh dataset created by the ANOVA technique, which only includes the chosen features, can be utilized to create an ML model. ANOVA can improve precision and effectiveness by focusing on the most crucial features.

3 FUTURE ENHANCEMENT

As future work, we plan to explore the possibility of extending the proposed framework to settings where the parties do not follow the honest but curious security model, which is beyond the scope of this work.

4 RESULT:

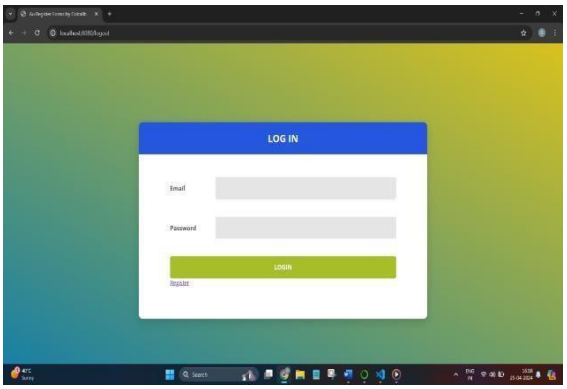


Fig 4.1 Login Page

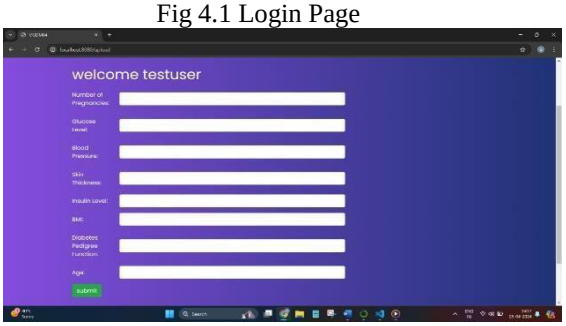


Fig 4.2 Welcome Page

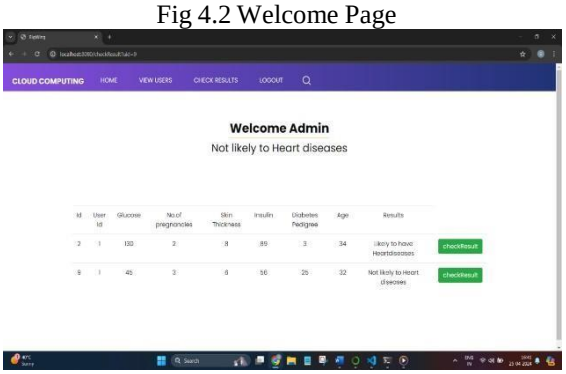


Fig 4.3 Result

This paper presents a privacy-preserving distributed algorithm using extremely randomized trees for healthcare. Evaluated with various healthcare and mental health datasets, including INTROMAT from Norway, it outperforms existing models. Results show up to 11.2% improvement in F1-score, 11.8% in ACC, and 0.232 in MCC for Depression, and up to 12.9% in F1-score, 13.2% in ACC, and 0.261 in MCC for Psykose datasets. Implemented on Amazon's AWS, it showcases scalability and low latency, handling missing values with linear overhead relative to party count. This framework delivers high-quality ML models while preserving privacy, potentially benefiting healthcare. Future work aims to expand security handling.

5 CONCLUSION

In this paper, we present the privacy-preserving distributed extremely randomized trees algorithm for learning without privacy concerns in the healthcare domain. We have evaluated our proposed algorithm extensively using two popular structured healthcare datasets and two mental health datasets associated with the Norwegian Introducing Mental health through Adaptive Technology (INTROMAT) project. Our approach outperforms the state of the art in distributed tree-based models by up to 11.2% in terms of F1-score, 11.8% in terms of ACC, and 0.232 in terms of MCC for the Depression augmented dataset, and by up to 12.9% in terms of F1-score, 13.2% in terms of ACC, and 0.261 in terms of MCC for the Psykose augmented dataset. Moreover, we present the implementation of our technique on Amazon's AWS cloud, as a proof of concept, to evaluate the latency and scalability of our framework. The proposed algorithm has linear overhead with respect to the number of parties and can also handle datasets with missing values. We demonstrated our framework's efficiency in terms of prediction performance, scalability, and overheads, as well as privacy. The proposed framework provides the possibility of developing high-quality and accurate machine learning models without privacy concerns and is expected to contribute to a better healthcare system in the long term. As future work, we plan to explore the possibility of extending the proposed framework to settings where the parties do not follow the honest-but curious security model, which is beyond the scope of this work. REFER.

6 REFERENCES

[1] C. G. D. S. E. Silva, G. C. Buginga, E. A. D. S. E. Silva, R. Arena, C. R. Rouleau, S. Aggarwal, S. B. Wilton, L. Austford, T. Hauer, and J. Myers, ''Prediction of mortality in coronary artery disease: Role of machine learning and maximal exercise capacity,'' Mayo Clinic Proc., vol. 97, no. 8, pp. 1472–1482, Aug. 2022.

[2] World Health Organization. (2009). Cardiovascular Diseases (CVDS). [Online]. Available: http://www.who.int/mediacentre/factsheets/fs317/e n/ index.html

[3] M. Ozcan and S. Peker, ''A classification and regression tree algorithm for heart disease modeling and prediction,'' Healthcare Anal., vol. 3, Nov. 2023, Art. No. 100130.

[4] M. M. Nishat, F. Faisal, I. J. Ratul, A. Al-Monsur, A. M. Ar-Rafi, S. M. Nasrullah, M. T. Reza, and M.R.H.Khan,''Acomprehensiveinves tigation of the performances of different machine learning classifiers with SMOTE-ENN oversampling technique and hyperparameter optimization for imbalanced heart failure dataset,'' Sci. Program., vol. 2022, pp. 1–17, Mar. 2022

[5] P. Ghosh, S. Azam, M. Jonkman, A. Karim, F. M. J. M. Shamrat, E. Ignatious, S. Shultana, A. R. Beeravolu, and F. De Boer, ''Efficient prediction of cardiovascular disease

using machine learning algorithms with relief and LASSO feature selection techniques,'' IEEE Access, vol. 9, pp. 19304–19326, 2021.
[6] A.K.Gárate-Escamila, A. Hajjam El Hassani, and E. Andrès, ''Classifica tion models for heart disease prediction using feature selection and PCA,'' Informat. Med. Unlocked, vol. 19, 2020, Art. No. 100330.
[7] S. Bashir, Z. S. Khan, F. H. Khan, A. Anjum, and K. Bashir, ''Improving heart disease prediction using feature selection approaches,'' in Proc. 16th Int. Bhurban Conf. Appl. Sci. Technol. (IBCAST), Jan. 2019, pp. 619–623.
[8] J. P. Li, A. U. Haq, S. U. Din, J. Khan, A. Khan, and A. Saboor, ''Heart disease identification method using machine learning classification in e healthcare,'' IEEE Access, vol. 8, pp. 107562–107582, 2020..

[8] B. Mandal, L. Li, G. S. Wang, and J. Lin, ''Towards detection of bus driver fatigue based on robust visual analysis of eye state,'' IEEE Trans. Intell. Transp. Syst., vol. 18, no. 3, pp. 545– 557, Mar. 2017.

[9] M. Ayar, A. Isazadeh, F. S. Gharehchopogh, and M. Seyedi, ''Chaotic based divide-and-conquer feature selection method and its applica tion in cardiac arrhythmia classification,'' J. Supercomput., vol. 78, pp. 5856–5882, Mar. 2022.
[10] S. Shilaskar and A. Ghatol, ''Feature selection for medical diagnosis: Evaluation for cardiovascular diseases,'' Expert Syst. Appl., vol. 40, no. 10, pp. 4146–4153, Aug. 2013.
[11] N. C. Long, P. Meesad, and H. Unger, ''A highly accurate firefly based algorithm for heart disease prediction,'' Expert Syst. Appl., vol. 42, no. 21, pp. 8221–8231, Nov. 2015.
[12] I. D. Mienye, Y. Sun, and Z. Wang, ''Improved sparse autoencoder based artificial neural network approach for prediction of heart disease,'' Infor mat. Med. Unlocked, vol. 18, Jan. 2020, Art. No. 100307. [13] C. B. C. Latha and S. C. Jeeva, ''Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques,'' Informat. Med. Unlocked, vol. 16, Jan. 2019, Art. No. 100203.
[14] R. K. Sevakula and N. K. Verma, ''Assessing generalization ability of majority vote point classifiers,'' IEEE Trans. Neural Netw. Learn. Syst., vol. 28, no. 12, pp. 2985–2997, Dec. 2017. [15] H.Li,Y.Cui,Y.Liu,W.Li,Y.Shi,C.Fang,H.Li,T.Gao,L. Hu,andY. Lu, ''Ensemble learning for overall power conversion efficiency of the all organic dyesensitized solar cells,'' IEEE Access, vol. 6, pp. 34118–34126, 2018.
[16] K. Budholiya, S. K. Shrivastava, and V. Sharma, ''an optimized XGBoost based diagnostic system for effective prediction of heart disease,'' J. King Saud Univ., Compute. Inf. Sci., vol. 34, no. 7, pp. 4514–4523, Jul. 2022.
[17] A. Gupta, R. Kumar, H. S. Arora, and B. Raman, ''MIFH: A machine intelligence framework for heart disease diagnosis,'' IEEE Access, vol. 8, pp. 14659–14674, 2020.
[18]A.Javeed,S.Zhou,L.Yongjian,I.Qasim,A.Noor,a ndR .Nour,''Anintel ligent learning system based on random search algorithm and optimized random forest model for improved heart disease detection,'' IEEE Access, vol. 7, pp. 180235–180243, 2019.
[19] D. Velusamy and K. Ramasamy, ''Ensemble of heterogeneous classifiers for diagnosis and prediction of coronary artery disease with reduced fea ture subset,'' Comput. Methods Programs Biomed. vol. 198, Jan. 2021, Art. No. 105770.