Identification of Cardio Diseases in Modern Healthcare using Machine Learning Classification

P.Priya[a], R.Aarnika[b], M.Dharani[c] and S.Santhi[d]

[a]
Department of Computer Science and Engineering, M.Kumarasamy College of
Engineering, Karur Tamil Nadu, India-639113
[b,c,d] Department of Computer Science and Engineering, M.Kumarasamy College of Engineering, Karur, Tamil Nadu, India-639113

_____

Abstract: The health region produces a massive quantity of facts. This statistics is not always made use to the full quantity and is frequently underutilized the usage of this big quantity of statistics, a ailment can be detected, predicated or maybe cured. A large hazard to human type is caused by sicknesses like heart disease, most cancers, tumour, and Alzheimer's disease prediction. Using machine getting to know strategies, the coronary heart ailment may be expected. Clinical data which includes blood strain, hypertension, diabetes, the quantity of each day cigarettes smoked, and so forth. Are used as input, so these traits are modeled to expect. This model can then be used to are expecting future clinical statistics. The algorithms like Decision Tree , k – Nearest Neighbor and Support Vector Machine are used. The accuracy of the model the use of every of the algorithm is calculated. Then the only with the good accuracy is taken because the version for predicting the coronary heart diseases.

Keywords: Data pre-processing, Classification Techniques, Feature Selection, Cardio Disease Prediction, Model Fine-tuning.

_____

1. Introduction

In our day by day lives, people keep on with a hectic, routine agenda that leads to pressure and tension. It results in illnesses like heart disease, most cancers, and many others. The assignment behind these sicknesses is the way to predict them. All of us have one of a kind cost of pulse fee and blood stress. However, it's been clinically demonstrated that the pulse ought to be among 60 and one hundred beats in line with minute and the blood pressure should be between 120/80 and one hundred forty/90. Coronary heart ailment is one of the important purposes of loss of life in the global. The coronary heart way cardio. Consequently, any heart ailment is assessed as a cardiovascular sickness. Coronary heart ailment or coronary artery sickness is the narrowing of the coronary arteries [24]. The coronary arteries deliver the coronary heart with oxygen and blood. It reasons ailment or loss of life in a large variety of human beings. It's miles one of the famous sort of heart sickness. Excessive blood glucose from diabetes can harm blood vessels and nerves that control coronary heart and blood vessels. If a person has had diabetes for an extended period of time, it's miles in all likelihood that they will have coronary heart disease in the destiny. The opposite hazard factors consist of age, gender, stress, and unhealthy diet. Chance of having a coronary heart sickness will increase while someone is growing old. Guys are at a better chance for coronary heart disease. So, in this paper based totally at the elements to expect the hazard of coronary heart disorder.

2.Technical background:

To predict a heart disease, the technologies like *Machine Learning, Deep Learning* and *Artificial Intelligence* are helpful.

2.1. Machine Learning:

Machine Learning can be defined as the ability of the automated learning without programming explicitly, it makes the ML model to learn like a human automatically. [1]
The different types of ML are
-   *Supervised Learning*
-   *Unsupervised learning*
-   *Semi-supervised learning*
-   *Reinforcement learning*

2.1.1. Supervised Learning :

Learning with the guidance and labeled data. We have labeled data in this form of model, where the correct input and exact output that is expected are clearly defined. This supervised learning model has two subtypes: regression and classification.[3]

2.1.2. Unsupervised learning:

_____

Learning from the data without a proper mark. The input is explicitly stated, but there is no indication of what we will receive as a result. Clustering and association are two of the unsupervised learning subtypes.[30]

2.1.3. Semi supervised learning:

Another form of learning is semi-*supervised learning*, which is a mixture of previous two learning types. [30]

2.1.4. Reinforcement Learning:

A reward based type, Where the model will based on the rewards, it's a type of feedback based learning. Here, the model will learn completely without any proper guidance.

2.2. Important Machine Learning Algorithms:

The majority of the algorithms in this project are classification algorithms [24].. Since classification algorithms allow you to work with data that has been classified. So, when we're working on prediction, we'll use the classification algorithms [5] that fall under Supervised learning. There are a variety of classification algorithms available. A few of them are mentioned further down.

2.2.1. Decision Tree :

Algorithm to find solution for classification problems, but it can also be used to solve regression problems. It's in the shape of a tree since it has decision nodes and leaf nodes. It can be used in data where input and output data are clearly separated for train. Activity of human decision making.[13] [27] and [28] .

2.2.2. K-Nearest Neighbor:

It's also a classification type applied to*Supervised Learning* methods. It has the potential to produce highly competitive outcomes. The key drawback of the KNN algorithm is that as the size of the data grows larger, it becomes slower.

2.2.3. Naive Bayes Classifier:

Based on Bayes theorem. This Nave Bayes Classifier can be used to create fast Machine Learning models. It's mainly used in a number of forecasting models. Some naive assumptions are made in the nave Bayes formulation. When we use it for classification purposes, it performs better than other versions. [26]

2.2.4. Support Vector Machine:

It's can classify, predict, and even detect outliers. It works by connecting two groups with a straight line.[44] [53]

2.2.5. Random Forest Classifier:

The random decision forests are another name for it. It is a form of ensemble learning [31]. It is achieved by creating several decision trees during the training phase and then combining the results.

In modern ML, the Random Forest Approach, this is the maximum used model. It's an example of ensemble-based methodology. It's made worse with the use of Decision Trees, which we've already looked at. [29]

2.2.6. Logistic Regression:

It's used to estimate the dependent variable among the independent variables. A discrete value between 0 and 1 is the product of logistic regression.

2.2.7. K- folds cross validation:

It's a technique for validating or evaluating machine learning models with handy amount of data. Sample numbers we need to divide is denoted by the letter k.

2.2.8. Artificial Neural Networks:

Artificial Neural Networks, or Neural Networks, are a form of artificial neural network. It was created using the biological neural networks model as inspiration. It is made up of a series of neurons, which are smaller units that are also known as nodes.[13][50]

2.3. Feature Selection Algorithms:

It is the process of reducing the variables as inputs into the prediction model that we are building. That is to say, it can also be defined as selecting the appropriate features. There are a number of function selection templates to choose from. The following are a few of them.[42]

2.3.1. Relief Feature Selection Algorithm:

The RFSA algorithm is a binary classification sensitive feature selection algorithm. It will be extremely beneficial in improving the classification algorithms' performance. [32][33]

2.3.2. Minimum Redundancy Maximum Relevance:

mRMR is commonly used with classification algorithms. As the name implies, the main objective is to minimise duplication while optimising relevance in the function we've selected. [32] [34]

2.3.3. Least Absolute Shrinkage and Selection Operator:

It is to perform the variable selection and regularization in order to accuracy improvement in prediction in statistics and in ML. It is initially made for the linear regression models, then it is also used for the classification models. [35]

2.3.4. Local Learning Based Feature Selection Algorithms:

By assigning weights to the functions, LLBFS reduces the complexity of Non-Linear Problems to Linear. The broad weighted and low weighted values are used to pick and reject candidates. [36]

2.3.5. Fast Conditional Mutual Information:

Another form of efficient feature selection model is the FCMIM. The design is based on Conditional Mutual Information in this case (CMI). [37] [38]

2.4. Deep Learning :

Machine learning focused on artificial neural networks ANN [40] includes deep learning as well. The model will be based on supervised, unsupervised, and semi-supervised learning methods in this case.[4]

2.5. Data Science :

Data Science is a discipline in which data is processed to obtain information and insight, which is then applied to a number of domains.

2.6. Artificial Intelligence :

Artificial Intelligence is the process of programming a computer to imitate human behavior. In this case, the computer would be intelligent in the same way as a person is. [25]

2.7. Data Mining:

Data mining means extracting information and ideas with data and applying them to useful purposes. Current programs, for example, HD finding with Data Mining. It is combined with a variety of hybrid intelligence and hybrid machine learning approaches to provide high precision in heart disease prediction.

The use of machine learning models in conjunction with data mining improves the performance of prediction and recommendation-based systems. [45][46][47][48]

Various technologies and methods, such as Fuzzy Logic, Prediction, Hybrid techniques [41], and Ensemble learning techniques, are used in conjunction with these technologies.[43]

3. Steps followed in Machine Learning :

The flow of steps in Machine Learning is displayed in figure 1 to find the flow of Machine Learning models.
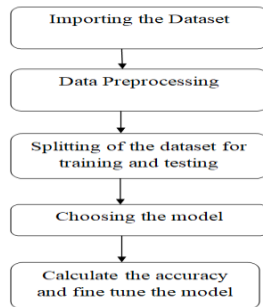

Figure 1: Sequence of steps in Machine Learning

● Choosing the dataset

It is necessary to choose a dataset that is appropriate for the model. It is available for free from a variety of websites on the internet. Kaggle and data.world are two examples of websites. Training and testing the model would be beneficial.

● Preprocessing the dataset

Following the selection of a dataset, the next step is to process it so that it is more appropriate for our model. Cleansing data, deleting unused data, and filing null values are all examples of data preprocessing. Since Machine Learning works with numerical data, this step also involves translating the characters to numerical values.

● Dataset split for test and train

The data is split for training and testing after it has been preprocessed for our model. Almost 70% of the data is used for preparation, while the rest is used for research.

● Fitting Model

After splitting the data, the next step is to pick the model that is best suited to our model. The accuracy will improve if you choose the right model. As a result, we must choose the appropriate model for our framework. [2]

● Calculating accuracy and tuning

After that, we must train and test our model. The model's accuracy is verified after it has been trained and tested. We need a best-fit model because, in Machine Learning, the model should not be the worst-fit or the most over-fit. As a result, we'll need to fine-tune our model to boost its accuracy.

● Change the model

If the model we've selected isn't accurate enough after fine tuning, we'll need to move to a different model that's better suited to our framework.

Mostly Used Algorithms in the previously studied models:

➢ *Naive Bayes*
➢ *Support Vector Machine*
➢ *Decision Trees*
➢ *Random Forest Classifiers*

4. Dataset Selection:

Term "dataset" refers to a grouping of similar data. It's available on a variety of websites, including Kaggle, Data World, and others. The dataset used in different forms of proposed works is the *Cleveland HD dataset*. 14 main features of the dataset are used in different proposed models. Age of the patient, their gender and sugar level etc. The feature name, feature code and feature description of the mostly used Cleveland heart disease dataset is displayed in table 1.

4.1. HD Cleveland dataset:

Table 1: Features of Cleveland Heart Disease Dataset

S.no	Feature Name	Feature Code	Description
1	Age	AGE	Age in years
2	sex	SEX	Male=1,Female=0
3	chest pain	CPT	Atypical angina=1 Typical angina=2 Asymptomatoic=3 Non-angial pain=4
4	resting blood pressure	RBP	mm hg, hospitalized
5	serum cholesterol	SCH	In mg/dl
6	fastingbloodsugar > 120mg/dl	FBS	fastingbloodsugar > 120mg/dl (T=1) (F=0)
7	resting electrocardiographic	RES	Normal=0 ST T=1 Hypertrophy=2
8	maximum heart rate	MHR	-----
9	exercise induced angina	EIA	yes=1 no=0
10	old peak=ST depression induced by exercise relative to rest	OPK	-----
11	The slope of the Peak Exercise ST Segment	PES	Up Sloping=1 Flat=2 Down Sloping=3
12	number of major vessels (0-3) Colored by fluoroscopy	VCA	
13	thallium scan	THA	Normal=3 Fixed defect=7 Reversible defect=7
14	label	LB	Heart disease patient=1 Healthy=0

5. Literature Study:

Various study works are listed. Some of the most important contributions that are used to forecast heart disease would be more useful in predicting heart disease sooner. Every year, a large number of new papers are published that attempt to predict heart disease using different technologies. Figure 2 explains the architecture studied from existing models.

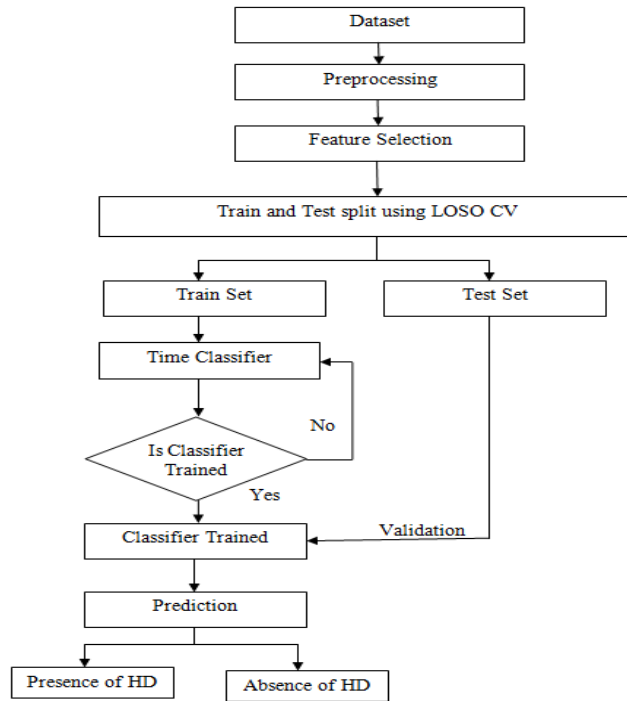### 5.1. Existing System Architecture:

Figure 2: Architecture of Existing system to find HD

5.2. Algorithm Usage:

There are various algorithms that are used in the HD identification. In a model proposed by Kahramanli *et al.* [6] uses the Artificial Neural Networks and Fuzzy logic to predict diabetes and heart disease. In an another model proposed by R. Das, I. Turkoglu*et al.* [7] uses ensemble type learning techniques for the predict of the HD. In an another model proposed by Palaniappan*et al.* [8] uses Naïve Bayes, DT and ANN technologies along with the data mining techniques for the effective diagnosis of the heart disease.

The model proposed by E. O. Olaniyi*et al.* [9] uses the three phase ANN technique for the heart disease prediction using arbitration. In an another model proposed by E. O. Olaniyi*et al.* [10] forms an integrated decision support system based on ANN and Fuzzy_AHP for heart failure risk prediction. In a model proposed by X. Liu *et al.* [11] forms A hybrid classification system for heart disease diagnosis based on the RFRS method for the feature selection. Mohan*et al.* [12] forms effective heart disease prediction using hybrid machine learning techniques. Also in [39]. In a model proposed by Praveen Kumar Reddy M*et al.* [14] uses SVM or Decision Tree for predicting heart diseases. These two algorithms were taken from the Scikitlearn library of Python. Here the Pandas library is used for the purpose of Manipulation and for the purpose of Visualization, Matplotlib library and WEKA tool is used. From, that model it can be found that Decision Tree is better to use with the categorical data. These are the some of the models and the algorithms that are used in the forecast of the HD.

In model [18] proposed by NidhiBhatla*et al.* the model gave good accuracy of 90% in Naive Bayes, but the model while using decision tree it gave the accuracy of 99.62%. Same as this model, a model newly proposed by Kiyong Noh *et al.* [19] used Naïve Bayes and Decision tree which also showed a good accuracy of 99%.E. Loukis *et al.* [23] proposed a model which also uses decision trees which is a representation of decision tree for the purpose of accurate prediction. They also used the SVM, but SVM gave poor results of 55% when decision trees gave the result of 100%.

AnimeshHazra*et al.* [21] proposed a model which is done by using the clustering and classification algorithms.

A.S.ThamujaNishandi [15] proposed a model for predicting heart diseases in Logistic Regression of ML by Jupytelab, as it uses Logistic Regression but it doesn't provide an efficient accuracy. The model proposed by used L1 norm SVM [51] [52]and validation is done by *K-fold cross validation* and various*feature selection* is done by some specific algorithms.

Bertsimas, J. Dunn *et al.* [20] proposed a model that is created by using the structured and unstructured data from the hospital.SSD (Single Shot Multi Box Detector) algorithm is a type of algorithm that is used for analyzing the disease in the human with the help of image processing. A review by MotilalTayade [22] on the image processing technology in medical field clearly explains about the technology health care industry in more detail.

5.3. Pseudo Code of the Existing models:

1. Begin
2. Preprocessing the dataset that is collected
3. Train the Classifiers
4. Using testing dataset, validate the model
5. Performance is Computed
6. End

5.4. Advantages:

Every model has its own set of benefits and drawbacks. The following are some of the major benefits that have been found in some of the most significant models that have already been completed.

In a model proposed by Kahramanli *et al.* [6] has the advantage of high accuracy because it uses the Artificial Neural Network techniques along with the Fuzzy Logic.Also [49] . Model by R. Das *et al.* [7] is also similar to the previous model as it has the advantage of the high accuracy. In an another model proposed by Palaniappan *et al.* [8] has the advantage that the high performance inaccuracy performance is achieved, so that model looked more better. . O. Olaniyi*et al.* [9] proposed a high accuracy model and they also proposed an another model with more higher accuracy [10]. X. Liu*et al.* [11] proposed a model that has a high accuracy of nearly 92% this is achieved because of the correct feature selection. The Hybrid ML model proposed by Mohan*et al.* [12] has a advantage of low computation time. Usage of Decision tree in [14] shown good results over SVM. In the model [17] the better feature selection is done so the accuracy measured for that model stands nearly 99% which is greater than all other models, so we can say that better feature selection will lead to the better accuracy.

Bertsimas, J. Dunn *et al.* [20] proposed model helps for the patient's to track the doctor about their availability and at the same time it is used to consult the doctor in name of WEB PROCTOR. This would be a greatest advantage for the heart patient. Because, it'll help them more as it minimizes the time of going to hospital for check up.

5.5. Limitations:
In a model proposed by Kahramanli *et al.* [6] has the limitations that the model needs the more execution time to predict the result. In the another model which is proposed by R. Das *et al.* [7] has the limitation that is computationally complex. In an another model proposed by Palaniappan *et al.* [8] the performance of NB and DT is more low. O.Olaniyi *et al.* [9] has the high computation time and an another model proposed by them is computationally complex [10]. The computation time is high in the model proposed by X. Liu *et al.* [11]. Mohan *et al.* [12] proposed a model using Hybrid ML model, but the accuracy of that model is very low when compared to the other model.

From the above models, we can identify that the model which has the good feature selection and the neural network based selection will give the high accuracy of the model. So, the model which is taken great from the feature selection has a high accuracy and low computation time. Selective model for prediction is required fro better performance. SVM doesn't work well for categorical data in [14].

Table 2 shows the comparison of various models with findings of its algorithms, advantages and its limitations.

5.6. Comparison of models:

Table 2: Comparison of Existing Models

Paper	Algorithm	Limitation	Advantage
[6]	Artificial Neural Network and Fuzzy Logic	More execution time	High Accuracy
[7]	Ensemble based learning	Computationally complex	High Accuracy
[8]	Naive Bayes, ANN and DT	Performance of NB and DT is low	High Performance
[9]	Artificial Neural Networks	Time is high for computation	Accuracy is high
[10]	ANN, Fuzzy Logic and AHP	Complex in computation	High Accuracy
[11]	Relief Rough Set Based Method	Time is taken high for computation	Achieved High Accuracy than any other model
[12]	Hybrid Machine Learning Method	Low Accuracy	Low computation time
[14]	SVM and Decision Tree	SVM doesn't gives good performance	Decision trees gave good results
[15]	Logistic Regression	Low Accuracy, due to less data	Difference between positive and negative rate shown clearly
[16]	Data Mining and Hybrid Intelligent Techniques	Hybrid Intelligent Techniques doesn't gave good results	Data Mining gave good results in prediction
[17]	L1 Norm SVM with feature selection	Without feature selection it will not give better accuracy	The accuracy was high nearly to 99%

6. Pseudo Code of the Proposedmodel:

1. Begin

2. Preprocessing the dataset that is collected
3. Using Feature Selection Algorithms, Select the features
4. Classifier training
5. Validate the model with train and test dataset
6. Calculation of Performance
7. End

7. Conclusion:

Our project's key goal is to predict heart diseases. Several previous articles are examined in this article. Every Machine Learning Algorithm used in various papers has its own set of benefits and drawbacks that will affect the model's performance. As a consequence, we must construct a model that is both effective and reliable.

References:

Rabi Narayan Behera, Kajaree Das, A Survey on Machine Learning: Concept, Algorithms and Applications in IJIRCCE, February 2017.

Sebastian Raschka, Model Evaluation, Model Selection, and Algorithm Selection in Machine Learning, November 2018

Iqbal Muhammad and Zhu Yan, Supervised Machine Learning Approaches: A Survey, IJSC_Paper_4_946-952

Ajay Shrestha, Ausif Mahmood, Review of Deep Learning Algorithms and Architectures, ACCESS.2019.2912200

Murugesan, M., Thilagamani, S. ,” Efficient anomaly detection in surveillance videos based on multi layer perception recurrent neural network”, Journal of Microprocessors and Microsystems, Volume 79, Issue November 2020, https://doi.org/10.1016/j.micpro.2020.103303

H. Kahramanli and N. Allahverdi, ''Design of a hybrid system for the diabetes and heart diseases,'' Expert Syst. Appl., vol. 35, nos. 1–2, pp. 82–89, Jul. 2008.

R. Das, I. Turkoglu, and A. Sengur, ''Effective diagnosis of heart disease through neural networks ensembles,'' Expert Syst. Appl., vol. 36, no. 4, pp. 7675–7680, May 2009.

S. Palaniappan and R. Awang, ''Intelligent heart disease prediction system using data mining techniques,'' in Proc. IEEE/ACS Int. Conf. Comput. Syst. Appl., Mar. 2008, pp. 108–115.

E. O. Olaniyi, O. K. Oyedotun, and K. Adnan, ''Heart diseases diagnosis using neural networks arbitration,'' Int. J. Intell. Syst. Appl., vol. 7, no. 12, p. 72, 2015.

Thilagamani, S., Nandhakumar, C. .” Implementing green revolution for organic plant forming using KNN-classification technique”, International Journal of Advanced Science and Technology, Volume 29 , Isuue 7S, pp. 1707–1712

X. Liu, X. Wang, Q. Su, M. Zhang, Y. Zhu, Q. Wang, and Q. Wang, ''A hybrid classification system for heart disease diagnosis based on the RFRS method,'' Comput. Math. Methods Med., vol. 2017, pp. 1–11, Jan. 2017.

Thilagamani, S., Shanti, N.,” Gaussian and gabor filter approach for object segmentation”, Journal of Computing and Information Science in Engineering, 2014, 14(2), 021006, https://doi.org/10.1115/1.4026458

M. Gudadhe, K. Wankhade, and S. Dongre, ''Decision support system for heart disease based on support vector machine and artificial neural network,'' in Proc. Int. Conf. Comput. Commun. Technol. (ICCCT), Sep. 2010, pp. 741–745.

Praveen Kumar Reddy M, T Sunil Kumar Reddy, S.Balakrishnan, Syed MuzamilBasha, Ravi Kumar Poluru, Heart Disease Prediction using Machine Learning Algorithms, IJITEE, vol.8, Aug 2019

Rhagini, A., Thilagamani, S. ,”Women defence system for detecting interpersonal crimes”,International Journal of Advanced Science and Technology, 2020, Volume 29,Issue7S, pp. 1669–1675

A. Hack, J. Lee, M. Memon, J. Khan, A. Malik, T. Ahmad, A. Ali, S. Nazir, I. Ahad, M. Shahid, "Feature Selection Based on the L1 Support Vector Machine Standard and an Effective Parkinson's Disease Detection System Using Voice Recordings," IEEE Access, Vol. 7, pp. 37718-37734, 2019.

K.Deepa, S.Thilagamani, "Segmentation Techniques for Overlapped Latent Fingerprint Matching", International Journal of Innovative Technology and Exploring Engineering (IJITEE), ISSN: 2278-3075, Volume-8 Issue-12, October 2019. DOI: 10.35940/ijitee.L2863.1081219

NidhiBhatla, Kiran Jyoti, "An Analysis of Heart Disease Predictionusing Different Data Mining Techniques", 2012, IJ RT, Vol 1, Issue 8

Kiyong Noh, HeonGyu Lee, Ho-Sun Shon, Bum Ju Lee, and Keun Ho Ryu,"Associative Classification Approach for Diagnosing Cardiovascular Disease",Springer, Vol:345, pp: 721- 727, 2006

Santhi, P., Priyanka, T.,Smart India agricultural information reterival system, International Journal of Advanced Science and Technology, 2020, 29(7 Special Issue), pp. 1169–1175.

AnimeshHazra, Arkomita Mukherjee, Amit Gupta, Mukherjee, Heart Disease Diagnosisand Prediction Using Machine Learning and Data Mining Techniques: A Review, Research Gate Publications,July 2017, pp.2137-2159.

Santhi, P., Lavanya, S., Prediction of diabetes using neural networks, International Journal of Advanced Science and Technology, 2020, 29(7 Special Issue), pp. 1160–1168

E. Loukis and M. MaragoudakisHeart murmur identification usingrandom forest in assistive environmentsPETRA'10, June 23 - 25,Samos, Greece, 2010

R. Detrano, A. Janosi, W. Steinbrunn, M. Pfisterer, J.-J. Schmid, S. Sandhu, K. H. Guppy, S. Lee, and V. Froelicher, ''International application of a new probability algorithm for the diagnosis of coronary artery disease,'' Amer. J. Cardiol., vol. 64, no. 5, pp. 304–310, Aug. 1989.

Deepa. K , LekhaSree. R , Renuga Devi. B , Sadhana. V , Virgin Jenifer. S ,"Cervical Cancer Classification", International Journal of Emerging Trends in Engineering Research, 2020, 8(3), pp. 804–807 https://doi.org/10.30534/ijeter/2020/32832020

Pouria Kaviani, SunitaDhotre, Short Survey on Naïve Bayes Algorithm, IJAERD, Vol.4 November 2017

Vijayakumar, P, Pandiaraja, P, Balamurugan, B & Karuppiah, M 2019, 'A Novel Performance enhancing Task Scheduling Algorithm for Cloud based E-Health Environment', International Journal of E-Health and Medical Communications (IJEHMC), Vol 10,Issue 2,pp 102-117.

Leonard A. Breslow and David W. Aha, Simplifying Decision Trees : A Survey, NCARAI technical report no. AIC-96-104

Paulo Angelo Alves Resende and Andre Costa Drummond, A Survey of Random Forest Based Methods for Intrusion Detection Systems, ACM Computing Surveys Vol.51, May 2018

P. Pandiaraja, N Deepa 2019 ," A Novel Data Privacy-Preserving Protocol for Multi-data Users by using genetic algorithm" , Journal of Soft Computing , Springer , Volume 23 ,Issue 18, Pages 8539-8553 ,

Xibin DONG, Zhiwen YU, Wenming CAO, Yifan SHI1, Qianli MA, A Survey on Ensemble Learning

N Deepa , P. Pandiaraja, 2020 ," Hybrid Context Aware Recommendation System for E-Health Care by merkle hash tree from cloud using evolutionary algorithm" , Journal of Soft Computing , Springer , Volume 24 ,Issue 10, Pages 7149–7161

A. M. D. Silva, Feature Selection, vol. 13. Berlin, Germany: Springer, 2015, pp. 1–13.

A. Unler, A. Murat, and R. B. Chinnam, ''Mr2PSO: A maximum relevance minimum redundancy feature selection method based on swarm intelligence for support vector machine classification,'' Inf. Sci., vol. 181, no. 20, pp. 4625–4641, Oct. 2011.

N Deepa , P. Pandiaraja, 2020 , " E health care data privacy preserving efficient file retrieval from the cloud service provider using attribute based file encryption ", Journal of Ambient Intelligence and Humanized Computing , Springer , https://doi.org/10.1007/s12652-020-01911-5

F. E. Harrell, Jr., ''Ordinal logistic regression,'' in Regression Modeling Strategies. Cham, Switzerland: Springer, 2015, pp. 311–325.

F. Fleuret, ''Fast binary feature selection with conditional mutual information,'' J. Mach. Learn. Res., vol. 5, pp. 1531–1555, Nov. 2004.

R. Alzubi, N. Ramzan, H. Alzoubi, and A. Amira, ''A hybrid feature selection method for complex diseases SNPs,'' IEEE Access, vol. 6, pp. 1292–1301, 2018

H. Kahramanli and N. Allahverdi, ''Design of a hybrid system for the diabetes and heart diseases,'' Expert Syst. Appl., vol. 35, nos. 1–2, pp. 82–89, Jul. 2008.

Vijayakumar, P ,Pandiaraja, P, , Karuppiah, M & Deborah, LJ 2017, 'An Efficient Secure Communication for Healthcare System using Wearable Devices', Journal of Computers and Electrical Engineering, Elsevier , Vol .No 63 , October 2017 , pp 232-245

X. Liu, X. Wang, Q. Su, M. Zhang, Y. Zhu, Q. Wang, and Q. Wang, ''A hybrid classification system for heart disease diagnosis based on the RFRS method,'' Comput. Math. Methods Med., vol. 2017, pp. 1–11, Jan. 2017.

A. U. Haq, J. Li, M. H. Memon, M. H. Memon, J. Khan, and S. M. Marium, ''Heart disease prediction system using model of machine learning and sequential backward selection algorithm for features selection,'' in Proc. IEEE 5th Int. Conf. Converg. Technol. (ICT), Mar. 2019, pp. 1–4.

K Sumathi, P Pandiaraja 2019," Dynamic alternate buffer switching and congestion control in wireless multimedia sensor networks" , Journal of Peer-to-Peer Networking and Applications , Springer , Volume 13,Issue 6,Pages 2001-2010.

G. G. N. Geweid and M. A. Abdallah, ''A new automatic identification method of heart failure using improved support vector machine based on duality optimization technique,'' IEEE Access, vol. 7, pp. 149595–149611, 2019.

M. Durairaj, K. Meena, S. Selvaraju, "Applying a data mining approach of rough sets on spermatological data analysis as predictors of in-vitro fertility of bull semen" International Journal of Computer, Mathematical Sciences and Applications, 2, 189-199, 2008

Pandiaraja, P, Vijayakumar, P, Vijayakumar, V & Seshadhri, R 2017, 'Computation Efficient Attribute Based Broadcast Group Key Management for Secure Document Access in Public Cloud', Journal of Information Science and Engineering, 33, No. 3, pp. 695-712

SellappanPalaniappan, RafiahAwang "Intelligent Heart Disease Prediction System Using Data Mining Techniques", IEEE, July 2015

A. Methaila, P. Kansal, H. Arya, and P. Kumar, "Early heart disease prediction using data mining techniques," in Proceedings of Computer Science & Information Technology (CCSIT-2014), vol. 24, pp. 53–59, Sydney, NSW, Australia, 2014.

O. W. Samuel, G. M. Asogbon, A. K. Sangaiah, P. Fang, and G. Li, "An integrated decision support system based on ANN and Fuzzy_AHP for heart failure risk prediction," Expert Systems with Applications, vol. 68, pp. 163–172, 2017

NorsaliniSalim., "Medical diagnosis using Neural Networks", Faculty of Information Technology, University Utara Malaysia, Sintok, Kedah, 2004.

P. S. Bradley and O. L. Mangasarian, ''Feature selection via concave minimization and support vector machines,'' in Proc. 15th Int. Conf. (ICML), 1998, pp. 82–90.

M. Moon and K. Nakai, ''Stable feature selection based on the ensemble L1-norm support vector machine for biomarker discovery,'' BMC Genomics, vol. 17, p. 1026, Dec. 2016.

N. Cristianini and J. Shawe-Taylor, An Introduction to Support Vector Machines. Cambridge, U.K.: Cambridge Univ. Press, 2000, p. 23.