Assessment of Machine Learning Classifiers for Heart Diseases Discovery

Chapter · January 2022

DOI: 10.1007/978-3-030-95947-0_31

4 authors:

Roseline Ogundokun
Landmark University

Sanjay Misra
Institute for Energy Technology

Peter Sadiku
University of Ilorin

Adeniyi Jide
Landmark University

Assessment of Machine Learning Classifiers for Heart Diseases Discovery

Roseline Oluwaseun Ogundokun[1(✉)], Sanjay Misra[2], Peter Ogirima Sadiku[3], and Jide Kehinde Adeniyi[1]

[1] Department of Computer Science, Landmark University Omu Aran, Omu-Aran, Nigeria
ogundokun.roseline@lmu.edu.ng
[2] Department of Computer Science and Communication, Ostfild University College, Halden, Norway
[3] Department of Computer Science, University of Ilorin, Ilorin, Nigeria

Abstract. Heart disease (HD) is one of the utmost serious illnesses that afflict humanity. The ability to anticipate cardiac illness permits physicians to deliver better knowledgeable choices about their patient's wellbeing. Utilizing machine learning (ML) to minimize and realize the symptoms of cardiac illness is a worthwhile decision. Therefore, this study aims to analyze the effectiveness of some supervised ML procedures for detecting heart disease in respect to their accuracy, precision, f1-score, sensitivity, specificity, and false-positive rate (FPR). The outcomes, which were obtained using python programming language were compared. The data employed in this investigation came from an open database of the National Health Service (NHS) heart disease which originated in 2013. Through the machine learning (ML) technique, a dimensionality reduction technique and five classifiers were employed and a performance evaluation between the three classifiers- principal component analysis (PCA), decision tree (DT), random forest (RF), and support vector machine (SVM). The NHS database contains 299 observations. The system was evaluated using confusion matrix measures like accuracy, precision, f1-score, sensitivity (TPR), specificity, and FPR. It is concluded that ML techniques reinforce the true positive rate (TPR) of traditional regression approaches with a TPR of 98.71% and f-measure value of 68.12%. The true positives rate which is the same as the sensitivity was used to evaluate the accuracy of the classifiers and it was deduced that the PCA + DT outperformed that of the other two with a sensitivity of 98.71% and since the value is on the high side, this implies that the classifier will be able to accurately detect a patient with HD in his or her body.

Keywords: Machine learning · Heart disease · Confusion matrix · Classification · Dimensionality reduction

Abbreviations

PCA	Principal Component Analysis
ML	Machine learning
DT	Decision Tree

M. Themistocleous and M. Papadaki (Eds.): EMCIS 2021, LNBIP 437, pp. 441–452, 2022.
https://doi.org/10.1007/978-3-030-95947-0_31

RF Random Forest
SVM Support Vector Machine
FE Feature extraction
FS Feature Selection
HD Heart Disease
DM Data Mining

1 Introduction

Cardiovascular illnesses are the biggest cause of mortality globally, as reported World Health Organization (WHO) [1], with 17.9 m persons dying per annum. Obesity, high blood pressure, hyperglycemia, and excessive saturated fat are all linked to an increased risk of heart disease [1]. Additionally, the American Heart Association [2] associates an increase in weight for instance 1–2 kg daily, sleeping difficulties, limb edema, persistent cough, and a fast heart rate with HD symptoms. Due to the symptoms' nature being similar to other diseases or mistaken with indications of age, diagnosis is a challenge for practitioners. Physicians now have a new chance to enhance patient diagnosis because of the rise in medical data gathering [3]. Physicians have expanded their use of computer technology to aid administrative activities in recent years.

Machine learning (ML) is becoming an essential tool in the healthcare sector to assist with patient diagnosis. ML is a diagnostic technique that is employed once a job is big and complex to analyze, for instance, converting a healthcare account of events into knowledge, making epidemic forecasts, or analyzing genetic data [4–6]. Machine learning methods have been utilized in recent research to detect and forecast various heart issues. Melillo et al. [7] were instrumental to the development of an automated classifier for affected roles with congestive heart failure (CHF) that distinguishes between those at low and high risk. The sensitivity and specificity of the classification and regression tree (CART) were calculated to be 93.3% and 63.5%, respectively. To discover the optimum set of characteristics and enhance performance, Al Rahhal et al. [8] developed a deep neural network (DNN) categorization of electrocardiogram (ECG) data. Guidi et al. [9] were instrumental to the development of a clinical decision support system (CDSS) for heart failure (HF) examination. They examined the performance of several ML classification techniques, including neural networks (NN), SVM, CART-based fuzzy rules, and RF. With an accuracy of 87.6%, the CART method and RF achieved the greatest results. Zhang et al. [10] used natural language processing (NLP) and the rule-based approach to find an NYHA class for HF from amorphous medical records, with an accuracy of 93.37%. Parthiban et al. [11] investigated an SVM method for diagnosing HD in diabetic individuals, achieving a 94.60% accuracy and accurately predicting characteristics like age, gore pressure, and gore sugar.

The large dimensionality of the dataset is a significant issue in machine learning [12]. Because analyzing numerous features takes a lot of retention and results in overfitting, weighting attributes, reducing repetitious data, and converting period, increasing the algorithm's accomplishment [13–17]. Different illnesses of health management, gene expression, medical imaging, and the Internet of Things may all be characterized by a limited number of characteristics. Feature extraction (FE) is used to modify and simplify data, while feature selection is used to decrease the dataset by eliminating irrelevant characteristics [18]. By capturing a large variance, principal component analysis (PCA) generates new components that contain the most important information of the characteristics [19].

Therefore, the purpose of this investigation is to liken the effectiveness of five supervised ML techniques which include DT, RF, NB, SVM, and KNN for detecting heart disease. The outcomes, which were obtained using python programming language were compared. The system was evaluated using confusion matrix measures like accuracy, precision, f1-score, DR, and FPR.

The remaining part of this article is structured as thus: Sect. 2 presents the literature review with related works discussed extensively. Section 3 discussed the material and methods utilized for the implementation of the study. Section 4 presented the results gotten from the system implementation and the interpretation of the results deduced. The article was concluded in Sect. 5 with future works suggested.

2 Literature Review

Artificial Intelligence (AI), Data Mining (DM), and ML algorithms and methods have started to be used in clinical settings in the past three years, including diagnostic radiography [20–22], cardiac electrophysiology [23], diabetes [24, 25], dermatology [26, 27], and psychoanalysis [28, 29]. In 2019, a surge in ML techniques was anticipated in the medical industry because of their practicality and accessibility as well as the remarkable outcomes achieved so far.

Prakash et al. published research on heart disease prediction in 2017 that used Optimality Criterion feature selection (OCFS) for prognostication and accurately detecting HD. On the assumption of selective information, an investigator advances their technique for choosing a rough feature set (RFS-IE). They evaluate the OCFS with the RFS-IE in terms of computing time, prognostication quality, and error rate utilizing a variety of datasets in their research. When compared to other methods, the OCFS technique takes the least amount of time to execute its process [30].

Seyedamin et al. published research in 2017. They experiment with various machine learning techniques and evaluated the accuracy of their findings. In this research, several machine learning methods were applied to a small data set and the results were compared. A classifier was created using SVM and a medical heart disease dataset. The aforementioned methods of Bagging, Boosting, and Stacking was used to enhance accuracy. MLP outperformed other methods with 84.15% accuracy when the stacking technique SVM was employed [31].

Nguyen Cong Long et al. published a study in 2015 on illness prediction using the firefly algorithm. Rough set theory is used to train the classifier. Other classification methods, such as Naive Bayes and SVM, are compared to the findings. The proposed approach increases accuracy to 87.2% while reducing convergence speed and processing time. The study's limitation is that when there are a significant number of characteristics, the rough set attribute becomes unmanageable [32].

Jesmin Nahar, Tasadduq Imama, and Kevin S. Tickle researched in 2012 that compared various classifiers for identifying heart disease. SVM offered promising accuracy when it comes to improving absolute accuracy as a performance metric. Their article also discussed automated and motivating feature selection techniques such as MFS and CFS. In terms of accuracy, both methods have shown to be extremely promising [33].

Jesmin Nahar et al. utilized an association rule mining (RM) classifier to infer major heart disease factors in 2013. The heart disease dataset was used to conduct a rule extraction experiment utilizing RM techniques such as predictive, Tertius, and apriori. The rule is chosen by Predictive Apriori built on its excessive accuracy [34].

H. Hannah Inbarani et al. proposed an innovative feature selection technique for illness prognostication. Their study was built on hybridization of Particle Swarm Optimization (PSO) and PSO-built Quick-reduct (PSO-QR). The findings of this study indicate that the suggested method outperforms current feature selection (FS) strategies on a variety of typical medical datasets [35].

Researchers [36] proposed five novel FS approaches on the origin of the performance impact of G-BLUP and Bayes C techniques. The authors predicted the body mass index (BMI) and high-density lipoprotein cholesterol (HDL) and it was discovered that guided FS of SNPs in the G-BLUP provided a versatile and computation efficiency alternative to Bayes C. The drawback in their research was that once the supervised selection was employed, predicted performance necessitates a great deal of rigorous assessment, otherwise results may not be obtained [36].

Sina Tabakhi postulated three FS models which are unsubstantiated, filter, and multivariate. The author examined an unproven FS method built on ant colony optimization (ACO). The researcher made a trade-off between the computing period and the rate of the findings. The UFSACO technique shows an increase in efficiency and efficacy, as well as an improvement over earlier similar methods [37].

M. Akhil Jabbar et al. presented a study on the classification of HD by utilizing artificial neural network (ANN) and feature subset selection (FSS). The authors presented FSS as a technique for reducing dimensionality and input data. This article presented a classification technique for HD classification that employed ANN and FS. The number of diagnostic tests required by physicians from patients was decreased when the number of components was reduced. The data set utilized in this study was from Andhra Pradesh, and the findings indicated that accuracy is improved over older categorization methods. Furthermore, the findings indicated that this method is more accurate and quicker [38].

Divia Tomar and Sonali Agarwal released a comprehensive article that downplays the relevance of different data mining methods such as classification, clustering, association, and regression in the field of medicine. They similarly provided an overview of various methods, as well as their benefits and drawbacks. They also drew attention to roadblocks and other issues with data mining methods used on medical data. This article was suggested as a good starting point for learning about the various data mining methods [39].

The main contribution in this study are as follows:

- Improving old manual system.
- Detection of heart disease.
- Introduction of PCA feature extraction.
- Improving efficiency and effectiveness.

3 Material and Method

3.1 Dataset

The NHS England dataset Catalogue is a publicly available online resource that is linked to HD. They're based on real-world hospital administrative data in England. A sample population of emergency admissions for HD is included in the simulated extract. The dataset is available and can be accessed: https://data.england.nhs.uk/dataset?_organization_limit=0&res_format=CSV.

3.2 Methodology

Feature Extraction

For machine learning concerns, feature extraction is a required step. Feature extraction discovers new m dimensions from a set of n original dimensions. This may be divided into two types of methods. They are the projection technique for unsupervised learning, which comprises principal component analysis (PCA), linear discriminate analysis, and others, and the compression approach for supervised learning, which uses mutual information and information theory [40].

PCA

PCA was chosen because it produces excellent results when dealing with linked characteristics. We selected PCA since we are dealing with test characteristics for heart disease diagnosis. It discovers similarities and contrasts between each characteristic by identifying patterns in the data set. It is a strong tool for data analysis. The NHS repository was used to choose the heart disease data collection. The original data is selected, as well as the average of the original data. It's time to calculate the covariance matrix. After that, the covariance matrix is used to choose the Eigenvectors and Eigenvalues. The main component of the heart disease data set is selected as the eigenvector having the greatest Eigenvalue. It demonstrates the strongest link between the data characteristics. The Eigenvalues are ordered from highest to lowest. The data with the highest level of relevance is selected, whereas the data with the lowest level of importance is rejected or deleted. This is done to reduce higher-dimensional data to lower-dimensional data [41–45].

3.3 Proposed System

To identify the appropriate characteristics, a framework is created that is coupled with feature extraction using PCA. This is accomplished via a step-by-step procedure. The outliers must first be eliminated. The observed data that varied significantly from the observed data are referred to as outliers. It's also known as "noise". Data noise or attribute noise are both possible sources of noise. To eliminate outliers, the data is cleaned as the first stage in the machine learning process. The feature extraction phase is the second step, in which the PCA is utilized to extract the key feature or the most significant features. The final stage is to categorize the HD using the five ML classifiers, and then determine whether or not a person has HD. The system flow diagram for the proposed system is shown in Fig. 1.

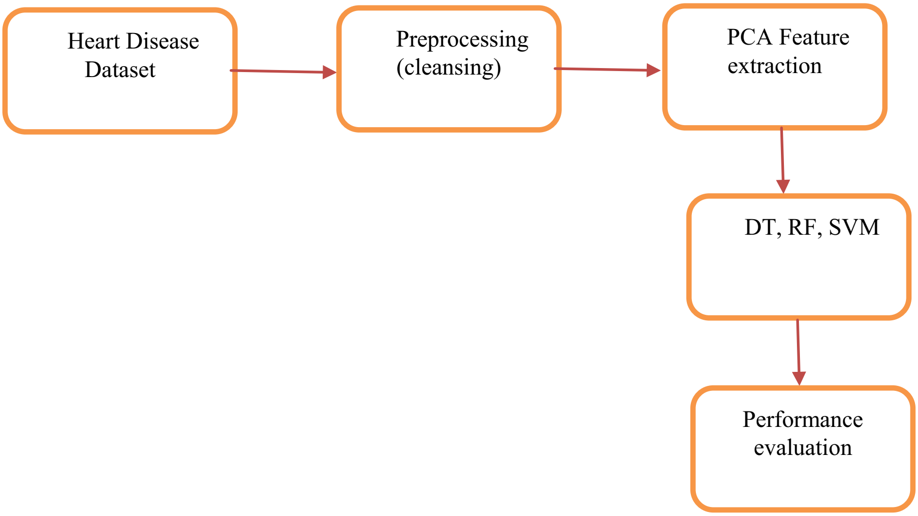

Fig. 1. Proposed System Flow

3.4 Performance Evaluation

The accuracy, sensitivity (detection rate), precision, and f-measure were employed to assess the performance of the study. The true positives (TP) signify an individual has been diagnosed of having HD and is having HD present in his body, the true negatives (TN) signify an individual that was diagnosed as not having HD and was not having HD, the false positives (FP) signify an individual that was diagnosed as having HD but was not having HD in his body and false negatives (FN) signifies an individual that was diagnosed of not having HD and is having HD [46].

4 Results and Discussion

An NHS HD dataset was obtained from the Kaggle repository, an open-source database. Irrelevant rows are removed based on a particular need of HD, which is to reduce the number of characteristics to the fewest that are most relevant for detecting HD, such as age, sex, and so on. The models were implemented using the Python programming language. The PCA is then performed, and the output is verified on the weighted dataset using the five ML classifiers as a subprocess, after which the results are produced. The results produced are evaluated using confusion matrix (CM) values which were used to calculate the performance matrices like accuracy, sensitivity, specificity, f-measure, precision, and false-positive rate. These CM values are shown in Figs. 2, 3, and 4. Table 1 shows the confusion matrix with the actual and predicted values. As it is seen in Table 1 HD binary classification comprises 2 classes: one is a positive class and the other one is the negative class.

Table 1. Confusion matrix

	Predicted HD patient	Predicted healthy patient
Actual HD patient	TP	FN
Actual healthy patient	FP	TN

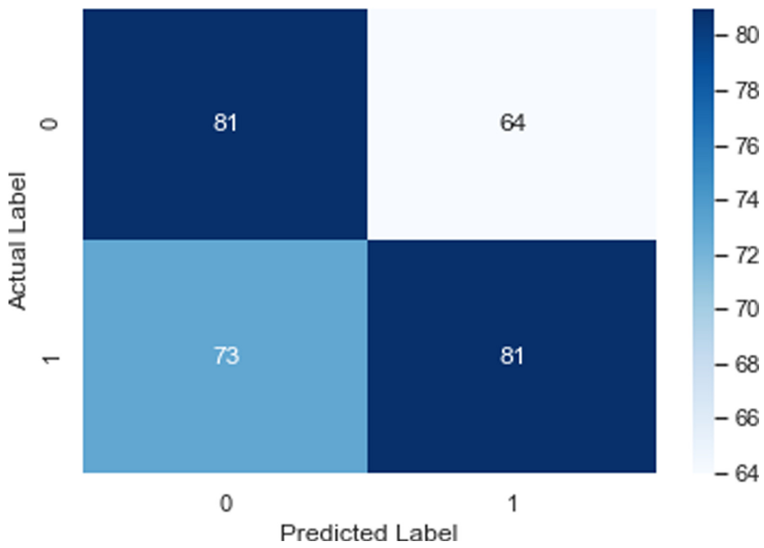

Fig. 2. Confusion matrix for classifier PCA+RF

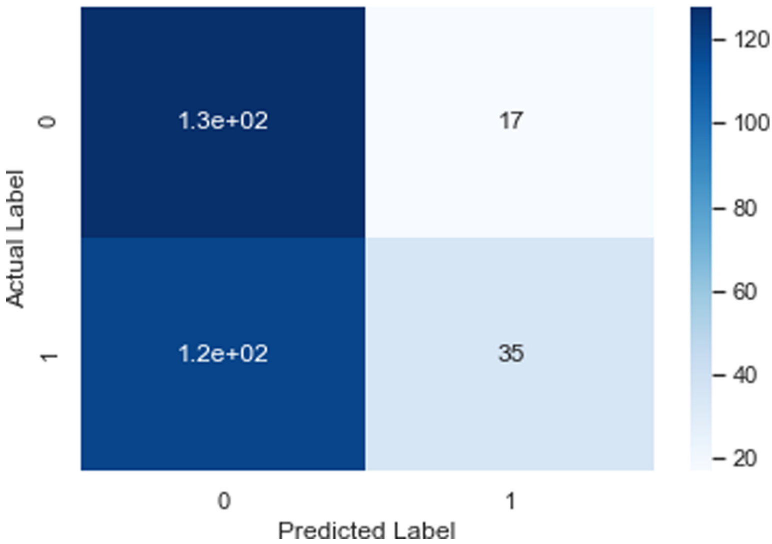

Fig. 3. Confusion matrix for classifier PCA+SVM

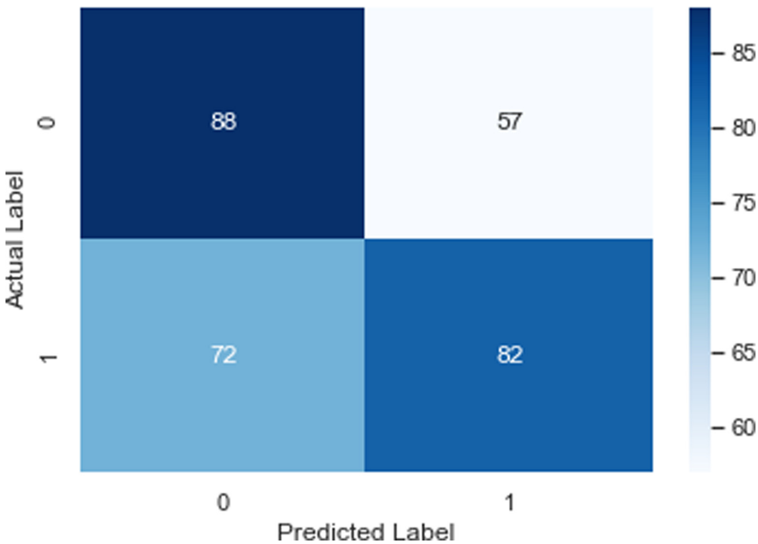

Fig. 4. Confusion matrix for classifier PCA+DF

Table 2. Performance evaluation for the classifiers

Measures	PCA+RF	PCA+DT	PCA+SVM
Sensitivity	55.86	**98.71**	60.69
Precision	52.6	52	**55**
Accuracy	54.18	52.31	**56.86**
F-measure	54.18	**68.12**	57.7

Table 3. Performance evaluation with the existing system

Authors	Methods	Sensitivity
Ayon, Slam, & Hossain [47]	DT	96.23
Shamrat, F. J. M., Raihan, M. A., Rahman, A. S., Mahmud, I., & Akter [48]	DT	98%
Budholiya, K., Shrivastava, S. K., & Sharma, [49]	XGBoost	85.71
Proposed method	**PCA+DT**	**98.71**

4.1 Discussion

The system employed the use of PCA feature extraction and three classification ML techniques. Table 2 shows the performance evaluation for the three classifiers with PCA FE and it was discovered that out of the three classifiers, PCA+SVM outperformed the remaining two in terms of precision and accuracy with 55% and 56.86% respectively, PCA+DT outperformance the remaining two in terms of sensitivity which is the detection rate of 98.71% and f-measure of 68.12% and this shows that the PCA+DT classifiers are the best ML technique for the detection of HD in the medical field as it was demonstrated in this study. Table 3 likewise demonstrate the performance evaluation of the proposed system with existing ones and it was discovered that the proposed system outperformed those of the existing systems with a sensitivity of 98.71% over that of Ayon, Slam & Hossain [47] having 96.23%, Shamrat, Raihan, Rahman, Mahmud & Akter [48] with 98% and Budholiya, K., Shrivastava, S. K., & Sharma, [49] with 85.71%. This shows that it is recommended that the feature extraction technique be encouraged to be used with ML techniques to have high sensitivity, accuracy, and precision values to perform effectively and efficiently.

5 Conclusion and Future Work

The use of PCA to enhance machine learning model detection is suggested in this article. The classifier aimed to determine whether a patient had heart disease. When system resources are taken into account, it is impossible to use all of the features. In this research, we were able to enhance the raw data findings by using the feature extraction method. Different machine learning methods, such as DT, RF, and SVM, are performed separately on an NHS HD dataset using Python, and the results are compared to determine which ML classifier works best and it was deduced that the best ML in this study is PCA+DT with a sensitivity or detection rate of 98.71% and an f-measure of 68.12% which are relatively a good system performance. The true positives rate which is the same as the sensitivity was used to evaluate the accuracy of the classifiers and it was deduced that the PCA+DT outperformed that of the other two with a sensitivity of 98.71% and since the value is on the high side, this implies that the classifier will be able to accurately detect a patient with HD in his or her body.

The use of a limited HD dataset that includes the FE method is a drawback of this research. Because FE performs better with a bigger dataset, the authors suggest using a larger dataset in the future to get greater accuracy, precision, f1-score, DR, and FPR results. Deep learning methods for classification may potentially be used in the study.

References

1. Cardiovascular Diseases (CVDs). http://www.who.int/cardiovascular_diseases/en/. Accessed 16 July 2019
2. American Heart Association. Classes of heart failure. https://www.heart.org/en/health-topics/heart-failure/what-is-heart-failure/classes-of-heart-failure. Accessed 11 Aug 2018
3. Heart failure. http://www.heart.org/HEARTORG/Conditions/HeartFailure/Heart-Failure_UCM_002019_SubHomePage.jsp. Accessed 19 June 2018
4. Shalev-Shwartz, S., Ben-David, S.: Understanding Machine Learning: From Theory to Algorithms. Cambridge University Press, Cambridge (2014)
5. Friedman, J., Hastie, T., Tibshirani, R.: The elements of Statistical Learning. Springer Series in Statistics, vol. 1, no. 10. Springer, New York (2001). https://doi.org/10.1007/978-0-387-84858-7
6. Marsland, S.: Machine Learning: An Algorithmic Perspective. Chapman and Hall/CRC, Boco Raton (2011)
7. Melillo, P., De Luca, N., Bracale, M., Pecchia, L.: Classification tree for risk assessment in patients suffering from congestive heart failure via long-term heart rate variability. IEEE J. Biomed. Health Inform. **17**(3), 727–733 (2013)
8. Al Rahhal, M.M., Bazi, Y., AlHichri, H., Alajlan, N., Melgani, F., Yager, R.R.: Deep learning approach for active classification of electrocardiogram signals. Inf. Sci. **345**, 340–354 (2016)
9. Guidi, G., Pettenati, M.C., Melillo, P., Iadanza, E.: A machine learning system to improve heart failure patient assistance. IEEE J. Biomed. Health Inform. **18**(6), 1750–1756 (2014)
10. Zhang, R., Ma, S., Shanahan, L., Munroe, J., Horn, S., Speedie, S.: Automatic methods to extract New York heart association classification from clinical notes. In: 2017 IEEE international conference on bioinformatics and biomedicine (bibm), pp. 1296–1299. IEEE (2017)
11. Parthiban, G., Srivatsa, S.K.: Applying machine learning methods in diagnosing heart disease for diabetic patients. Int. J. Appl. Inf. Syst. (IJAIS) **3**(7), 25–30 (2012)
12. Domingos, P.: A few useful things to know about machine learning. Commun. ACM **55**(10), 78–87 (2012)
13. Wettschereck, D., Dietterich, T.G.: An experimental comparison of the nearest-neighbor and nearest-hyperrectangle algorithms. Mach. Learn. **19**(1), 5–27 (1995). https://doi.org/10.1007/BF00994658
14. Wettschereck, D., Aha, D.W., Mohri, T.: A review and empirical evaluation of feature weighting methods for a class of lazy learning algorithms. Artif. Intell. Rev. **11**(1), 273–314 (1997). https://doi.org/10.1023/A:1006593614256

15. Yang, M.S., Nataliani, Y.: A feature-reduction fuzzy clustering algorithm based on feature-weighted entropy. IEEE Trans. Fuzzy Syst. **26**(2), 817–835 (2017)
16. Chen, R., Sun, N., Chen, X., Yang, M., Wu, Q.: Supervised feature selection with a stratified feature weighting method. IEEE Access **6**, 15087–15098 (2018)
17. Imani, M., Ghassemian, H.: Feature extraction using weighted training samples. IEEE Geosci. Remote Sens. Lett. **12**(7), 1387–1391 (2015)
18. Liu, H., Motoda, H. (eds.): Feature Extraction, Construction, and Selection: A Data Mining Perspective, vol. 453. Springer, Heidelberg (1998). https://doi.org/10.1007/978-1-4615-5725-8
19. Guyon, I., Gunn, S., Nikravesh, M., Zadeh, L.A. (eds.): Feature Extraction: Foundations and Applications, vol. 207. Springer, Heidelberg (2008)
20. Rajpurkar, P., et al.: ChexNet: radiologist-level pneumonia detection on chest x-rays with deep learning. arXiv preprint arXiv:1711.05225 (2017)
21. Grewal, M., Srivastava, M.M., Kumar, P., Varadarajan, S.: RadNet: radiologist level accuracy using deep learning for hemorrhage detection in CT scans. In: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), pp. 281–284. IEEE (2018)
22. Li, Z., et al.: Thoracic disease identification and localization with limited supervision. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 8290–8299 (2018)
23. Rajpurkar, P., Hannun, A.Y., Haghpanahi, M., Bourn, C., Ng, A.Y.: Cardiologist-level arrhythmia detection with convolutional neural networks. arXiv preprint arXiv:1707.01836 (2017)
24. Ting, D.S.W., et al.: Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. JAMA **318**(22), 2211–2223 (2017)
25. Alade, O., Sowunmi, O., Misra, S., Maskeliūnas, R., Damaševičius, R.: A neural network based expert system for the diagnosis of diabetes mellitus. In: Antipova, T., Rocha, Á. (eds.) Information Technology Science, pp. 14–22. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-74980-8_2
26. Esteva, A., et al.: Dermatologist-level classification of skin cancer with deep neural networks. Nature **542**(7639), 115–118 (2017)
27. Adegun, A.A., Viriri, S., Ogundokun, R.O.: Deep learning approach for medical image analysis. Comput. Intell. Neurosci. **2021**(2021), 6215281 (2021)
28. Al Hanai, T., Ghassemi, M.M., Glass, J.R.: Detecting depression with audio/text sequence modeling of interviews. In: Interspeech, pp. 1716–1720 (2018)
29. Huang, Y.H., Wei, L.H., Chen, Y.S.: Detection of the prodromal phase of bipolar disorder from psychological and phonological aspects in social media. arXiv preprint arXiv:1712.09183 (2017)
30. Long, N.C., Meesad, P., Unger, H.: A highly accurate firefly-based algorithm for heart disease prediction. Expert Syst. Appl. **42**(21), 8221–8231 (2015)
31. Nahar, J., Imam, T., Tickle, K.S., Chen, Y.P.P.: Computational intelligence for heart disease diagnosis: a medical knowledge-driven approach. Expert Syst. Appl. **40**(1), 96–104 (2013)
32. Nahar, J., Imam, T., Tickle, K.S., Chen, Y.P.P.: Association rule mining to detect factors that contribute to heart disease in males and females. Expert Syst. Appl. **40**(4), 1086–1093 (2013)
33. Inbarani, H.H., Azar, A.T., Jothi, G.: Supervised hybrid feature selection based on PSO and rough sets for medical diagnosis. Comput. Methods Programs Biomed. **113**(1), 175–185 (2014)
34. Bermingham, M.L., et al.: Application of high-dimensional feature selection: evaluation for genomic prediction in man. Sci. Rep. **5**(1), 1–12 (2015)
35. Tabakhi, S., Moradi, P., Akhlaghian, F.: An unsupervised feature selection algorithm based on ant colony optimization. Eng. Appl. Artif. Intell. **32**, 112–123 (2014)

36. Jabbar, M.A., Deekshatulu, B.L., Chandra, P.: Classification of heart disease using artificial neural network and feature subset selection. Glob. J. Comput. Sci. Technol. Neural Artif. Intell. **13**(3), 4–8 (2013)
37. Tomar, D., Agarwal, S.: A survey on data mining approaches for healthcare. Int. J. Bio-Sci. Bio-Technol. **5**(5), 241–266 (2013)
38. Hand, D.J.: Principles of data mining. Drug Saf. **30**(7), 621–622 (2007)
39. Roiger, R.J.: Data Mining: A Tutorial-Based Primer. Chapman and Hall/CRC, Boco Raton (2017)
40. Kavitha, R., Kannan, E.: An efficient framework for heart disease classification using feature extraction and feature selection techniques in data mining. In: 2016 International Conference on Emerging Trends in Engineering, Technology, and Science (ICETETS), pp. 1–5. IEEE (2016)
41. www.cs.cmu.edu/afs/cs/academic/class/15385-s12/...slides/lec-18.ppt
42. Tilki, Ö.: PCA-based face recognition: an application, Master's thesis (2014). www.doc.ic.ac.uk/~dfg/ProbabilisticInference/IDAPILecture15.pdf
43. Karegowda, A.G., Manjunath, A.S., Jayaram, M.A.: Comparative study of attribute selection using gain ratio and correlation-based feature selection. Int. J. Inf. Technol. Knowl. Manag. **2**(2), 271–277 (2010)
44. www.isical.ac.in/~k.ramachandra/slides/Feature%20Extraction.pptx
45. http://dai.fmph.uniba.sk/courses/ml/sl/pca.pdf
46. Abdulsalam, S., et al.: Performance evaluation of ANOVA and RFE algorithms for classifying microarray dataset using SVM. In: Themistocleous, M., Papadaki, M., Kamal, M.M. (eds.) EMCIS 2020. LNBIP, vol. 402, pp. 480–492. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-63396-7_32
47. Ayon, S.I., Islam, M.M., Hossain, M.R.: Coronary artery heart disease prediction: a comparative study of computational intelligence techniques. IETE J. Res. 1–20 (2020)
48. Shamrat, F.J.M., Raihan, M.A., Rahman, A.S., Mahmud, I., Akter, R.: An analysis on breast disease prediction using machine learning approaches. Int. J. Sci. Technol. Res. **9**(02), 2450–2455 (2020)
49. Budholiya, K., Shrivastava, S.K., Sharma, V.: An optimized XGBoost based diagnostic system for effective prediction of heart disease. J. King Saud Univ.-Comput. Inf. Sci. (2020)