HEART DISEASE PREDICTION BASED ON PHYSIOLOGICAL PARAMETERS USING ENSEMBLE CLASSIFIER AND PARAMETER OPTIMIZATION

Agung Muliawan[1], Achmad Rizal[2], Sugondo Hadiyoso[3*]
Institut Teknologi dan Sains Mandala, Indonesia[1]
Telkom University, Indonesia[23]
sugondo@telkomuniversity.ac.id

ABSTRACT
This study describes the prediction of heart disease using ensemble classifiers with parameter optimization. As input, a public dataset was taken from UCI machine learning repository, which refers to the dataset at UCI Machine learning. The dataset consists of 13 variables that are considered to influence heart disease. Particle swarm optimization (PSO) was used for feature selection and principal component analysis (PCA) for feature extraction to reduce the features' dimensions. The application of parameter optimization on several machine learning methods such as SVM (Radial Basis Function), Deep learning, and Ensemble Classifier (bagging and boosting) to get the highest accuracy comparison. The results of this study using PSO dimensionality reduction in the public dataset of heart disease resulted in the slightest accuracy compared to PCA. In contrast, the highest accuracy was obtained from optimizing Deep Learning parameters with an accuracy of 84.47% and optimization of SVM RBF parameters with an accuracy of 83.56%. The highest accuracy in the ensemble classifier using bagging on SVM of 83.51%, with a difference of 0.5% from SVM without using bagging.
Keywords: Heart Disease, Machine Learning, Dimensionality Reduction, Parameter Optimization, Ensemble Classifier

1. Introduction

Heart disease is one of the diseases with the highest risk of death. According to data from the World Health Organization (WHO), in 2012 showed, 17.5 million people, or 31% of the world's population, died from heart disease (Al-Mawali, 2015). According to 2018 data, the age-adjusted cardiovascular disease (CVD) death rate in the United States was 217.1 per 100,000. Every 36 seconds, a CVD-related death occurs. In the US, a stroke victim dies every three minutes and 33 seconds. Based on data from 2018, 405 people die from a stroke daily (Virani et al., 2021). However, it is anticipated that advances in the upcoming year will combine with other important risk factors, including hypertension, cholesterol, and diabetes mellitus, to become a significant risk factor (Farkouh et al., 2013). Because of this, efforts are needed to improve the prevention of heart disease, such as consulting a cardiologist in carrying out medical actions and maintaining a healthy body (Winnige et al., 2021). However, in reality, people are reluctant to carry out regular heart health checks due to a lack of public awareness of the importance of health. This is an obstacle to the early detection of heart disease in the community (Nardin et al., 2020).

The rapid development of technology using machine learning (ML) to solve society's problems has become very important, especially those related to medical informatics or biomedical computing and predicting heart disease (Ahmad & Polat, 2023; Nagavelli et al., 2022). Research related to heart disease can generally be divided into two. The first is biological signal processing to detect heart disease, and the second is data mining related to variables that trigger heart disease. Research mainly involves electrocardiogram signals processing for the detection of heart disease (Cheng et al., 2020; Hadiyoso & Rizal, 2017; Pestana et al., 2020) or processing heart sounds (Rizal & Suratman, 2020; Li et al., 2022; Zeinali & Niaki, 2022). Heart imaging techniques through echocardiography are also an alternative for the detection of heart disease (Liastuti et al., 2022; Mabrouk et al., 2016). Another imaging technique for analyzing blood vessels related to heart disease is angiography imaging (Khan Mamun & Elfouly, 2023). The photoplethysmogram (PPG) analysis method also has the potential to detect heart failure (Ave et al., 2015; Fahoum et al., 2023). The second category tends to predict heart disease based on

specific parameters such as complaints, lifestyle, or the results of a person's physical examination (Sharma & Parmar, 2020; Hossain et al., 2023; Bharti et al., 2021). Even the prediction of cardiovascular disease can also be performed by analyzing mental illness (Cunningham et al., 2019). The advantage of this method is that it can predict before heart disease occurs.

Methods for predicting or detecting heart disease by considering the variables of psychological examination, blood, age, and gender have been reported in several studies. However, there is still an opportunity to improve accuracy and reduce the variables used in predictions. Therefore, in this research proposes a feature dimension reduction method using principal component analysis (PCA) and particle swarm optimization (PSO), which will be combined with parameter optimization in machine learning (SVM, deep learning, Ensemble Classifier bagging, and bosting). It is thought that this combination of methods can improve detection accuracy even though feature attributes are reduced. This study aims to find a machine learning model with the highest accuracy by testing the dimensionality reduction algorithm.

2. Research Methods

This study used rapid miner tools to carry out the data mining process. Rapid miner was chosen because it has many data mining modeling and visualization features that are easy to read. In this research, Rapid Miner supports the normalization process, parameter optimization and ensemble classifier. The rapid miner process is divided into two main functions: without using dimensionality reduction in the application of machine learning methods and using dimensionality reduction in the application of machine learning methods. The dimensionality reduction testing process is needed to see that some features that are not required or have the most negligible weight can be reduced (Thrun et al., 2023; Vachharajani & Pandya, 2022). The following is the software and hardware used in this study:
Software
- Operating System Windows 11 64bit BIOS F.70
- Rapid miner 9.1
- Microsoft Office 2013
Hardware
- PC HP 245 G-7
- Processor : AMD Ryzen 5 3500 U up to 2.1GHz
- VGA : Radeon Vega Mobile
- RAM : DDR-3 8GB

The first step is to input the dataset, which is then normalized. Furthermore, whether the data that has been normalized is carried out or not, dimensionality reduction is carried out with PSO or PCA to test the level of accuracy produced. Then each data was tested for accuracy using SVM RBF parameter optimization, deep learning, and ensemble classifier (bagging and boosting). Figure 1 shows the flowchart for the proposed model. While Figure 2 shows the implementation of the proposed model.

2.1 Dataset

The research data use sources from the Kaggle public dataset, the UCI Machine Learning Repository. The data used is an open access dataset, some researchers have used it (Moturi et al., 2020; Sharma & Parmar, 2020). The Cleveland Heart Disease dataset is used in this work which is extracted from the UCI repository. Totally 14 attributes are used in the diagnosis process and 303 data instances. (details are shown in Table 1) with 303 data instances. Dataset processing starts from taking the raw data in UCI Machine Learning with the CSV extension and then converting it to Excel to simplify the import process into Rapid Miner. The "target" attribute in the form of a nominal 0-1 will be converted into "yes" and "no". The data that has been imported (read excel) will be set to the target role and the target role becomes the label.

2.2 Dimensionality Reduction

This study compares two types of dimensional reduction methods, PCA and PSO. The orthogonal basis of the data can be converted into a lower-dimensional subspace using the Principal Components Analysis (PCA) method (Yao et al., 2012). It is possible to reduce the

number of features needed for efficient data representation. By creating a data distribution model in the modified space, this method can reduce the properties of finite variables (Pimentel et al., 2014).

The Particle Swarm Optimization (PSO) algorithm proposed by Kennedy and Eberhart in 1995 was adapted from the foraging behaviour of birds and fish (particles) (Miraswan & Maulidevi, 2016; Zhenyu Meng et al., 2022). All these particles will move in space (optimal) at a certain speed and continue to change each particle in the search space indicated for the search until it reaches the destination (Jamian et al., 2014).
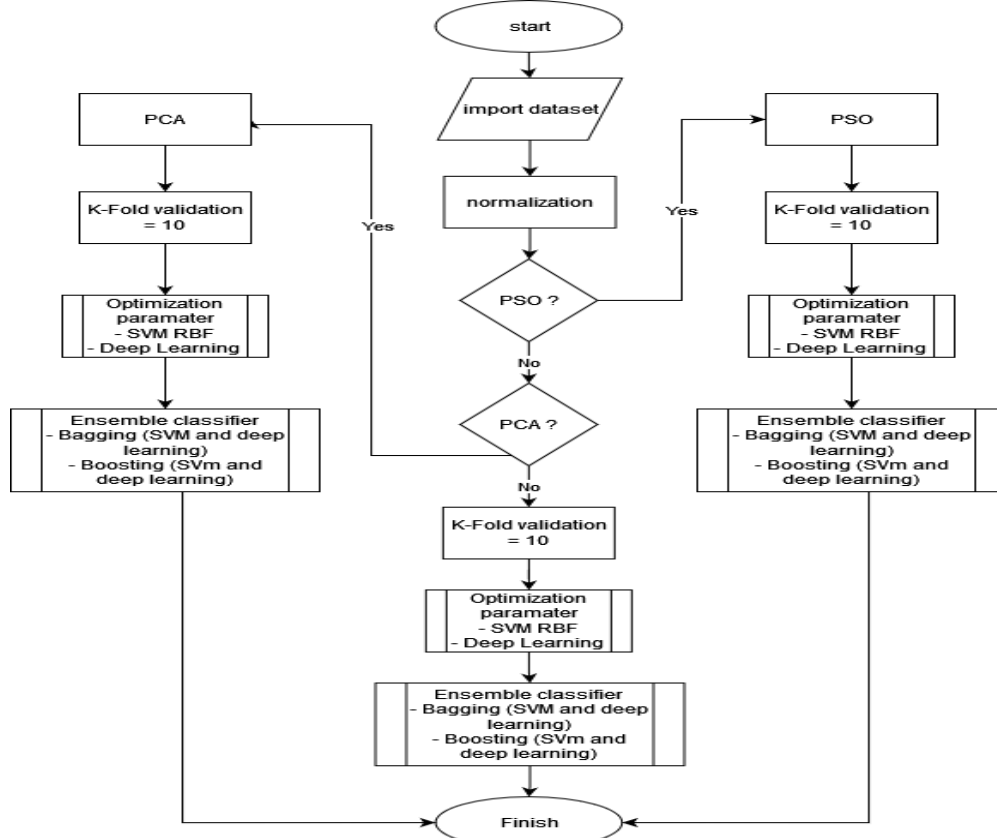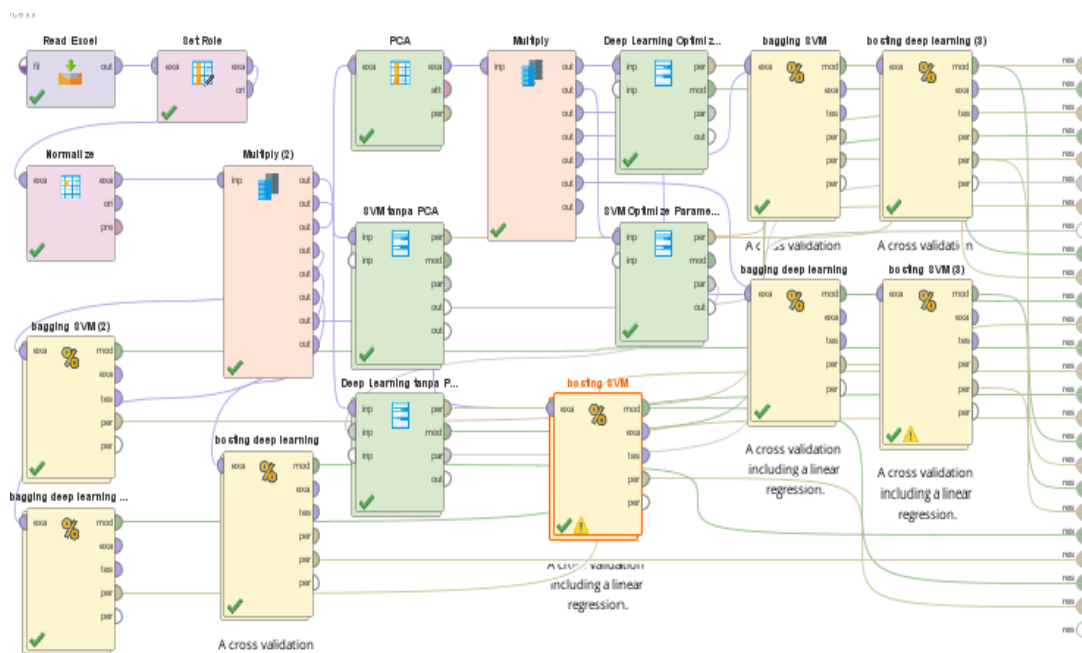

Fig. 1. Flowchart For The Proposed Model

Fig. 2. The Model

Table 1 - Variables And Type Of Data

No	Variable	Type	Values
1.	Age	Integer	29-77 year
2.	Sex	Binominal	0 (female); 1 (male)
3.	Chest paint	Integer	1 (Typical angina); 2 (Atypical angina); 3 (Non angina); 4 (Asymptotic)
4.	Blood pressure	Integer	80-200 mmHg
5.	Cholesterol	Integer	85-603 md/dl
6.	Blood glucose level	Binominal	0 (<=120mg/dl); 1 (>120mg/dl)
7.	Electrocardiography	Integer	0 (normal); 1 (ST-T problem); 2 (Ventricle hypertrophy)
8.	Heart rate	Integer	60-202 bpm
9.	Angina induction	Binominal	0 (No); 1 (Yes)
10.	Old peak	Real	0 - 6.2
11.	Slope	Binominal	1 (up); 2 (flat); 3 (down)
12.	Number of major blood vessel	Integer	0 - 3
13.	Thallium defect	Binominal	3 (Normal); 6 (fixed defect); 7 (reversible defect)

PSO places simple things known as particles in the search space of a certain problem or function. At their present location, the particles subsequently form a fitness function. By taking into account a particular component of the past of the optimal position of one or more swarm members with some random behavior, each particle can decide movement in the search space. When every particle is present, the subsequent iteration starts (Miraswan & Maulidevi, 2016). A three-dimensional vector D that represents the size of the search space and is connected to the following values characterizes each member of the swarm as follows:

1. Actual position, xi
2. Previous best position, pi
3. Particle velocity, vi

The actual position (xi) can be seen as a point in the search space. The present position is regarded as a solution to this problem in each iteration of the algorithm. If the place receives the highest fitness score, the coordinates are stored in the pi vector. The best overall results can be made public in a variable called Pg for comparison with the outcomes of the following iteration. The objective is to remember the best location after recording it (Jamian et al., 2014). In this study, PSO was used to generate three characteristics which mentioned above.

The following are the general steps in the PSO process in Data Mining:

1. Particle Population Initialization: The first step is to initialize the particle population. Each particle represents a potential solution in the search space, which may be a set of model parameters or a set of relevant features
2. Particle Quality Assessment: Each particle is evaluated for its quality in achieving optimization goals.
3. Determination of the Best Particle (Pbest): Each particle stores information about its own best position (Pbest) based on its quality assessment
4. Determination of the Global Particle Best (Gbest): In addition to storing Pbest, PSO also searches for the best solution found by all the particles in the population.
5. Particle Position Update: Each particle updates its position based on personal experience (Pbest) and collective experience (Gbest). This process describes how particles move in search of a better solution.
6. Evaluation and Criteria: PSO iteration continues with particle position updates. Evaluation continues, and stopping criteria are applied. The stopping criteria can be the number of

iterations that have been performed, the accuracy reaching a certain threshold, or an insignificant improvement in quality.
7. Optimal Solution: After the iteration is complete, the solution represented by the best Gbest or Pbest is considered the optimal or best solution found by the PSO algorithm.

2.3 Classifiers
2.3.1 Support Vector Machine
SVM is a learning system used in the hypothesis of linear functions in high-dimensional feature space; the computer will be trained with an algorithm based on optimization theory with statistical learning theory (Srivastava & Bhambhu, 2010). SVM can work on non-linear data by using a kernel approach to the initial features of the data set (Awad & Khanna, 2015). Kernel functions map lower dimensions to higher dimensions (Abbaszadeh et al., 2019). In this study, the RBF kernel concept or Radial Basic Function is used in the classification process to get better accuracy with the formula:

$$K = \left(x_i, x_j\right) = exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right)$$

x_i and x_j are pairs of two training data. $\sigma^2 > 0$ is a constant. The kernel function must use dot product substitution in the feature space, which is very dependent on this kernel function in determining the new features produced.

2.3.2 Deep Learning
Deep learning is an artificial neural network algorithm that uses data as input and processes it with a hidden layer. Furthermore, a non-linear transformation of the input data is carried out to calculate the output value (Li et al., 2019).
This study uses parameter optimization (learning rate) with a min value of 0.01, max 1.0, steps 10, and uses GPU utilization in processing so that the data computing process can take place faster than using CPU and RAM alone.

2.3.3 Ensemble Classifier
Ensemble classifier is a data-level approach aimed at improving class balance (Liu et al., 2022). The strategy using the ensemble algorithm aims to enhance the algorithm without altering the data. The data level approach and the algorithm level approach are two possible trajectories (Murugananthan & Durairaj, 2019). Boosting and bagging are two common ensemble algorithms (Jafarzadeh et al., 2021). An algorithm with better classification performance is AdaBoost. Bagging is a straightforward but efficient ensemble method that has been utilized in numerous real-world applications to improve the accuracy of classification algorithms. Following are the general steps in how an ensemble classifier works:
1. Creating Base Learner Models: The first step is to build a number of different base models (base learners). These models can come from various machine learning algorithms such as Decision Trees, Random Forests, Logistic Regression, Support Vector Machines, Neural Networks, and others.
2. Base Learner Model Training: Each base model is trained on the same training data. These models generate individual predictions based on the characteristics of the data and the rules they learn during training
3. Prediction Combination: The results of the base models are combined to produce an ensemble prediction. Combining methods can vary, but the most common methods are through majority voting (bagging), adding weights (boosting), or selecting predictions from the model that is considered the best (stacking).
4. Evaluation and Accuracy: Ensemble predictions are evaluated on different test data to measure their performance.
The working principle of the ensemble classifier is based on the assumption that various basic models have different strengths and weaknesses. By combining predictions from these various models, the weaknesses of one model can be offset by the strengths of the other models, overall resulting in more accurate and stable predictions (Rokach, 2010).

3. Results and Discussions

The data was first reduced in dimensions using PSO or PCA in this study. In the next stage, classification is carried out using SVM and DL. Then the SVM and DL were optimized using bagging and boosting methods. These results are then compared with the accuracy without the dimension reduction process.

Table 2 - Accuracy (%) using PSO

Algorithm		Ensemble Classifier			
		Bagging		Boosting	
SVM (RBF)	DL	SVM	DL	SVM	DL
76.32	76.54	76.25	67.66	72.61	74.22

Table 3 - Accuracy (%) using PCA

Algorithm		Ensemble Classifier			
		Bagging		Boosting	
SVM (RBF)	DL	SVM	DL	SVM	DL
84.20	84.18	83.22	82.54	82.76	83.13

From the table above, we can see that the use of feature extraction can increase the classification accuracy quite significantly. The highest accuracy results were obtained using SVM parameter optimization (RBF) with 84.20%. For the ensemble classifier, it can be seen that the accuracy difference is quite far from PSO to PCA, which has increased quite a lot. Next, we will classify without using dimensional reduction or all of the existing input variables listed in Table 4.

Table 4 - Accuracy (%) Without Dimension Reduction

ML & Parameters optimization		Ensemble Classifier			
		Bagging		Boosting	
SVM (RBF)	DL	SVM	DL	SVM	DL
83.56	84.47	83.51	82.51	83.17	81.22

Table 5 - Optimization parameters

Algorithm	Parameter	Value
SVM (RBF)	C and Gamma	Min: 0.01
		Max: 1
		Steps: 10
		Scale: linear
Deep learning	Gamma	Min: 0.01
		Max: 1
		Steps: 10
		Scale: linear

Suppose, viewed more specifically, using a small dataset (13 input variables and 303 instances) results in significant accuracy. If the dataset has more input variables and more instances, the accuracy can be higher. The exciting thing is that the highest accuracy results obtained from deep learning parameter optimization of 84.47% with a learning rate of 0.1 can beat the accuracy of SVM parameter optimization (RBF).

From the results obtained, the reduction of the dimension of the feature does not necessarily increase the accuracy of the classification. For feature reduction using PSO, the accuracy is much lower than PCA. This is because the optimization, in general, does not always produce the maximum value but only improves performance. Meanwhile, classification without a dimension reduction process for some cases increases accuracy; this is understandable because the initial characteristics will provide complete information on the data. Meanwhile, feature reduction may eliminate important information for classification.

The classifier ensemble process generally does not produce higher accuracy than the parameter optimization process. This is because the learning algorithm is used without considering the value of the kernels in the algorithm. This causes a decrease in the accuracy value than using parameter optimization. In the parameter optimization process, trials are conducted to find the kernel value which has the smallest Root Mean Square Error (RMSE) value to be used as a reference for the use of kernel values such as SVM, which has the disadvantage of being difficult to determine the optimal parameter optimization value so that parameter optimization

tests are carried out on the SVM kernel (dot, RBF, polynomial). Jung et al. (Hojin Nam, Minseo Rhee, Jeung-Sun Lee, 2021) used the same dataset as this study.

Another study on feature analysis in chronic heart disease using the same dataset in this study was conducted by Bialy et al. (El-Bialy et al., 2015). The difference in this research is using the C4.5 Tree algorithm by grouping the process to be compared with other data, then the data with the highest similarity will be grouped into one part. The results of this study have a classification accuracy of 78.06%, which is greater than the average classification accuracy using a separate dataset of 75.48%. The results of this study were compared with (El-Bialy et al., 2015) using the same dataset. It was found that the value of increasing accuracy by optimizing parameters with the ensemble classifier SVM (Baging and bosting) and Deep Learning (83.34% and 84.47%) was compared. without using parameter optimization (76.30% and 75.48). Optimizing ensemble classifier parameters has the potential to be explored in further research. From the results of the performance test of the proposed method, it can be seen that the parameter optimization using DL increases the accuracy, although it is not significant. Meanwhile, dimension reduction in features and parameter optimization cannot significantly increase accuracy. The highest accuracy is achieved using DL with parameter optimization without feature reduction. An ensemble classifier cannot produce higher accuracy because its parameters are not optimized. Exploring opportunities are still open to achieve higher accuracy with better parameter optimization in the proposed method.

4. Conclusion

In this study, a combination of dimensional reduction and optimization of classifier methods is proposed to predict heart disease. PCA and PSO are used in dimension reduction, while the classifiers used are SVM (Radial Basis Function), Deep learning, and Ensemble Classifier (bagging and boosting). The test results show that the highest level of accuracy is found in deep learning parameter optimization with an accuracy value of 84.47% with a difference of 0.29% on PCA. The ensemble classifier is still below the highest accuracy, with an accuracy value of 83.17% on boosting SVM. This occurs because the use of ensemble classifiers (baging and bosting) is suitable for certain data characteristics such as data that is large enough to produce a significant increase in better accuracy.

The implications of research on clinical health applications can support doctors in carrying out early detection using parameters that have been observed, It is hoped that this research can be a reference for further study in predicting heart disease so that it can detect the risk of heart disease early. Suggestions from the further analysis can be made by increasing the number of input variables and instances so that accuracy can increase. The use of other algorithms can also be done in further research, such as KNN, Naïve Bayes, Decision Tree, etc. The use of parameter optimization is also needed in testing the accuracy of each machine learning method.

References

Abbaszadeh, B., Haddad, T., & Yagoub, M. C. E. (2019). Probabilistic prediction of Epileptic Seizures using SVM. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, 7, 3442–3445. https://doi.org/10.1109/EMBC.2019.8856286

Ahmad, A. A., & Polat, H. (2023). Prediction of Heart Disease Based on Machine Learning Using Jellyfish Optimization Algorithm. Diagnostics, 13(14), 2392. https://doi.org/10.3390/diagnostics13142392

Al-Mawali, A. (2015). Non-communicable diseases: Shining a light on cardiovascular disease, Oman's biggest killer. Oman Medical Journal, 30(4), 227–228. https://doi.org/10.5001/omj.2015.47

Ave, A., Fauzan, H., S, R. A., & Zakaria, H. (2015). Early Detection of Cardiovascular Disease with Photoplethysmogram (PPG) Sensor. *The 5th International Conference on Electrical Engineering and Informatics 2015*, 676–681.

Awad, M., & Khanna, R. (2015). Support Vector Machines for Classification. In Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers (Nomor January, hal. 1–248). https://doi.org/10.1007/978-1-4302-5990-9

Bharti, R., Khamparia, A., Shabaz, M., Dhiman, G., Pande, S., & Singh, P. (2021). Prediction of Heart Disease Using a Combination of Machine Learning and Deep Learning. Computational Intelligence and Neuroscience, 2021, 1–11. https://doi.org/10.1155/2021/8387680

Cheng, Y., Hu, Y., Hou, M., Pan, T., He, W., & Ye, Y. (2020). Atrial fibrillation detection directly from compressed ECG with the prior of measurement matrix. *Information (Switzerland)*, *11*(9), 1–15. https://doi.org/10.3390/INFO11090436

Cunningham, R., Poppe, K., Peterson, D., Every-Palmer, S., Soosay, I., & Jackson, R. (2019). Prediction of cardiovascular disease risk among people with severe mental illness: A cohort study. *PLoS ONE*, *14*(9), 1–13. https://doi.org/10.1371/journal.pone.0221521

El-Bialy, R., Salamay, M. A., Karam, O. H., & Khalifa, M. E. (2015). Feature Analysis of Coronary Artery Heart Disease Data Sets. *Procedia Computer Science*, *65*(Iccmit), 459–468. https://doi.org/10.1016/j.procs.2015.09.132

Fahoum, A. S. Al, Al-haija, A. O. A., & Alshraideh, H. A. (2023). Identification of Coronary Artery Diseases Using Photoplethysmography Signals and Practical Feature Selection Process. *Bioengineering*, *10*, 1–13.

Farkouh, M. E., Boden, W. E., Bittner, V., Muratov, V., Hartigan, P., Ogdie, M., Bertolet, M., Mathewkutty, S., Teo, K., Maron, D. J., Sethi, S. S., Domanski, M., Frye, R. L., & Fuster, V. (2013). Risk factor control for coronary artery disease secondary prevention in large randomized trials. *Journal of the American College of Cardiology*, *61*(15), 1607–1615. https://doi.org/10.1016/j.jacc.2013.01.044

Hadiyoso, S., & Rizal, A. (2017). Electrocardiogram signal classification using higher-order complexity of hjorth descriptor. *Advanced Science Letters*, *23*(5), 3972–3974. https://doi.org/10.1166/asl.2017.8251

Hojin Nam, Minseo Rhee, Jeung-Sun Lee, Y.-G. J. B. K. (2021). Effective Diagnosis of Coronary Artery Disease using Case-based Reasoning. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, *12*(5), 449–457. https://doi.org/10.17762/turcomat.v12i5.991

Hossain, M. I., Maruf, M. H., Khan, M. A. R., Prity, F. S., Fatema, S., Ejaz, M. S., & Khan, M. A. S. (2023). Heart disease prediction using distinct artificial intelligence techniques: performance analysis and comparison. Iran Journal of Computer Science, 6(4), 397–417. https://doi.org/10.1007/s42044-023-00148-7

Jafarzadeh, H., Mahdianpari, M., Gill, E., Mohammadimanesh, F., & Homayouni, S. (2021). Bagging and Boosting Ensemble Classifiers for Classification of Multispectral, Hyperspectral and PolSAR Data: A Comparative Evaluation. Remote Sensing, 13(21), 4405. https://doi.org/10.3390/rs13214405

Jamian, J. J., Abdullah, M. N., Mokhlis, H., Mustafa, M. W., & Bakar, A. H. A. (2014). Global particle swarm optimization for high dimension numerical functions analysis. *Journal of Applied Mathematics*, *2014*, 1–15. https://doi.org/10.1155/2014/329193

Khan Mamun, M. M. R., & Elfouly, T. (2023). Detection of Cardiovascular Disease from Clinical Parameters Using a One-Dimensional Convolutional Neural Network. Bioengineering, 10(7), 796. https://doi.org/10.3390/bioengineering10070796

Li, F., Zhang, Z., Wang, L., & Liu, W. (2022). Heart sound classification based on improved mel-frequency spectral coefficients and deep residual learning. Frontiers in Physiology, 13, 1–16. https://doi.org/10.3389/fphys.2022.1084420

Li, Y., Shi, D., & Bu, F. (2019). Automatic recognition of rock images based on convolutional neural network and discrete cosine transform. *Traitement Du Signal*, *36*(5), 463–469. https://doi.org/10.18280/ts.360512

Liastuti, L. D., Siswanto, B. B., Sukmawan, R., Jatmiko, W., Nursakina, Y., Yusticia, R., Putri, I., Jati, G., & Nur, A. A. (2022). Detecting Left Heart Failure in Echocardiography through Machine Learning : A Systematic Review. *Rev. Cardiovasc. Med.*, *23*(12), 1–10.

Liu, L., Wu, X., Li, S., Li, Y., Tan, S., & Bai, Y. (2022). Solving the class imbalance problem using ensemble algorithm: application of screening for aortic dissection. BMC Medical Informatics and Decision Making, 22(1), 82. https://doi.org/10.1186/s12911-022-01821-w

Mabrouk, A., Omar, S., Bansal, M., & Sengupta, P. P. (2016). Advances in Echocardiographic

Imaging in Heart Failure With Reduced and Preserved Ejection Fraction. *Circulation Research*, *July*, 357–374. https://doi.org/10.1161/CIRCRESAHA.116.309128

Miraswan, K. J., & Maulidevi, N. U. (2016). Particle swarm optimization and fuzzy logic control in gas leakage detector mobile robot. *Proceedings of the 2015 International Conference on Automation, Cognitive Science, Optics, Micro Electro-Mechanical System, and Information Technology, ICACOMIT 2015*, 150–155. https://doi.org/10.1109/ICACOMIT.2015.7440195

Moturi, S., Vemuru, S., & Rao, S. N. T. (2020). Classification Model for Prediction of Heart Disease using Correlation Coefficient Technique. *International Journal of Advanced Trends in Computer Science and Engineering*, *9*(2), 2117–2123. https://doi.org/10.1007/978-981-10-5699-4_67

Murugananthan, V., & Durairaj, U. K. (2019). RUS boost tree ensemble classifiers for occupancy detection. *International Journal of Recent Technology and Engineering*, *8*(2 Special Issue 2), 272–277. https://doi.org/10.35940/ijrte.B1048.0782S219

Nardin, M., Verdoia, M., Negro, F., Rolla, R., Tonon, F., & De Luca, G. (2020). Impact of active smoking on the immature platelet fraction and its relationship with the extent of coronary artery disease. *European Journal of Clinical Investigation*, *50*(2), 1–10. https://doi.org/10.1111/eci.13181

Nagavelli, U., Samanta, D., & Chakraborty, P. (2022). Machine Learning Technology-Based Heart Disease Detection Models. Journal of Healthcare Engineering, 2022, 1–9. https://doi.org/10.1155/2022/7351061

Pestana, J., Belo, D., & Gamboa, H. (2020). Detection of Abnormalities in Electrocardiogram (ECG) using Deep Learning. *Proceedings of the 13th International Joint Conference on Biomedical Engineering Systems and Technologies*, 236–243. https://doi.org/10.5220/0008967302360243

Pimentel, M. A. F., Clifton, D. A., Clifton, L., & Tarassenko, L. (2014). A review of novelty detection. *Signal Processing*, *99*, 215–249. https://doi.org/10.1016/j.sigpro.2013.12.026

Rizal, A., & Suratman, F. Y. (2020). Classification of Normal and Murmur Hearts Sound using the Fractal Method. *International Journal of Advanced Trends in Computer Science and Engineering*, *9*(5), 8178–8183. https://doi.org/10.30534/ijatcse/2020/181952020

Rokach, L. (2010). Ensemble-based classifiers. Artificial Intelligence Review, 33(1–2), 1–39. https://doi.org/10.1007/s10462-009-9124-7

Sharma, S., & Parmar, M. (2020). Heart Diseases Prediction using Deep Learning Neural Network Model. *International Journal of Innovative Technology and Exploring Engineering*, *9*(3), 2244–2248. https://doi.org/10.35940/ijitee.c9009.019320

Srivastava, D. K., & Bhambhu, L. (2010). Data classification using support vector machine. *Journal of Theoretical and Applied Information Technology*, *12*(1), 1–7.

Thrun, M. C., Märte, J., & Stier, Q. (2023). Analyzing Quality Measurements for Dimensionality Reduction. Machine Learning and Knowledge Extraction, 5(3), 1076–1118. https://doi.org/10.3390/make5030056

Vachharajani, B., & Pandya, D. (2022). Dimension reduction techniques: Current status and perspectives. Materials Today: Proceedings, 62, 7024–7027. https://doi.org/10.1016/j.matpr.2021.12.549

Virani, S. S., Alonso, A., Aparicio, H. J., Benjamin, E. J., Bittencourt, M. S., Callaway, C. W., Carson, A. P., Chamberlain, A. M., Cheng, S., Delling, F. N., Elkind, M. S. V., Evenson, K. R., Ferguson, J. F., Gupta, D. K., Khan, S. S., Kissela, B. M., Knutson, K. L., Lee, C. D., Lewis, T. T., … Tsao, C. W. (2021). Heart Disease and Stroke Statistics - 2021 Update: A Report From the American Heart Association. *Circulation*, *143*(8), E254–E743. https://doi.org/10.1161/CIR.0000000000000950

Winnige, P., Vysoky, R., Dosbaba, F., & Batalik, L. (2021). Cardiac rehabilitation and its essential role in the secondary prevention of cardiovascular diseases. World Journal of Clinical Cases, 9(8), 1761–1784. https://doi.org/10.12998/wjcc.v9.i8.1761

Yao, F., Coquery, J., & Lê Cao, K.-A. (2012). Independent Principal Component Analysis for biologically meaningful dimension reduction of large biological data sets. BMC Bioinformatics, 13(1), 24. https://doi.org/10.1186/1471-2105-13-24

Zeinali, Y., & Niaki, S. T. A. (2022). Heart sound classification using signal processing and machine learning algorithms. Machine Learning with Applications, 7, 100206. https://doi.org/10.1016/j.mlwa.2021.100206

Zhenyu Meng, Yuxin Zhong, Guojun Mao, & Yan Liang. (2022). PSO-sono: A novel PSO variant for single-objective numerical optimization. *Information Sciences*, *586*, 176–191.