



Paper Type: Original Article

A Combination of EDA, Machine Learning, and Artificial Neural Networks for Accurate Prediction of Heart Disease

Ali Rashedi Gazari^{1*}, Shayan Rokhva¹ , Toktam Khatibi¹ , Elham Akhondzade Noughabi¹ , Babak Teimourpour¹ 

¹ Department of Information Technology Engineering, Tarbiat Modares University, Tehran, Iran; alirashedigazari@gmail.com; shayanrokhva1999@gmail.com; toktam.khatibi@modares.ac.ir; elham.akhondzade@modares.ac.ir; b.teimourpour@modares.ac.ir.

Citation:

Received: 25 September 2024

Revised: 02 November 2024

Accepted: 10 March 2025

Rashedi Gazari, A., Rokhva, Sh., Khatibi, T., Akhondzade Noughabi, E., & Teimourpour, B. (2025). A combination of EDA, machine learning, and artificial neural networks for the accurate prediction of heart disease. *Annals of Healthcare Systems Engineering*, 2(1), 38-46.


Abstract


Cardiovascular diseases cause millions of deaths each year, with cases continuing to rise, making early prediction increasingly important. Although data science and Artificial Intelligence (AI) have been utilized to address this issue, further studies that enhance predictability and generalization are crucial, as they significantly reduce mortality rates and healthcare costs. This study employs Exploratory Data Analysis (EDA), a variety of conventional Machine Learning (ML) algorithms, and an Artificial Neural Network (ANN) to predict heart disease accurately and fill research gaps. A dataset from Kaggle, containing 1025 training samples and 303 test samples, with 14 attributes, including 13 predictive variables and a binary target indicating heart disease presence, was used. Normalization, feature importance analysis, K-fold cross-validation, and grid search were meticulously applied to improve model performance, generalization, and robustness. These methodologies led to impressive results, with most models achieving 100% accuracy, precision, recall, and F1-score on the test data, without signs of overfitting, data leakage, or bias. Principal Component Analysis (PCA) was also conducted to evaluate the richness of the features and their potential for dimension reduction. Lastly, in-depth discussions were made to clarify the study's outcomes, compare results with the most related studies, and comprehensively examine real-world applicability.

Keywords: Heart disease, Exploratory data analysis, Machine learning, Artificial intelligence, Healthcare.

1 | Introduction

Heart disease remains a leading cause of death worldwide, with its incidence rising each year [1]. Early detection is vital for enhancing healthcare outcomes, making predictive analytics pivotal for timely disease prevention, effective treatment, and significant cost reduction [2], [3], [4]. Machine Learning (ML), a branch

 Corresponding Author: alirashedigazari@gmail.com

 <https://doi.org/10.22105/ahse.v2i1.29>



Licensee System Analytics. This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0>).

of Artificial Intelligence (AI), is revolutionizing medical diagnostics by offering robust, accurate predictions that considerably guide clinical decisions. ML models can leverage historical health data to solve complex classification problems, such as predicting heart disease, and uncover complex correlations and hidden patterns often overlooked by human analysts and traditional statistical methods [5]. Apart from the high performance, ML algorithms also enable machines to extract features, learn from data, and therefore, analyze and predict outcomes automatically, leading to automation, increased productivity, and reduced time for real-world scenarios [6], [7], [8].

Deep Learning (DL), a subset of ML, involves neural networks with multiple layers and deep structures, making it particularly effective for large datasets and complex problems where traditional ML algorithms may reach performance limits [9]-[11]. The performance of DL models generally tends to increase with a substantial amount of data and deeper structures [11], [12]. Notably, both ML and DL approaches can be employed for prediction depending on the application. However, models that demonstrate superior performance, generalization, and robustness on test data, representing unseen scenarios, should be prioritized for real-world applicability [12], [13].

To identify the scientific gap, a series of recent studies were reviewed [2], [9], [14]-[18]. While leveraging both conventional ML and state-of-the-art DL models is a common practice for predicting heart disease, which is widely discussed in the literature, the primary challenge remains in achieving highly accurate and robust predictions while leveraging different metrics. Prediction of heart disease is the gap that this paper aims to fill.

Given the proven effectiveness of both conventional ML and state-of-the-art DL models in various sectors [4], [8], [11], [19], particularly healthcare [2], [15], [16], this study aims to leverage data comprehension, Exploratory Data Analysis (EDA), diverse conventional ML algorithms, and an Artificial Neural Network (ANN) for accurate and robust prediction of heart disease. To be more precise, the main contributions of this study are as follows:

- I. Utilizing a heart disease dataset from Kaggle, comprising 1025 training instances and 303 test instances, with 13 predictive features and one binary target indicating heart disease.
- II. Performing a comprehensive analysis of the most highly correlated features linked to the disease, and also examining their distribution and visualizing them through a correlation matrix.
- III. Leveraging a variety of conventional ML models and an ANN for the best possible prediction, also boosting their performance and generalization through normalization, grid searching, and k-fold cross-validation, in turn, surpassing numerous previous studies.
- IV. Comparing the performance, potential strengths, and weaknesses of the employed methods, based on various metrics such as accuracy, precision, recall, and F1 score, while reporting the utilized hyperparameters.
- V. Applying Principal Component Analysis (PCA) to evaluate the richness of the dataset's features and effectively summarizing them in fewer dimensions while computing information loss.
- VI. An accurate and thorough discussion regarding the utilized criteria, employed models, study outcomes, and how they positively contribute to the health sector.

2 | Methodology

2.1 | Data Comprehension and Exploratory Data Analysis

To achieve accurate and comprehensive results, it is essential to have a thorough understanding of the dataset being analyzed [8], [20]. This section examines the dataset, focusing on its features, distributions, and variable correlations. The insights gained will enhance model implementation, clarify the relationship between predictive variables and heart disease, and improve the interpretation of PCA results. The dataset includes 1,025 training samples and 303 test samples, which were intentionally separated to avoid data leakage and bias

[17]. Notably, no missing value was identified. Since understanding each feature is critical for contextualizing the research, *Table 1* provides definitions and types of each variable.

For improved comprehension, *Fig. 1* displays the correlation matrix as a heatmap, illustrating the relationships between every pair of variables. This analysis is crucial, as frequently high correlations among predictive variables can indicate dependency and information redundancy, potentially introducing noise and exacerbating the curse of dimensionality. Additionally, if the correlation of any predictive variable with the target is either extremely low or high, these features may pose a challenge since variables with minimal impact or excessive correlation can act as noise or merely replicate the target, leading to unnecessary dimensionality and compromising the model's robustness and performance [21], [22]. Fortunately, the analysis did not reveal significant concerns, allowing for the inclusion of all features in the modeling process.

According to *Fig. 1*, Chest Pain type (CP) and maximum heart rate achieved by patients (Thalach) exhibit the strongest positive correlations, indicating that as these parameters increase, so does the risk of heart disease. Conversely, exercise-induced angina (Exang) and ST depression induced by exercise (Oldpeak) revealed the most converse correlations, reflecting an inverse relationship with the target. In this context, *Fig. 2* provides a more detailed visualization of these, confirming that the four aforementioned attributes are noticeably different between those who do and those who do not suffer from heart diseases, resulting in some precious insights. For instance, although in the healthcare sector, where the stakes are high, analyzing each feature is paramount, these insights can offer valid intuitions about influential attributes [14], [23], assisting decision-makers when time or resources are limited.

Notably, all figures in Section 2, related to EDA, are based on the complete dataset, showing the distribution of train and test data combined. In other words, although the data were initially separated into training and testing sets, they were integrated for exploratory analysis. Afterward, the data were re-separated into their original formats to ensure that model creation, training, and implementation were conducted with distinct datasets, preventing data leakage and bias, thereby enhancing the validity and generalizability of our findings [12].

Table 1. Definition and status of the features in the dataset.

Attribute	Description	Variable Type
Age	Represents the age of an individual	Continuous
Sex	Indicates the gender of an individual	Categorical
CP	Categorize the type of CP experienced	Categorical
Trestbps	Measures the resting blood pressure	Continuous
Chol	Indicates the serum cholesterol level	Continuous
FBS	Shows whether fasting blood sugar is above 120 or not	Categorical
Restecg	Displays the results of the resting electrocardiogram	Categorical
Thalach	Records the maximum heart rate achieved by patients	Continuous
Exang	Whether the patient experienced angina (CP) during exercise	Categorical
Oldpeak	Measures ST depression induced by exercise relative to rest.	Continuous
Slope	Describes the slope of the peak exercise ST segment.	Categorical
Ca	Counts the number of major vessels colored by fluoroscopy	Categorical
Thal	Refers to thalassemia, a blood disorder	Categorical
Target	Whether one has heart disease or not	Categorical

2.2 | Implementation of Techniques and Model Building

After re-separating train and test to minimize bias, both datasets were normalized using the mean and standard deviation from only the training data, crucial for enhanced performance of some ML algorithms like K-Nearest Neighbors (KNN) and Logistic Regression, also ensuring that models remain unaware of test data distribution, enhancing robustness and validity [20]. Various conventional ML algorithms and an ANN were

then implemented. Model performance and generalization were fine-tuned through grid searching and 5-fold cross-validation, ensuring optimized model configurations and robust outcomes.

2.3 | Principal Component Analysis

PCA, an unsupervised learning technique, was performed to demonstrate how effectively the data dimensions can be reduced while retaining significant information. We progressively reduced the 13 predictive variables to 12, 11, 10, and so forth, down to 2 features while calculating the information loss at each step. This analysis will illustrate the extent to which the data can be compressed into lower dimensions while being informative. This analysis primarily provides insights into the importance of dimensionality reduction, especially in high-dimensional data scenarios.

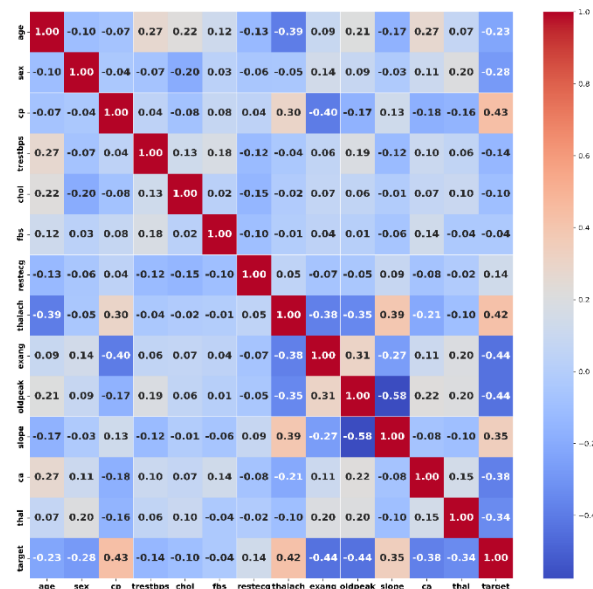


Fig. 1. Correlation matrix with a heatmap.

3 | Results and Analysis

3.1 | Model Implementation for Accurate Classification

After applying and optimizing models, they were evaluated on the test unseen dataset using four metrics: Accuracy, precision, recall, and F1 score. The parameters yielding the reported performances, along with their results, are summarized in Table 2. Most models demonstrated nearly 100% accuracy, highlighting the effectiveness of the employed models on the current dataset.

Table 2. Applied Models, Parameters Used, and Performance on The Test Data.

ML Model	Employed Hyper-Parameters	Accuracy	Precision	Recall	F1 Score
Decision Tree (DT) classifier	max_depth:9 ; min_sample_leaf:3 ; min_sample_split:6	100%	100%	100%	100%
Support Vector Classifier (SVC)	C:10; gamma=0.1; kernel:"rbf"	100%	100%	100%	100%
KNN Classifier	n_neighbor: 4	100%	100%	100%	100%
Random Forest Classifier (RFC)	max_depth:9 ; min_sample_leaf:3 ; min_sample_split:4 ; n_estimators:100	100%	100%	100%	100%
AdaBoost Classifier	learning_rate:0.999 ; n_estimators:400	99.670%	99.397%	100%	99.697%
Bagging Classifier	max_samples:0.9 ; n_estimators:30	100%	100%	100%	100%

Table 2. Continued.

ML Model	Employed Hyper-Parameters	Accuracy	Precision	Recall	F1 Score
XGBoost Classifier	learning_rate:0.1 ; max_depth:7 ; n_estimators:100	100%	100%	100%	100%
Logistic Regression	C=0.5 ; penalty:"L1" ; solver:"saga"	86.140%	84.746%	90.909%	87.719%
ANN (MLP)	5 layers and their neurons: [13,13,6,6 (Hidden layers),1 (Classifier)] ;activation functions: [ReLU*4 , Sigmoid] ; optimizer: "Adam" ; loss function: Cross Entropy Loss ; learning rate: 0.001 ; L2_regularization: 0.001 ; epoch: 100 ; batch_size: 10	100%	100%	100%	100%

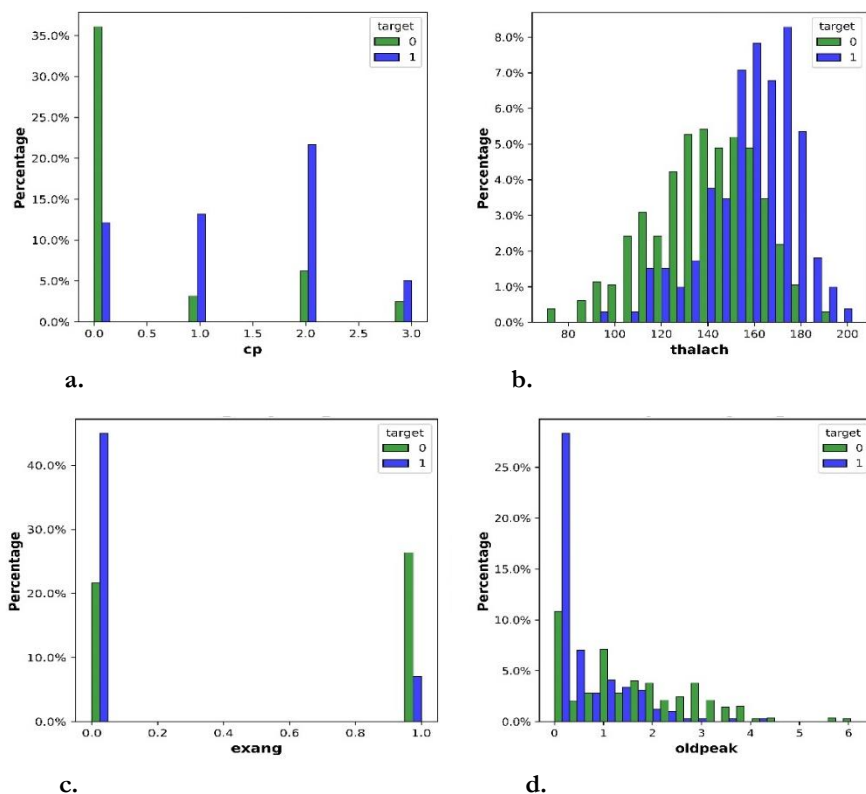


Fig. 2. Studying the most highly correlated variables, directly or conversely, on the target; a. cp, b. thalach, c. exang, d. oldpeak by target.

3.2 | Results of Principal Component Analysis

PCA inevitably results in some degree of information loss due to the dimension reduction and decreased variances. The extent of this loss is influenced by the variance of the features and their correlations. So, *Table 3* demonstrates the information loss when reducing data from 13 dimensions to fewer variables. For example, condensing 13 predictive variables into 11 retains approximately 93.84% of the information, but reducing it further results in 90% conserved information. Since effectively summarizing data dimensions is paramount in the health sector, given the high stakes involved, this analysis is immensely beneficial.

Table 3. The amount of information loss using principal component analysis.

From 13 Features to X Features	Information Loss	X=7	26.52 %
X=12	2.84 %	X=6	33.26 %
X=11	6.16 %	X=5	40.74 %
X=10	10.18 %	X=4	48.42 %
X=9	15.03 %	X=3	57.42 %
X=8	20.61 %	X=2	66.66 %

4 | Discussion

4.1 | Studying Criteria and Misclassifications

In this study, we evaluated model performance using accuracy, precision, recall, and F1 score for two main reasons. First, while accuracy is commonly used, it can be misleading with imbalanced datasets. However, our dataset was relatively balanced in terms of target, reducing this concern. Secondly, accuracy alone does not reveal misclassification types or locations. To address this, we included precision, recall, and F1 score. The recall is crucial as it is sensitive to false negatives, where diseased individuals are misclassified as healthy, potentially leading to severe consequences like death. Thus, the cost of false negatives is higher than that of false positives [24]. Therefore, recall is prioritized over precision in our study. As shown in *Table 2*, Logistic Regression, despite having the lowest overall performance, achieved a better recall compared to other criteria, which is good. Additionally, the F1 score is valuable and studied as it balances both types of misclassifications by being the harmonic mean of precision and recall.

4.2 | Studying Nuances, Outcomes, and Usability

The study's outcomes are highly beneficial for the health sector. Studying the correlation matrix, feature distribution, and analysis of important features provides valuable insights, making it easier for healthcare professionals to make informed suggestions. For instance, *Fig. 1* and *Fig. 2* indicate that doctors should advise patients to visit them if they experience intense CP or a rapidly increasing heartbeat during physical activities. In other words, according to the data, they are influential factors to consider.

Furthermore, analyzing all features in practical applications can often be costly and time-consuming. Therefore, focusing on the most important features is crucial, and EDA plays a pivotal role. Therefore, we concentrate on high correlations between predictive variables and the target and also among predictive variables themselves to evaluate feature dependency and avoid the curse of dimensionality. PCA was also shown to be a viable solution to address these concerns by summarizing data dimensions, despite information loss. Since we achieved 100% without dimension reduction in this study, PCA only revealed some theoretical insights. Nonetheless, insights can be highly beneficial and practical in other applications.

Moreover, the study's impressive results, with nearly 100% accurate predictions across all criteria and no overfitting or bias, are noteworthy. Despite the relatively small sample size, 1025 for training and 303 for testing, the findings have practical implications for reducing mortality rates. Additionally, the features proved effective for heart disease prediction, yielding excellent outcomes in both simple models like KNN and DT, and complex models like SVC and ANN. The only consideration is to utilize a considerably larger sample size, ideally in millions, and also more features, as they give rise to enhanced generalization and robustness.

Since almost all models performed exceptionally well, they allow doctors and health organizations to choose different models based on their tastes and preferences. For example, some doctors may prefer DT to intricate models like ANN or SVC due to their transparency and ease of understanding, while some others may be keen on ANN, knowing that they are more complex, and their performance will not be saturated with large datasets.

4.3 | Comparison with Similar Works

For this dataset, code execution on a Kaggle repository achieved 100% accuracy with RFC and 84% accuracy with Logistic Regression, respectively, validating a part of the present work [17]. However, there are similar studies on heart disease prediction. For example, another study used a data science and ML approach, applying various ML algorithms from the Scikit learn library to predict heart disease. In that study, Logistic Regression achieved the best performance among conventional ML models, with 86.49 % accuracy, 91% recall, 82% precision, and an 86% F1 score [25]. The fact that Logistic Regression performed best in that study, while it underperformed considerably in the current study, suggests that all models are worth exploring. Yet, the

model that demonstrates the best performance and generalization on unseen test data should be used for practical applications [8], [25].

The most similar and comprehensive study [14] also focused on accurately predicting cardiovascular disease using various ML techniques. This study achieved 100% accuracy and 100% sensitivity with models such as KNN, DT, and Random Forests. The fact that a relatively simple model like KNN performed as well as the more complex Random Forest underscores the importance of examining a range of models. Additionally, while we examined feature importance using a correlation matrix, this study explored feature importance derived from ML models. Comparing the key features identified by different methods can provide valuable insights. For instance, our study found “Cp” and “Thalach” to have the strongest positive correlations with the disease based on the correlation matrix. Conversely, [14] shows slightly different results. While DT, RF, and Logistic Regression all identify “Cp” as the most influential feature, the AdaBoost classifier highlights “Chol” as the most important. Furthermore, although we identified “Thalach” as the second most influential feature, this was only validated by the AdaBoost classifier. In contrast, DT, RF, and Logistic Regression pointed to other features as being more influential.

4.4 | Suggestions for Future Works

- I. To create a generalized model for a vast population, it is crucial to study large datasets with millions of observations consisting of more attributes and potentially various nationalities for superior analysis.
- II. Exploring the curse of dimensionality and dimension reduction methods, especially in the health sector, where frequent high-dimensional data needs summarization for human interpretation.
- III. Integrating diverse data types, such as tabular data, CT/MRI images, heartbeat alterations, and brain signals. This combination can be groundbreaking research as it enhances prediction accuracy and sheds light on complex patterns and hidden correlations with better EDA.

4 | Conclusion

Given the persistent prevalence of heart disease as a leading cause of death globally, this study aimed to predict heart disease using EDA, various conventional ML algorithms, and a state-of-the-art ANN. The dataset, sourced from the Kaggle repository, comprised 1025 training and 303 test samples with 13 predictive variables. Initial EDA included correlation matrices and data distribution assessments. Techniques such as normalization, grid searching, and 5-fold cross-validation were employed to enhance performance and robustness. Subsequently, multiple ML models were applied. While logistic regression underperformed with an accuracy of 84.16%, most models achieved almost 100% performance, with nearly no misclassification among the 303 test samples. The study also reported optimized parameters for these models in addition to excellent performance of models in terms of precision, recall, F1-score, and accuracy. PCA was conducted to evaluate the summarization of attributes, and the results were thoroughly examined. An in-depth discussion clarified the study’s outcomes, compared results with the most similar research, and explored the real-world applications of the findings. This approach not only demonstrated commendable performance theoretically but also highlighted potential avenues for saving lives.

References

- [1] Mohan, S., Thirumalai, C., & Srivastava, G. (2019). Effective heart disease prediction using hybrid machine learning techniques. *IEEE Access*, 7, 81542–81554. <https://doi.org/10.1109/ACCESS.2019.2923707>
- [2] Al-Alshaikh, H. A., P., P., Poonia, R. C., Saudagar, A. K. J., Yadav, M., AlSagri, H. S., & AlSanad, A. A. (2024). Comprehensive evaluation and performance analysis of machine learning in heart disease prediction. *Scientific reports*, 14(1), 7819. <https://doi.org/10.1038/s41598-024-58489-7>
- [3] Baashar, Y., Alkaws, G., Alhussian, H., Capretz, L. F., Alwadain, A., Alkahtani, A. A., & Almomani, M. (2022). Effectiveness of artificial intelligence models for cardiovascular disease prediction: Network meta-analysis. *Computational intelligence and neuroscience*, 2022(1), 5849995. <https://doi.org/10.1155/2022/5849995>

- [4] Baghdadi, N. A., Farghaly Abdelaliem, S. M., Malki, A., Gad, I., Ewis, A., & Atlam, E. (2023). Advanced machine learning techniques for cardiovascular disease early detection and diagnosis. *Journal of big data*, 10(1), 144. <https://doi.org/10.1186/s40537-023-00817-1>
- [5] Babu, S. V., Ramya, P., & Gracewell, J. (2024). Revolutionizing heart disease prediction with quantum-enhanced machine learning. *Scientific reports*, 14(1), 7453. <https://www.nature.com/articles/s41598-024-55991-w>
- [6] Shah, D., Patel, S., & Bharti, S. K. (2020). Heart disease prediction using machine learning techniques. *Springer nature computer science*, 1(6), 345. <https://doi.org/10.1007/s42979-020-00365-y>
- [7] 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), 8387680. <https://doi.org/10.1155/2021/8387680>
- [8] Rokhva, S., Teimourpour, B., & Soltani, A. H. (2024). Computer vision in the food industry: Accurate, real-time, and automatic food recognition with pretrained MobileNetV2. *Food and humanity*, 3, 100378. <https://doi.org/10.1016/j.foohum.2024.100378>
- [9] Bhavakar, G. S., Das Goswami, A., Vasantao, C. P., Gaikwad, A. K., Zade, A. V., & Vyawahare, H. (2024). Heart disease prediction using machine learning, deep Learning, and optimization techniques, semantic review. *Multimedia tools and applications*, 83(39), 86895–86922. <https://doi.org/10.1007/s11042-024-19680-0>
- [10] Rokhva, S., Teimourpour, B., & Soltani, A. H. (2024). *AI in the food industry: Utilizing EfficientNet B7 & transfer learning for accurate and real-time food recognition*. <https://dx.doi.org/10.2139/ssrn.4903767>
- [11] Talaei Khoei, T., Ould Slimane, H., & Kaabouch, N. (2023). Deep learning: Systematic review, models, challenges, and research directions. *Neural computing and applications*, 35(31), 23103–23124. <https://doi.org/10.1007/s00521-023-08957-4>
- [12] Alijani, S., Fayyad, J., & Najjaran, H. (2024). Vision transformers in domain adaptation and domain generalization: a study of robustness. *Neural computing and applications*, 36(29), 17979–18007. <https://doi.org/10.1007/s00521-024-10353-5>
- [13] Freiesleben, T., & Grote, T. (2023). Beyond generalization: A theory of robustness in machine learning. *Synthese*, 202(4), 109. <https://doi.org/10.1007/s11229-023-04334-9>
- [14] Ali, M. M., Paul, B. K., Ahmed, K., Bui, F. M., Quinn, J. M. W., & Moni, M. A. (2021). Heart disease prediction using supervised machine learning algorithms: Performance analysis and comparison. *Computers in biology and medicine*, 136, 104672. <https://doi.org/10.1016/j.combiomed.2021.104672>
- [15] Zhou, C., Dai, P., Hou, A., Zhang, Z., Liu, L., Li, A., & Wang, F. (2024). A comprehensive review of deep learning-based models for heart disease prediction. *Artificial intelligence review*, 57(10), 263. <https://doi.org/10.1007/s10462-024-10899-9>
- [16] Ogundepo, E. A., & Yahya, W. B. (2023). Performance analysis of supervised classification models on heart disease prediction. *Innovations in systems and software engineering*, 19(1), 129–144. <https://doi.org/10.1007/s11334-022-00524-9>
- [17] kaggle. (2024). *Heart disease prediction using RFC and LR 100%*. <https://www.kaggle.com/code/devbatrax/heart-disease-prediction-using-rfc-and-lr-100>
- [18] Sepehri, M. M., Naderi, S., & Naderi, M. (2017). Through improving the model for patients with chest pain in the heart emergency department. *Payavard-salamat*, 11(2), 235-246. **(In Persian)**. <http://payavard.tums.ac.ir/article-1-6240-en.html>
- [19] Farki, A., & Noughabi, E. A. (2023, May). Real-time blood pressure prediction using apache spark and kafka machine learning. In *2023, the 9th international conference on web research (ICWR)* (pp. 161-166). IEEE. <https://doi.org/10.1109/ICWR57742.2023.10138962>
- [20] Salehi, A., Aghdasi, M., Khatibi, T., & SheikhMohammadI, M. (2023). Data quality in process mining: A systematic review. *Sciences and techniques of information management*, 9(3), 103–160. <https://doi.org/10.22091/stim.2022.7800.1737>
- [21] Mamdouh Farghaly, H., & Abd El-Hafeez, T. (2023). A high-quality feature selection method based on frequent and correlated items for text classification. *Soft computing*, 27(16), 11259–11274. <https://doi.org/10.1007/s00500-023-08587-x>
- [22] Hall, M. A. (1999). *Correlation-based feature selection for machine learning*. <https://hdl.handle.net/10289/15043>

-
- [23] Woodman, R. J., & Mangoni, A. A. (2023). A comprehensive review of machine learning algorithms and their application in geriatric medicine: Present and future. *Aging clinical and experimental research*, 35(11), 2363–2397. <https://doi.org/10.1007/s40520-023-02552-2>
 - [24] Géron, A. (2022). *Hands-on machine learning with scikit-learn, keras, and tensorflow*. O'Reilly Media. <https://books.google.com/books?id=X5ySEAAAQBAJ>
 - [25] kaggle. (2024). *Heart disease predictions*. <https://kaggle.com/code/desalegngeb/heart-disease-predictions>