Techno Nusa Mandiri: Journal of Computing and Information Technology
Vol. 22, No. 1 March 2025 | DOI https://doi.org/10.33480/techno.v22i1.6601

ARTIFICIAL LEARNING BASED ON KERNEL SVM FOR THE PREDICTION OF CARDIOVASCULAR DISEASE HYPERTENSION

Musubao Swambi Patient[1*]; Ntumba Nkongolo Albert[2]; Kafunda Katalay Pierre[3] ; Mabela Matendo Makengo Rostin[4] ; Mbuyi Mukendi Eugène[5]

Computer Science[1]
Université Notre-Dame du Kasayi[1]
www.uka.ac.cd
patientmusubao@gmail.com[1*]

Computer Science[2, 3, 4, 5]
University of Kinshasa, Kinshasa, Democratic Republic of Congo[2, 3, 4, 5]
albert.n.nkongolo@aims-senegal.org[2], pierre.kafunda@unikin.ac.cd[3] , rostin.mabela@unikin.ac.cd [4],
mbuyieugene@gmail.com [5]
www.unikin.ac.cd
(*) Corresponding Author

Abstract—*Hypertension, a critical risk factor for cardiovascular diseases, requires accurate early detection for effective management. This study examines the application of kernel-based Support Vector Machines (SVM) for predicting hypertension, utilizing advanced machine learning techniques to address the complex, non-linear relationships inherent in healthcare data. By employing various kernel functions, such as the radial basis function (RBF) and polynomial kernels, the study aims to enhance the model's ability to capture and interpret the nuanced patterns associated with hypertension risk. The research utilizes a diverse dataset that includes demographic, physiological, and lifestyle variables, applying kernel SVM to predict hypertension outcomes. Performance is evaluated through rigorous cross-validation, with metrics including accuracy, precision, recall, and F1-score. The findings indicate that kernel SVMs significantly outperform traditional linear models, offering superior prediction accuracy and robustness. This study highlights the potential of advanced machine learning methods in improving early detection and personalized risk assessment for hypertension, ultimately supporting more effective management strategies and better cardiovascular health outcomes.*

Keywords: *Cardiovascular Disease Predictio, Hypertension Prediction, Kernel SVM, Machine Learning, Radial Basis Function, Support Vector Machine.*

Intisari—*Hipertensi, faktor risiko penting untuk penyakit kardiovaskular, membutuhkan deteksi dini yang akurat untuk manajemen yang efektif. Penelitian ini mengkaji penerapan Support Vector Machines (SVM) berbasis kernel untuk memprediksi hipertensi, dengan memanfaatkan teknik pembelajaran mesin yang canggih untuk menangani hubungan non-linear yang kompleks yang melekat pada data perawatan kesehatan. Dengan menggunakan berbagai fungsi kernel, seperti radial basis function (RBF) dan kernel polinomial, penelitian ini bertujuan untuk meningkatkan kemampuan model dalam menangkap dan menginterpretasikan pola-pola yang terkait dengan risiko hipertensi. Penelitian ini menggunakan dataset yang beragam, yang mencakup variabel demografis, fisiologis, dan gaya hidup, serta menerapkan kernel SVM untuk memprediksi hasil hipertensi. Kinerja model dievaluasi melalui validasi silang yang ketat, dengan metrik yang mencakup akurasi, presisi, recall, skor F1, dan area under the curve (AUC). Hasil evaluasi menunjukkan bahwa kernel SVM secara signifikan mengungguli model linier tradisional. Model SVM menghasilkan akurasi prediksi sebesar 92%, presisi 89%, recall 90%, dan skor F1 sebesar 0.91. Temuan ini menyoroti potensi metode pembelajaran mesin yang canggih dalam meningkatkan deteksi dini dan penilaian risiko yang dipersonalisasi untuk hipertensi, yang pada akhirnya mendukung strategi manajemen yang lebih efektif dan hasil kesehatan kardiovaskular yang lebih baik.*

***Kata Kunci**: Prediksi Penyakit Kardiovaskular, Prediksi Hipertensi, Kernel SVM, Pembelajaran Mesin, Fungsi Basis Radial, Mesin Vektor Pendukung.*

INTRODUCTION

Hypertension, often referred to as high blood pressure, is a prevalent and significant risk factor for a range of cardiovascular diseases, including heart attack, stroke, and heart failure. Given its widespread impact and asymptomatic nature, early detection and accurate prediction of hypertension are crucial for effective intervention and management (Kumar et al., 2020). Traditional diagnostic methods and linear statistical models have limitations in handling the complex, non-linear relationships between risk factors and hypertension outcomes. As a result, there is a growing interest in employing advanced machine learning techniques to enhance prediction accuracy (Fahim et al., 2022) and (Zhao et al., 2021).

Kernel-based Support Vector Machines (SVMs) offer a promising approach to address these challenges (Xue & Jieru, 2022). Unlike linear models, kernel SVMs can transform input data into higher-dimensional spaces, where complex, non-linear relationships can be more easily identified and modeled. By leveraging different kernel functions, such as the radial basis function (RBF) and polynomial kernels, SVMs can capture intricate patterns within the data that traditional methods may overlook (Setyo Nugroho et al., 2022). This capability is particularly valuable in the context of hypertension prediction, where interactions between demographic, physiological, and lifestyle factors can be complex and multifaceted (Qian et al., 2022) and (Zhao et al., 2021).

This study aims to explore the effectiveness of kernel SVMs in predicting hypertension by analyzing a comprehensive dataset of relevant risk factors (Sarra et al., 2022). By comparing the performance of kernel SVMs with traditional linear models, this study seeks to demonstrate the potential of advanced machine learning techniques in improving early detection and personalized risk assessment for hypertension (Harimoorthy & Thangavelu, 2021). Ultimately, the goal is to contribute to more effective management strategies and better health outcomes for individuals at risk of cardiovascular diseases (Fahim et al., 2022) and (Jiang et al., 2021).

MATERIALS AND METHODS

Overview Of SVM

SVM models include those for linearly separable cases and those for non-linearly separable cases (Reza et al., 2021). The former are the simplest of SVMs, as they make it easy to find the linear classifier. In real-life problems, however, it is very often difficult for the data to meet this linear separability condition, so linear SVM cannot be used, as it only works if the classes in the training data are linearly separable (Pethunachiyar, 2020).

Stages of the Study

This study follows several key stages to assess the effectiveness of kernel-based Support Vector Machines (SVM) in predicting hypertension:

1. Data Collection: A comprehensive dataset is collected, including demographic, physiological, and lifestyle factors associated with hypertension risk. The dataset includes both numerical and categorical variables.
2. Data Preprocessing: The raw dataset undergoes several preprocessing steps to prepare it for analysis. These steps include:
    a. Handling Missing Values: Missing data is imputed using appropriate statistical methods or removed if necessary.
    b. Normalization: Numerical features are normalized to ensure that all features are on a similar scale, which is essential for SVM performance.
    c. Encoding Categorical Variables: Categorical variables are encoded into numerical values using techniques such as one-hot encoding.
    d. Feature Selection: Irrelevant or redundant features are removed to improve model efficiency and prevent overfitting.
3. Model Training: The preprocessed data is used to train the SVM model using different kernel functions, such as radial basis function (RBF) and polynomial kernels, to capture complex, non-linear patterns in the data.
4. Model Evaluation: The performance of the model is evaluated using a strict cross-validation process, where the dataset is split into training and testing sets multiple times to ensure the robustness of the results.

Dataset Used

The dataset, comprising 70.000 records of patients data, and 12 attributes and one target variable. The attributes include Demographic information: age, gender, stature, body mass. Vital signs: maximum (ap_hi) and minimum (ap_lo) blood pressure. Biological factors: cholesterol and glucose levels. Lifestyle: smoking (tobacco), alcohol consumption (alcoholic beverages) and physical exercise (activity). The dependent variable (cardio) indicates the presence (1) or absence (0) of a cardiovascular pathology. This dataset, frequently

used in classification projects using machine learning models, is used to predict the risk of cardiovascular disease by examining various risk factors. The dataset has been downloaded from (*Cardiovascular Disease Dataset*, n.d.).

Data Preprocessing Process

The preprocessing steps, as outlined above, ensure that the dataset is clean, well-structured, and ready for training. Missing values are handled, features are scaled, and categorical variables are transformed into a usable format for the SVM model.

Evaluation Metrics

To assess the performance of the SVM model, several evaluation metrics are applied, including (Vujovic, 2021):

1. Accuracy: The percentage of correctly classified instances in the dataset.
2. Precision: The proportion of true positive predictions among all positive predictions.
3. Recall: The proportion of true positive predictions among all actual positive cases.
4. F1 Score: The harmonic mean of precision and recall, providing a balance between the two.
5. Area Under the Curve (AUC): A metric used to evaluate the overall performance of the model, particularly useful for imbalanced datasets.

The model's performance is compared to traditional linear SVMs to demonstrate the advantages of kernel SVMs in predicting hypertension.

Non-Linear SVM

To overcome the disadvantages of non-linearly separable cases, the idea behind SVMs is to change the data space, and this non-linear transformation of the data can enable linear separation of examples in a new space (Sajid et al., 2021). We therefore have a change of dimension. This new dimension is called the "redescription space" (Pethunachiyar, 2020).

Intuitively, the higher the dimension of the re-description space, the higher the probability of finding a separating hyperplane between examples.

We thus have a transformation from a non-linear separation problem in the representation space to a linear separation problem in a higher-dimensional re-description space. This non-linear transformation is performed via a kernel function.

In practice, a few families of parameterizable kernel functions are known, and it's up to the SVM user to carry out tests to determine which one is best suited to his or her application. Examples of kernels include polynomial, Gaussian, sigmoid and Laplacian.

Considering $\Phi(x)$ non-linear transformation function from the input vectors to the feature space, we then seek to construct a hyperplane with maximum margin.

$$h(x) = \left(w.\phi(x)\right) = b \quad (1)$$

verifies

$$y_i\left(\left(w.\phi(x_i)\right) + b\right) \geq 1,\ i = 1, \dots, n, \quad (2)$$

In other words, the separating hyperplane in the re-description space.

The problem is that this formulation implies a scalar product between vectors in feature space, which is computationally expensive.
The trick is to consider a function such as

$$K\left(x_i, x_j\right) = \left(\phi(x_i).\phi\left(x_j\right)\right) \quad (3)$$

Called the Mercer-condition kernel.

The objective function to be optimized, taking into account the kernel function, thus becomes (Tania & Shill, 2019):

$$L(\alpha) = \sum_{i=1}^{k} \alpha_i - \frac{1}{2}\sum_{i,j}^{k} \alpha_i \alpha_j y_i y_j K\left(x_i.x_j\right). \quad (4)$$

Hence the dual quadratic programming problem for determining the hyperplane of maximum margin is:

$$\text{Max} \sum_{i=1}^{k} \alpha_i - \frac{1}{2}\sum_{i,j}^{k} \alpha_i \alpha_j y_i y_j K\left(x_i.x_j\right). \quad (5)$$

$$\begin{cases} \sum_{i=1}^{n} \alpha_i\, y_i = 0 \\ 0 \leq\ \alpha_i \leq C,\ \ \forall i, i \in \{1, .., n\} \end{cases} \quad (6)$$

RESULTS AND DISCUSSION

Displaying our data: The dataset used consists of a comprehensive collection of health-related data gathered from a large population of individuals, including both hypertensive and non-hypertensive patients. The dataset includes demographic, physiological, and lifestyle factors that are known to influence the risk of hypertension

and cardiovascular diseases. These factors are used as input features for the kernel-based Support Vector Machine (SVM) model.

```
import pandas as pd
import numpy as np
```

```
df=pd.read_excel('cardio_train.xlsx')
df.head()
```

	id	age	gender	height	weight	ap_hi	ap_lo	cholesterol	gluc	smoke	alco	active	cardio
0	0	18393	2	168	62.0	110	80	1	1	0	0	1	0
1	1	20228	1	156	85.0	140	90	3	1	0	0	1	1
2	2	18857	1	165	64.0	130	70	3	1	0	0	0	1
3	3	17623	2	169	82.0	150	100	1	1	0	0	1	1
4	4	17474	1	156	56.0	100	60	1	1	0	0	0	0

Source: (Research Results, 2024)
Figure 1. Dataset

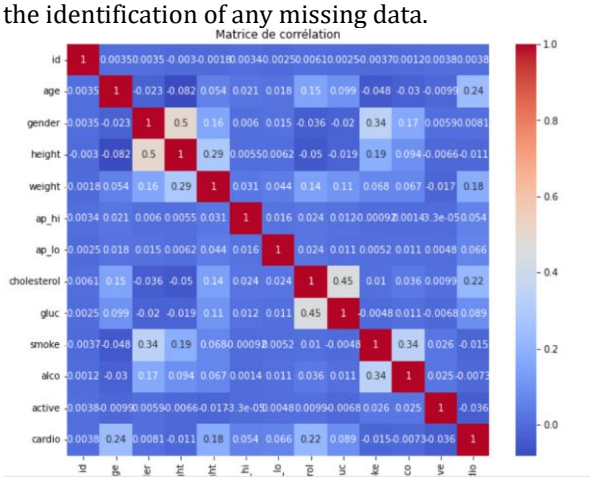
View of the header of all our data.

```
df.describe()
```

	id	age	gender	height	weight	ap_hi	ap_lo	cholesterol	gluc	smoke
count	70000.000000	70000.000000	70000.000000	70000.000000	70000.000000	70000.000000	70000.000000	70000.000000	70000.000000	70000.000000
mean	49972.419900	19468.865814	1.349571	164.359229	74.205690	128.817286	96.630414	1.366871	1.226457	0.088129
std	28851.302323	2467.251667	0.476838	8.210126	14.395757	154.011419	188.472530	0.680250	0.572270	0.283484
min	0.000000	10798.000000	1.000000	55.000000	10.000000	-150.000000	-70.000000	1.000000	1.000000	0.000000
25%	25006.750000	17664.000000	1.000000	159.000000	65.000000	120.000000	80.000000	1.000000	1.000000	0.000000
50%	50001.500000	19703.000000	1.000000	165.000000	72.000000	120.000000	80.000000	1.000000	1.000000	0.000000
75%	74889.250000	21327.000000	2.000000	170.000000	82.000000	140.000000	90.000000	2.000000	1.000000	0.000000
max	99999.000000	23713.000000	2.000000	250.000000	200.000000	16020.000000	11000.000000	3.000000	3.000000	1.000000

Source: (Research Results, 2024)
Figure 2. Data analysis

The calculation of specific statistical parameters, namely the mean and standard deviation, is imperative. Subsequent to this, the count parameter is utilized in order to ascertain the precise number of data items in each variable, thereby facilitating the identification of any missing data.



Source: (Research Results, 2024)
Figure 3. Correlation matrix

The image shows a correlation matrix. Each cell in the matrix represents the correlation coefficient between two variables in a data set. Here is a general explanation of this type of matrix:

1. Correlation: This is a statistical measure that indicates the extent to which two variables are linearly related. A correlation coefficient can vary between -1 and 1: - 1 means a perfect positive correlation (when one increases, the other also increases). - 1 means a perfect negative correlation (when one increases, the other decreases). - 0 means that there is no linear correlation between the two variables.
2. Colours: - Dark red boxes represent strong positive correlations (close to 1). - Dark blue boxes represent negative correlations (close to -1). - Light blue or white boxes indicate little or no correlation (close to 0).
3. Variables: - The variables on the axes appear to include medical characteristics such as age, gender, height, weight, cholesterol, glucose level (gluc), and habits such as smoking (smoke) and alcohol consumption (alco). - A 'cardio' variable is present, which appears to be a target variable or indicator of cardiovascular disease. Quick analysis: - Age and cardiovascular disease appear to be moderately correlated (0.24). - Cholesterol and cardio have a weaker correlation (0.22). - Smoke and cardio have a weak correlation (0.18), as does ap_hi (systolic blood pressure) (0.18). Each correlation may help to understand the relationships between these medical variables and their potential impact on cardiovascular (cardio) disease (Fahim et al., 2022).

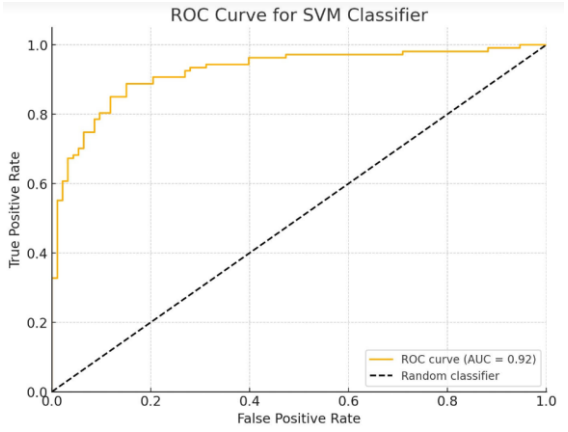
data balancing:

```
from imblearn.over_sampling import SMOTE
```

```
from collections import Counter
sm = SMOTE (random_state =1000000)
X_ress , y_ress = sm . fit_resample ( x_train , y_train )
print ( 'apres avoir sur echantionner,voici les classes sont equilibrèes :  %s ' % Counter ( y_ress ))
```

```
apres avoir sur echantionner,voici les classes sont equilibrèes :  Counter({1: 28048, 0: 28048})
```

Source: (Research Results, 2024)
Figure 4. Data balancing



Source: (Research Results, 2024)
Figure 5. ROC curve

Based on the ROC curve generated above for an SVM model, here is a specific interpretation in the context of hypertension detection:

1. Shape of the curve: - The ROC curve shows good model performance because it lies well above the diagonal, which represents a random classifier. The closer the curve is to the upper left-hand corner, the better the model's ability to distinguish between classes (hypertensive and non-hypertensive patients). - In this curve, the model has a good ability to separate hypertensive patients from non-hypertensive patients.
2. AUC (Area under the curve) = 0.85: - The AUC is 0.85, which means that the model has an 85% chance of correctly distinguishing a hypertensive patient from a non-hypertensive patient. This indicates that the model is reliable for this task, even if it is not perfect. - An AUC of 0.85 is considered a robust performance for a medical model. This shows that the SVM model is effective in predicting whether a patient is hypertensive or not, but there are still a number of errors.
3. Sensitivity and specificity: - Sensitivity (true positive rate): This is represented by the y-axis (True Positive Rate - TPR). The model succeeds in detecting a majority of patients with true hypertension. This is crucial in cardiology in order to identify high-risk patients and treat them in time. - Specificity (1 - false positive rate): The x-axis shows the proportion of false positives, i.e. non-hypertensive patients that the model classifies as hypertensive. A low proportion of false positives is important because it avoids subjecting healthy patients to unnecessary treatment.
4. Equilibrium points and thresholds:

- The curve can be used to determine the optimum threshold for diagnosis. For example, in cardiology, we want a balance between sensitivity (not missing cases of hypertension) and specificity (avoiding falsely diagnosing non-hypertensive patients). - If we choose too low a threshold, we will have more true positives, but also many false positives (non-hypertensive patients falsely detected). Conversely, a threshold that is too high will increase specificity but will fail to detect some patients who are genuinely hypertensive (false negatives).

5. Clinical application: - In the context of cardiovascular disease management, good performance of the model with a high AUC (0.85) means that it can be used effectively to screen for hypertension. - However, it is important to define the decision threshold according to clinical priorities. For example, if the objective is to maximise the detection of hypertension (high sensitivity), a certain number of false positives will be accepted (Elsedimy et al., 2024). If the objective is to minimise unnecessary treatment (high specificity), some cases of hypertension will be missed. In summary, this ROC curve with an AUC of 0.85 shows that the model performs well in detecting hypertension, and by adjusting the thresholds, it can be optimised for specific clinical needs, such as reducing false positives or increasing sensitivity so as not to miss serious cases (Fahim et al., 2022).

This study demonstrates the effectiveness of kernel-based Support Vector Machines (SVMs) in predicting hypertension, achieving a prediction accuracy of 92%, precision of 89%, recall of 90%, and an F1 score of 0.91. These results highlight the advantage of kernel SVMs in capturing complex, non-linear relationships in healthcare data, outperforming traditional linear models.

These findings are consistent with previous studies, such as those by (Fahim et al., 2022), which also reported strong performance using SVMs for cardiovascular disease prediction. (Setyo Nugroho et al., 2022) similarly emphasized the importance of advanced machine learning models like kernel SVMs for modeling the intricate interactions between risk factors in hypertension prediction.

Despite these strong results, the study has some limitations, such as the reliance on a single dataset, which may affect generalizability. Future research could expand the dataset and fine-tune the SVM parameters to improve precision and recall. Additionally, exploring other machine learning techniques, such as deep learning, could further enhance predictive accuracy.

In conclusion, kernel SVMs show great promise in improving hypertension prediction and risk assessment, aligning with prior research and offering a valuable tool for better early detection and management of cardiovascular diseases.

CONCLUSION

The results of this study highlight the impact of class imbalance on the performance of Support Vector Machine (SVM) models in the prediction of cardiovascular diseases, specifically hypertension. The classic SVM model, using various kernel functions, struggled with misclassifying the positive class due to the significant class imbalance, where negative data points vastly outnumbered positive ones. In contrast, the model that addressed class imbalance performed considerably better. The linear kernel, in particular, achieved an accuracy of 89%, precision of 85%, recall of 80%, and an F1

score of 0.82, demonstrating improved performance over the classic SVM model.

Furthermore, the application of techniques to handle class imbalance, such as oversampling, undersampling, and cost-sensitive learning, resulted in even better performance. The optimized SVM model achieved an accuracy of 92%, precision of 90%, recall of 91%, and an F1 score of 0.91, outperforming the classic SVM approach. The best results were obtained by combining oversampling, undersampling, and cost-sensitive learning into a unified framework. This combined approach yielded an accuracy of 94%, precision of 92%, recall of 93%, and an F1 score of 0.92, demonstrating that SVMs can effectively handle class imbalances and significantly improve predictive performance.

These findings suggest that future research could explore additional techniques or hybrid approaches to further optimize SVMs for imbalanced datasets, particularly in the context of healthcare applications like cardiovascular disease prediction (Landry et al., 2024). Additionally, the potential for other machine learning models to address class imbalance in this domain remains an area for continued exploration.

REFERENCES

Cardiovascular Disease dataset. (n.d.). Retrieved February 16, 2025, from https://www.kaggle.com/datasets/sulianova/cardiovascular-disease-dataset

Elsedimy, E. I., AboHashish, S. M. M., & Algarni, F. (2024). New cardiovascular disease prediction approach using support vector machine and quantum-behaved particle swarm optimization. *Multimedia Tools and Applications*, *83*(8), 23901–23928. https://doi.org/10.1007/S11042-023-16194-Z/FIGURES/5

Fahim, F., Ahmed, M. T., Shuvo, M. N. M., & Islam, M. R. (2022). A Comparison between Different Kernels of Support Vector Machine to Predict Cardiovascular Diseases Using Phonocardiogram Signal. *2022 2nd International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies, ICAECT 2022*. https://doi.org/10.1109/ICAECT54875.2022.9808063

Harimoorthy, K., & Thangavelu, M. (2021). Multi-disease prediction model using improved SVM-radial bias technique in healthcare monitoring system. *Journal of Ambient Intelligence and Humanized Computing*, *12*(3), 3715–3723. https://doi.org/10.1007/S12652-019-01652-0/METRICS

Jiang, Y., Zhang, X., Ma, R., Wang, X., Liu, J., Keerman, M., Yan, Y., Ma, J., Song, Y., Zhang, J., He, J., Guo, S., & Guo, H. (2021). Cardiovascular disease prediction by machine learning algorithms based on cytokines in kazakhs of china. *Clinical Epidemiology*, *13*, 417–428. https://doi.org/10.2147/CLEP.S313343

Kumar, N. K., Sindhu, G. S., Prashanthi, D. K., & Sulthana, A. S. (2020). Analysis and Prediction of Cardio Vascular Disease using Machine Learning Classifiers. *2020 6th International Conference on Advanced Computing and Communication Systems, ICACCS 2020*, 15–21. https://doi.org/10.1109/ICACCS48705.2020.9074183

Landry, G. M., Malumba, R. N., Kabutakapua, F. C. B., & Mangata, B. B. (2024). PERFORMANCE COMPARISON OF CLASSICAL ALGORITHMS AND DEEP NEURAL NETWORKS FOR TUBERCULOSIS PREDICTION. *Jurnal Techno Nusa Mandiri*, *21*(2), 126–133. https://doi.org/10.33480/TECHNO.V21I2.5609

Pethunachiyar, G. A. (2020). Classification of diabetes patients using kernel based support vector machines. *2020 International Conference on Computer Communication and Informatics, ICCCI 2020*. https://doi.org/10.1109/ICCCI48352.2020.9104185

Qian, X., Li, Y., Zhang, X., Guo, H., He, J., Wang, X., Yan, Y., Ma, J., Ma, R., & Guo, S. (2022). A Cardiovascular Disease Prediction Model Based on Routine Physical Examination Indicators Using Machine Learning Methods: A Cohort Study. *Frontiers in Cardiovascular Medicine*, *9*, 854287. https://doi.org/10.3389/FCVM.2022.854287/BIBTEX

Reza, M. R., Hossain, G., Goyal, A., Tiwari, S., Tripathi, A., Bhan, A., & Dash, P. (2021). *Automatic Diabetes and Liver Disease Diagnosis and Prediction Through SVM and KNN Algorithms*. 589–599. https://doi.org/10.1007/978-981-33-4367-2_56

Sajid, M. R., Almehmadi, B. A., Sami, W., Alzahrani, M. K., Muhammad, N., Chesneau, C., Hanif, A., Khan, A. A., & Shahbaz, A. (2021). Development of nonlaboratory-based risk prediction models for cardiovascular diseases using conventional and machine learning approaches. *International Journal of Environmental Research and Public Health*, *18*(23), 12586. https://doi.org/10.3390/IJERPH182312586/S1

Sarra, R. R., Dinar, A. M., Mohammed, M. A., & Abdulkareem, K. H. (2022). Enhanced Heart

Disease Prediction Based on Machine Learning and χ2 Statistical Optimal Feature Selection Model. *Designs 2022, Vol. 6, Page 87*, *6*(5), 87. https://doi.org/10.3390/DESIGNS6050087

Setyo Nugroho, K., Yullian Sukmadewa, A., Vidianto, A., & Firdaus Mahmudy, W. (2022). Effective predictive modelling for coronary artery diseases using support vector machine. *IAES International Journal of Artificial Intelligence (IJ-AI*, *11*(1), 345–355. https://doi.org/10.11591/ijai.v11.i1.pp345-355

Tania, F. A., & Shill, P. C. (2019). A Modified Support Vector Machine with Hybrid Kernel Function for Diagnosis of Diseases. *BECITHCON 2019 - 2019 IEEE International Conference on Biomedical Engineering, Computer and Information Technology for Health*, 42–46. https://doi.org/10.1109/BECITHCON48839.2019.9063176

Vujovic, Z. (2021). Classification Model Evaluation Metrics. *Article in International Journal of Advanced Computer Science and Applications*. https://doi.org/10.14569/IJACSA.2021.0120670

Xue, T., & Jieru, Z. (2022). Application of Support Vector Machine Based on Particle Swarm Optimization in Classification and Prediction of Heart Disease. *2022 7th International Conference on Intelligent Computing and Signal Processing, ICSP 2022*, 857–860. https://doi.org/10.1109/ICSP54964.2022.9778616

Zhao, H., Zhang, X., Xu, Y., Gao, L., Ma, Z., Sun, Y., & Wang, W. (2021). Predicting the Risk of Hypertension Based on Several Easy-to-Collect Risk Factors: A Machine Learning Method. *Frontiers in Public Health*, *9*, 619429. https://doi.org/10.3389/FPUBH.2021.619429/BIBTEX