Mesopotamian Journal of Artificial Intelligence in Healthcare
Vol.2023, **pp**. 21–26
DOI: https://doi.org/10.58496/MJAIH/2023/004; ISSN: 3005-365X
https://mesopotamian.press/journals/index.php/MJAIH

Research Article

Predicting Coronary Artery Disease Utilizing Support Vector Machines: Optimizing Predictive Model

Sarmad Muwfak Khazaal[1],*, Haitham Maarouf[1]

[1] *Computer Science Department, Modern University for Business and Science, Beirut, Lebanon*

ARTICLE INFO

Article History
Received 21 Jan 2023
Accepted 20 Mar 2023
Published 25 Mar 2023

Keywords
Artificial Intelligence
Machine Learning
Support Vector Machine
Confusion Matrix
ROC Curve

ABSTRACT

In the medical field, specialists aspire through the use of artificial intelligence to obtain accuracy in the results and details of patients. Such as automatic detection and prediction of diseases and obtaining medical results by applying artificial intelligence techniques such as machine learning algorithms to clinical data. Therefore, because coronary artery disease (CAD) is one of the most prevalent types of diseases in the world and with the highest rate of injuries, one of the machine learning algorithms will be used in this research to diagnose this disease through artificial intelligence. Patient clinical factors were used, a data set of 303 people. The Support Vector Machine (SVM) algorithm is used and applied to the clinical factors of patients, which are 56 variables, as they are considered one of the most important clinical factors that can be used to detect CAD. This algorithm is considered to have good predictive ability in medical clinical characteristics. The Support Vector Machine algorithm model, which was used in this research, provided the highest prediction accuracy of (96.7%), with an AUC value of (71.5%). This indicates the effectiveness of the SVM algorithm in classifying the clinical patient data set that was adopted in this research. After applying it to the research data set, the SVM algorithm showed its ability to classify and predict well. The accuracy of the algorithm model was (96.7%). This indicates that this algorithm can help cardiologists in different ways during their daily practices to predict coronary artery disease, which is one of the most common heart diseases.

1. INTRODUCTION

In recent years, artificial intelligence has proven its ability to be applied in various fields of life, especially in the field of diagnosis and medical prediction, which is one of the important areas in human life. Where specialists in artificial intelligence worked on the use of prediction and medical diagnosis because of its importance in medical decision-making. Thus, relying on it to estimate whether or not a person is infected with the disease, as artificial intelligence in this field is witnessing rapid growth. Where biotechnology, genomics, and artificial intelligence technologies help: in the availability of important medical data, advances in medicine techniques and pathological analyzes, and in reducing the time and cost for the examinations conducted by the patient, as well as in making the right decisions about the patient's condition and giving the appropriate medicine to him [1]. Applications of artificial intelligence, such as machine learning, are an important branch of computer science that simulates tasks performed by the human brain. Machine learning techniques are an important aspect of healthcare clinical practices that are frequently used in prediction. Because of the development witnessed by the branch of artificial intelligence, there has become a barrier between doctors and developers of machine learning technologies in terms of understanding the approach of these technologies and how they work [2].

Artificial intelligence has developed greatly in various human medicine sciences and through machine learning. And at the present time, through deep learning through the availability of a learning data set, through which artificial intelligence can diagnose various diseases, such as the endoscopic colonoscopy program, and obtain pictures and video clips about the disease without the need for human intervention [3]. The use of nested models to classify patient data by using one of the nested ensemble nu-Support Vector Classification (NE-nu-SVC) machine learning techniques that help effectively diagnose heart diseases such as coronary artery disease [4]. As well as the use of artificial intelligence in image segmentation for two- and three-dimensional printing in the field of clinical medicine. Through which tumors can be detected accurately and through these medical images, which is called the use of image segmentation based on artificial intelligence for 3D printing. The Yolov3 algorithm diagnoses prostate tumor by training it on MRI images [5]. As well as the use of artificial intelligence techniques in alternative medicine and the verification of these techniques by implementing them on a clinical data set of patients to support alternative medicine techniques [6-10].

---

*Corresponding author. Email: sarmad.m.khazaal@gmail.com

2. SUPPORT VECTOR MACHINE

SVM is one of the most famous supervised machine learning algorithms and is considered the most widely involved in many domains, especially in the medical domain. This algorithm is characterized by its ability to achieve classification and regression tasks with a small dataset with a complicated series. In addition, this algorithm aims to change the dataset into a new space in which this data diverges in a way that can be classified and sorted, where this is done by partitioning the data employing the hyperplane. The last one is a straight line that diverges the dataset for two specific variables and based on the training data, whether this data can separate linearly or not. Put differently; this line predicts that a person may or may not suffer from a disease. This mechanism is accomplished by relying on one of the points of the dataset of the two variables that are less than or higher than the straight line in order to create the prediction process. More noticeably, this algorithm is effective with a data set that is entirely linearly separated, the support vector points are as close as possible to the hyperplane, and the margins are large. Accordingly, the outcome of this algorithm will be more accurate when the number of variables in the training dataset is moreover. In the training stage, the dataset that is highly efficient in performance is trained, where the best super-level is generated based on the points of the dataset located on the supporting vectors and not on the entire training data. Unfortunately, this algorithm is not satisfactorily useful with a large dataset. For this reason, its performance was unsatisfactory in this work, as it achieved a fair measure of accuracy compared to the rest of the algorithms.

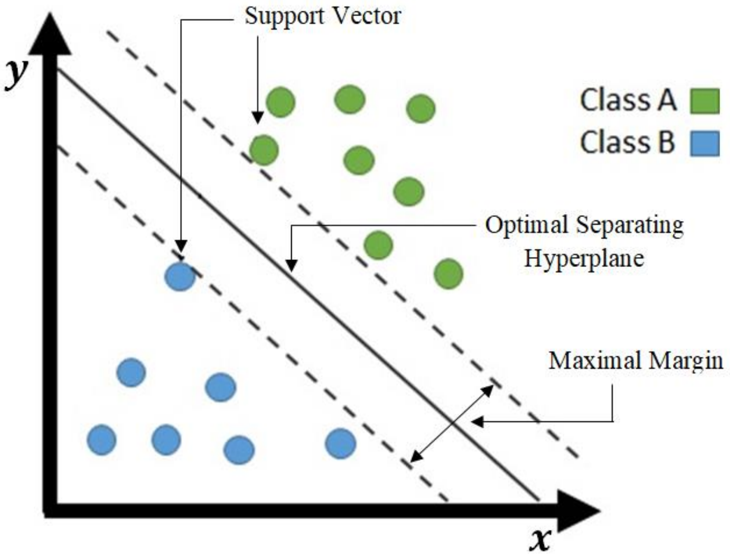

Fig. 1. Supporting vector, margins, and hyperplane in SVM

Algorithm: Support Vector Machine

```
Input: Import dataset.
Training data with labeled examples $(X, y)$.
Preprocessing: If needed, preprocess the data
(e.g., normalization, feature scaling).
Select Kernel: Choose a kernel function
(linear, polynomial, radial basis function, etc)..
                    Train/Test Split: Split the data into a training set and a test set.
$N_{in}$(the number of input vector)
$N_{sv}$(the number of support vector)
$N_{ft}$(the number of features in a support vector)
$SV[N_{sv}]$(support vector machine)
$IN[N_{in}]$(input vector array), $b*$ (bias)
Training:
a. Initialize the SVM model with parameters
(e.g., regularization parameter C, kernel parameters).
b. Solve the optimization problem to find the optimal hyperplane
that maximizes the margin between classes.
c. Use optimization techniques (e.g., Sequential Minimal
```

```
$Optimization)\ to\ find\ the\ optimal\ values\ for\ the\ Lagrange\ multipliers\ (alphas).$
$d. Compute\ the\ bias\ term\ (b)\ based\ on\ the\ support\ vectors.$
Predication:
$F1`, F2(decision\ function\ outputs)$
for $i \rightarrow 1$ to $N_{in}$ by 2 do
    $F1 = 0, F2 = 0$
    for $j \rightarrow 1$ to $N_{sv}$ by 1 do
        $dist1 = 0, dist2 = 0$
        for $k \rightarrow 1$ to $N_{ft}$ by 1 do
            $dist1 += (SV[j].feature[k] - IN[i].feature[k])^2$
            $dist2 += (SV[j].feature[k] - IN[i+1].feature[k])^2$
        end
        $k1 = \exp(-yxdist1)$
        $k2 = \exp(-yxdist2)$
        $F1 += SV[j].\alpha * xk1$
        $F2 += SV[j].\alpha * xk1$
    end
    $F1 = F1 + b^*$
    $F2 = F2 + b^*$
end
Output: Outcomes
```

Notice: The final trained SVM model along with the chosen hyperparameters.

In this algorithm, there are support vectors, which are a set of two or more data attributes that are as close as possible to the hyperplane and that will be relied upon to construct the hyperplane. In this work, this algorithm is employed to classify the heart disease dataset through the hyperplane, which diverges these data into two groups, the first is placed after the support vector point, and the second represents the data, based on which all data belonging to the second group are organised and set after the support vector point. Figure 1 illustrates the operating mechanism of this applied algorithm with an illustration of the hyperplane. This figure employs the margin technique to determine the nearest supporting vector point, which determines the distance between one point and another within the hyperplane, as it contributes to classifying the data with more remarkable accuracy. This technique is of two styles: the hard margin, where the dataset is separable linearly, and the other is the soft margin, meaning that its dataset is not separable. Furthermore, the support vector machine algorithm finds the most appropriate vector points with the most acceptable margin in analysing the behaviors of the dataset while generating the best location for the hyperlevel separating the dataset. Occasionally a group of obstacles may arise in implementing this algorithm, as it takes work to separate two categories of a specific variable. Hence, it requires shuffling the dataset through the use of transformation, which transfers the dataset to another area in order to separate the data in a way that facilitates the performance of this algorithm. Moreover, the kernel method is used, which is a mathematical form that is operated with nonlinear data, which is mainly used to solve the issue of nonlinear separation. Thus, this will solve the problem of overfitting. In the end, satisfactory performance will be obtained from this algorithm. To clarify the collection of the above-mentioned, algorithm 2 can be traced, which explains the steps of executing this algorithm.

3. RESULTS

In the SVM algorithm, eight clinical factors were selected, which showed that there is a correlation between them and the variable age, namely (Weight, HTN, BP, CR, BUN, ESR, St Depression), where the Support Vector machine algorithm was trained based on the data of these variables specifically. It turned out that 47 people were predicted by the algorithm correctly that they are sick, and they are actually sick, that is, they were classified correctly, and this is based on the variables that were chosen in this algorithm, through which the naming variable was classified into patients with the disease and not infected with the disease. It turns out that one person was correctly predicted by the algorithm to be not sick, and he is actually not sick, that is, he was classified correctly. It turned out that 12 people were incorrectly predicted by the algorithm as not sick, and they are actually sick, i.e. they were incorrectly misclassified by this algorithm. And that one person was incorrectly predicted by the algorithm to be ill, when in fact he was not ill, i.e., misclassified. The AUC value of this algorithm was 71.5% and it reflects the accuracy of the model as a whole in classification.

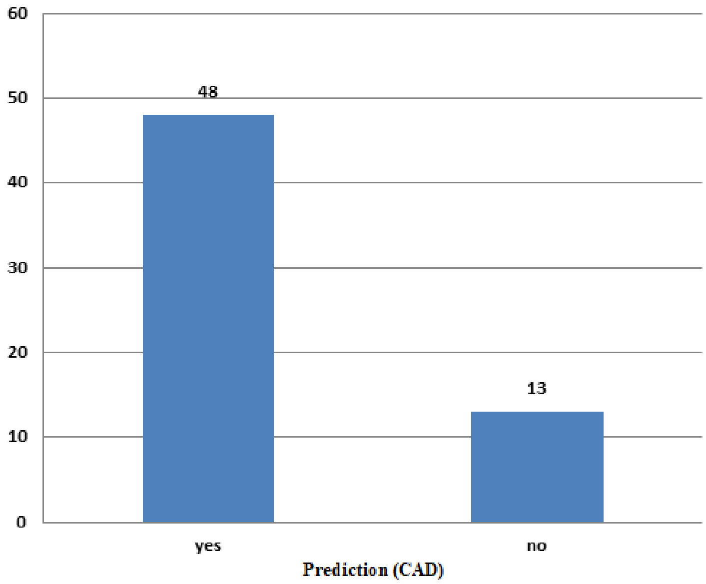

Fig. 2. Predicting sick versus non-patients by CAD variable using SVM algorithm

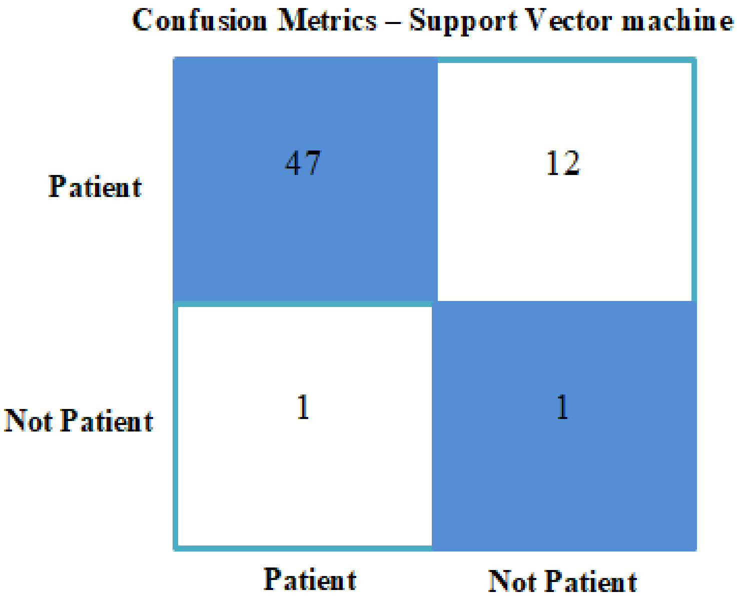

Fig. 3. Shows the confusion matrix of the SVM algorithm model

TABLE I. The result of applying the performance operator

	Outcomes					
Support Vector Machine	Accuracy%	Sensitivity %	Specificity %	Precision %	F1-Score %	AUC%
	96.7	97.9	92.3	97.9	97.9	71.5

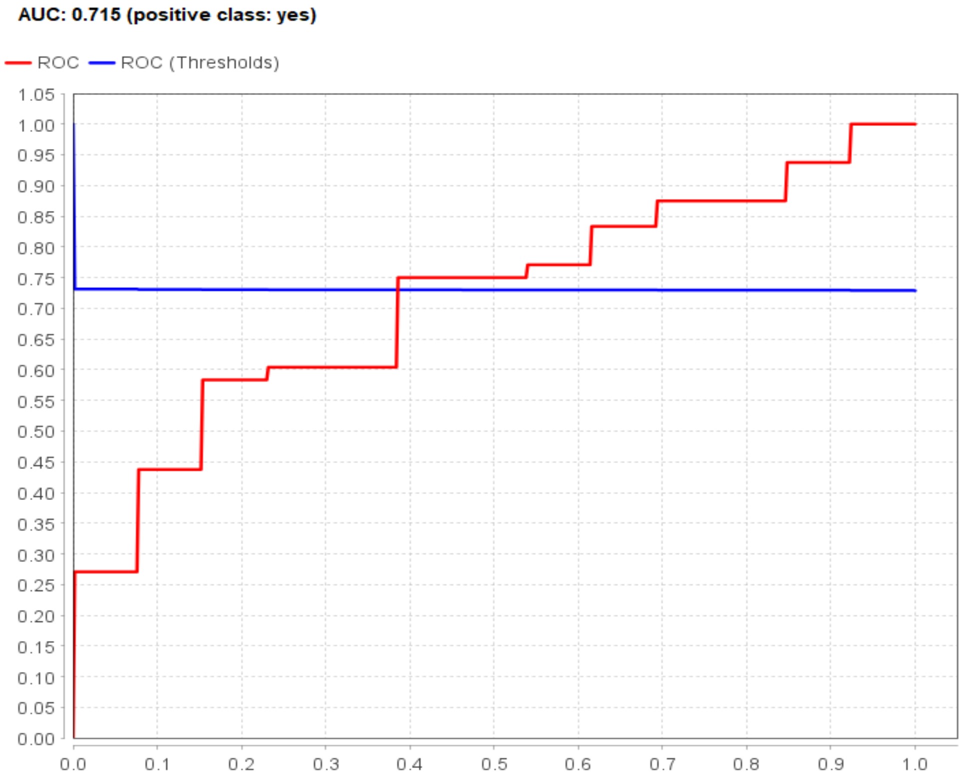

Fig. 4. The result of applying the ROC curve in relation to the variable CAD

4. CONCLUSIONS

The application of clinical machine learning algorithms is useful for healthcare physicians, as it helps in solving medical tasks such as diagnosis, predicting disease risks, prescribing medication, and other tasks. Where it was found that these techniques have an important impact on people's health. Since coronary artery disease is considered a serious disease, it has attracted attention from cardiologists and researchers. Accordingly, this research presented one of the machine learning techniques, which is the SVM algorithm, and it is considered one of the important classification algorithms in the classification of clinical data, as the prediction accuracy of the SVM algorithm model reached (96.7%) and the AUC value was (71.5%). The value of AUC indicates that the predictive model of the algorithm is good and effective, and this indicates the ability of the algorithm to diagnose coronary artery disease. This algorithm can be considered and relied upon to improve the methods of diagnosis and prediction of coronary artery disease. It is possible to repeat the use of this algorithm on a different data set, which can lead to more useful results that support the power of machine learning techniques with its ability to predict different medical diseases. In addition, medical machine learning techniques reduce time and cost in diagnosing patients with coronary artery disease. Thus, these techniques are considered to have a potential role in preventing coronary artery disease or reducing its risk. In addition, the support of computer diagnostic techniques applied to different patient data sets that are performed in medical and academic care centers by cardiologists, with an evaluation of their performance and impact.

Funding

The author had no institutional or sponsor backing.

Conflicts of Interest

The author's disclosure statement confirms the absence of any conflicts of interest.

Acknowledgment

The author extends appreciation to the institution for their unwavering support and encouragement during the course of this research.

References

[1] H. Djihane, S. Slatnia, O. Kazar, H. Saouli, and A. Merizig, "Artificial Intelligence in Healthcare: A Review on Predicting Clinical Needs," *International Journal of Healthcare Management,* vol.15, pp.267–275, 2022. doi:10.1080/20479700.2021.1886478.

[2] K. Burak, R. Cuocolo, D. P. D. Santos, A. Stanzione, and L. Ugga, "Must-Have Qualities of Clinical Research on Artificial Intelligence and Machine Learning," *Balkan Medical Journal*,. vol.40, pp.3-12, 2023. doi:10.4274/balkanmedj.galenos.2022.2022-11-51.

[3] S. Marco, D.Massimi, Y. Mori, L. Alfarone, and A. Fugazzaet al, "Artificial Intelligence-Aided Endoscopy and Colorectal Cancer Screening," *Diagnostics*, vol.13, March 2023. doi:10.3390/diagnostics13061102.

[4] A. Moloud, U. R. Acharya, N. Sarrafzadegan, and V. Makarenkov, "NE-Nu-SVC: A New Nested Ensemble Clinical Decision Support System for Effective Diagnosis of Coronary Artery Disease," *IEEE Access,* vol.7, pp.167605–167620, 2019. doi:10.1109/ACCESS.2019.2953920.

[5] Jia, Guang, Xunan Huang, Sen Tao, Xianghuai Zhang, Yue Zhao, Hongcai Wang, Jie He, et al. 2022. "Artificial Intelligence-Based Medical Image Segmentation for 3D Printing and Naked Eye 3D Visualization." *Intelligent Medicine*, vol.2, pp.48–53, February 2022. doi:10.1016/j.imed.2021.04.001.

[6] C. Hongmin, S. Moon, J. Park, S. Bak, and Y. Ko et al, "The Use of Artificial Intelligence in Complementary and Alternative Medicine: A Systematic Scoping Review," *Frontiers in Pharmacology*, vol.13, April 2022. doi:10.3389/fphar.2022.826044.

[7] I. E. Salem and K. H. Al-Saedi, "A Sample Proposal Enhancing the Security of the Cloud Computing System Through Deep Learning and Data Mining," *Al-Salam Journal for Engineering and Technology*, vol.3, no.1,pp. 1–10. https://doi.org/10.55145/ajest.2024.03.01.001

[8] AH. Al-Mistarehi, M. M. Mijwil, Y. Filali, M. Bounabi, G. Ali, and M. Abotaleb, "Artificial Intelligence Solutions for Health 4.0: Overcoming Challenges and Surveying Applications," *Mesopotamian Journal of Artificial Intelligence in Healthcare*, vol.2023, pp.15–20. https://doi.org/10.58496/MJAIH/2023/003

[9] M. M. Mijwil, K. K. Hiran, R. Doshi, M. Dadhich, AH. Al-Mistarehi, and I. Bala, "ChatGPT and the Future of Academic Integrity in the Artificial Intelligence Era: A New Frontier," Al-Salam Journal for Engineering and Technology, vol. 2, no. 2, pp.116-127, April 2023. https://doi.org/10.55145/ajest.2023.02.02.015.

[10] M. M. Mijwil,G. Ali, and E. Sadıkoğlu, "The Evolving Role of Artificial Intelligence in the Future of Distance Learning: Exploring the Next Frontier," *Mesopotamian Journal of Computer Science*, vol.2023, pp.98-105, May 2023. https://doi.org/10.58496/MJCSC/2023/012