Heart disease prediction using support vector machine

Balakrishnan Duraisamy | Rakesh Sunku | Krithik Selvaraj | Vishnu Vardhan Reddy Pilla | Manoj Sanikala

Abstract Heart disease prediction through online consultation using machine learning refers to the application of advanced algorithms and techniques to analyze medical data collected during online consultations to predict the likelihood of an individual developing heart disease. Machine learning models are trained using historical data that includes various risk factors such as age, gender, blood pressure, cholesterol levels, and medical history. These models then utilize the input provided by patients during online consultations, such as symptoms, lifestyle habits, and additional medical tests, to generate personalized predictions about the probability of heart disease occurrence. By leveraging the power of machine learning, this approach aims to assist healthcare professionals in making more accurate diagnoses and providing timely recommendations for preventive measures or further medical intervention, ultimately improving patient outcomes and reducing the burden on healthcare systems. In this paper, a machine learning technique called Support Vector Machine (SVM) is used for heart disease prediction. Heart disease prediction through online consultation using SVM involves utilizing SVM as a machine learning algorithm to predict the likelihood of an individual having heart disease based on their consultation information provided online.

Keywords: healthcare system, heart disease prediction, deep learning, machine learning, SVM

1. Introduction

Predicting heart disease (Baviskar et al., 2023) solely through online consultation is challenging, as it typically requires a comprehensive evaluation involving medical tests, physical examinations, and access to a patient’s complete medical history. However, there are some general guidelines and risk factors that can be discussed during an online consultation to provide preliminary insights. Keep in mind that this should not replace a thorough evaluation by a healthcare professional, but it can help in identifying potential risks or symptoms that require further investigation.

During an online consultation, the following information and factors can be discussed:

1. Medical History: The healthcare provider can inquire about your personal and family medical history, including any previous heart conditions, diabetes, high blood pressure, high cholesterol, or other relevant health issues.
2. Symptoms: Discuss any symptoms you may be experiencing, such as chest pain or discomfort, shortness of breath, palpitations, fatigue, dizziness, or swelling. Describing the nature, duration, and severity of these symptoms can provide valuable information.
3. Lifestyle Factors: Discuss your lifestyle habits, including smoking, alcohol consumption, diet, and exercise routine. Unhealthy lifestyle choices can contribute to the development of heart disease.
4. Risk Factors: Certain risk factors can increase the likelihood of developing heart disease. These include age, gender (men are generally at higher risk), high blood pressure, high cholesterol levels, obesity, diabetes, family history of heart disease, and a sedentary lifestyle. It’s important to discuss these factors during the consultation.
5. Remote Monitoring: If anyone has access to wearable devices or remote monitoring tools that track vital signs like heart rate, blood pressure, or activity levels, sharing this information with the healthcare provider can provide additional insights.

However, it’s important to emphasize that online consultations have limitations, especially when it comes to the accuracy and comprehensiveness of the assessment. They should never replace in-person evaluations and diagnostic tests performed by qualified healthcare professionals. If you suspect or have concerns about heart disease, it’s advisable to consult with a healthcare provider who can guide you through the appropriate diagnostic procedures and recommend suitable treatment options.

Machine learning techniques (Pan et al., 2022; Alsaleh et al., 2023; Ali et al., 2021; Ouyang 2022) play a crucial role in heart disease prediction by enabling accurate risk assessment and personalized recommendations. These techniques...
analyze large volumes of data, including patient demographics, medical history, symptoms, and diagnostic test results, to uncover complex patterns and relationships. By training models on historical data, machine learning algorithms can identify relevant risk factors and predict the likelihood of heart disease occurrence. This assists healthcare professionals in making informed decisions, such as early intervention and tailored treatment plans, ultimately leading to improved patient outcomes. Machine learning empowers medical practitioners with efficient and reliable tools to assess heart disease risk, contributing to more effective prevention, diagnosis, and management strategies.

In this paper, a machine learning technique called Support Vector Machine (SVM) is introduced for heart prediction. SVM is particularly effective in cases where the data is well-separated and there is a need to find an optimal decision boundary. In heart disease prediction, SVM can analyze various patient-related factors such as age, gender, cholesterol levels, blood pressure, and other relevant clinical features. By learning from historical data, SVM can identify patterns and relationships between these factors and the presence or absence of heart disease. SVM’s ability to handle high-dimensional feature spaces and its capacity to capture non-linear relationships using kernel functions make it well-suited for complex heart disease prediction tasks. SVM-based models can provide accurate risk assessments and aid in early detection, allowing healthcare professionals to implement preventive measures and personalized treatment plans for patients at risk of heart disease.

A new approach was suggested, utilizing machine learning techniques to identify important features and enhance the accuracy of cardiovascular disease prediction (Mohan et al., 2019). The prediction model incorporates various feature combinations and established classification techniques. By employing the hybrid random forest with a linear model (HRFLM), a notable performance improvement was achieved, reaching an accuracy rate of 88.7% in heart disease prediction. However, HRFLM has high computational complexity.

A system based on a machine learning technique (Li et al., 2020) was proposed for the efficient and accurate diagnosis of heart disease. The system utilizes classification algorithms and incorporates a novel fast conditional mutual information feature selection algorithm to address the feature selection challenge. However, the prediction accuracy is still low.

A heart disease prediction model (HDPM) (Fitriyani et al. 2020) that proved to be successful was introduced for a Clinical Decision Support System (CDSS). This model incorporates Density-Based Spatial Clustering of Applications with Noise (DBSCAN) to identify and remove outliers, a combination of Synthetic Minority Over-sampling Technique-Edited Nearest Neighbor (SMOTE-ENN) to address imbalanced training data, and XGBoost for accurate heart disease prediction. However, SMOTE is that it can potentially introduce synthetic samples that are noisy or unrealistic.

To enhance prediction accuracy, hybrid classifiers (Ashri et al., 2021) was introduced by employing an ensembled model with majority voting. Additionally, a suggested preprocessing technique and feature selection, based on a genetic algorithm, aimed to improve prediction performance while minimizing time consumption. To address the issue of overfitting, the 10-folds cross-validation technique was utilized. However, the genetic algorithm has a premature convergence problem.

A framework called MalCaDD (Machine Learning based Cardiovascular Disease Diagnosis) (Rahim et al., 2021) was introduced to accurately predict cardiovascular diseases. The framework addresses missing values using the mean replacement technique and tackled data imbalance using the Synthetic Minority Over-sampling Technique (SMOTE). Furthermore, the framework incorporated the Feature Importance technique for selecting relevant features. Ultimately, the framework proposed an ensemble of Logistic Regression and K-Nearest Neighbor (KNN) classifiers to achieve higher prediction accuracy. However, KNN is sensitive to k values.

2. Materials and Methods

To create a full-fledged webpage that implements online doctor consultancy it constitutes appointment booking, video calling with the doctor, delivering of medicine from the nearby medical shop, connecting with the diagnostic centers, and heart disease prediction will be enabled for free. we must implement a full-fledged online doctor consultancy. If your health concerns are complex and your regular doctor believes they require urgent attention from a specialized physician, they may refer you to one. Alternatively, you can opt for an online consultation with a medical specialist. If you’re dissatisfied with the outcomes, you have the option to schedule an appointment with the doctor.

Creating an online room for patients and doctors has Audio, Video calling, and chatting features in virtual mode. The doctor can advise consulting the hospital physically. We are adding features like online consultation and physical mode to our project. If we have any previous health records you have to upload those records in our portal for the best understanding of the patient’s health condition. The doctor can able to view the previous documents related to the patient or issue new ones through a connection with the online diagnostic center. We cannot recommend or provide any consultancy for emergency patients.

2.1. Appointment Booking

There will be a home page that contains three logins. By using this log in the patient who is suffering can able to book an appointment with the doctor by choosing the domain of the doctor and symptoms, he can able to choose the date and time
and he can able to upload the previous documents in that. There will be two options given to the patient which he can able to meet the doctor in person or can occur virtually by using this website.

2.2. Video Call

After booking the appointment by the patient, the chat room will be created at that particular time and a link will be displayed to both the doctor and the patient. In which the doctor as well as the patient can able to interact through video calls and audio calls and he can able to chat by using the chat window.

2.3. Pharmacy Login

After analysis of the patient, If the doctor decided to give a prescription without any further diagnostics, the diction will have a box in which he can enter the names of the medicine and the quantity of the medicine. From then the prescription will be generated. Further, the prescription will be forwarded to the medical shop which contains the pharmacy login. Then medicines available in the medical shop will be filled and then the medicine will be packed and then it will be sent to the patient with the help of delivery boys the bill will be generated and he can able to pay money in cash or online.

2.4. Diagnostic Center

After the interaction with the doctor if the doctor decides to further diagnose the patient, then he will book the appointment with the diagnostic center the patient needs to go to the diagnostic center he needs to have the diagnostic center and the report will be generated and a copy will be sent to the doctor as well as the patient and he can further meet the doctor in an online and similar fashion will be continued.

2.5. Heart Disease Prediction

On the home page, there will be an option open for the patients in which they can able to check the probability of attacking heart diseases by which some of the parameters like bp, sugar, and etc. SVM is used for this prediction. SVM is a supervised machine learning algorithm used for classification tasks. It aims to find an optimal hyperplane that separates the data into different classes while maximizing the margin between the classes. In the case of heart disease prediction, we want to classify patients as having heart disease or not based on certain features.

Let’s consider a binary classification problem where we have a dataset with n training samples. Each sample is represented by a set of features xᵢ and is associated with a class label yᵢ (1 for heart disease, 0 for no heart disease). To find a hyperplane that separates these two classes.

The decision function of an SVM can be written as:

\[ f(x) = sign(w \cdot x + b) \quad (1) \]

Here, w is the weight vector and b is the bias term. The sign function returns either +1 or -1, indicating the predicted class of sample x.

To find the optimal hyperplane, SVM maximizes the margin between the two classes. The margin is the perpendicular distance between the hyperplane and the nearest training samples (support vectors). Let’s denote the positive support vector as x⁺ and the negative support vector as x⁻.

The margin is defined as:

\[ margin = \frac{2}{||w||} \quad (2) \]

where \( ||w|| \) is the Euclidean norm of the weight vector w. Maximizing the margin is equivalent to minimizing \( ||w||^2 \).

However, not all samples lie perfectly on one side of the hyperplane. To handle this, SVM introduces slack variables \( \xi_i \) for each training sample. These variables allow for a soft margin, allowing some samples to be misclassified or lie within the margin. The objective function for SVM can be written as:

\[
\begin{align*}
\text{minimize:} & \quad \frac{1}{2} ||w||^2 + C \sum \xi_i \\
\text{subject to:} & \quad y_i(w \cdot x_i + b) \geq 1 - \xi_i \text{ and } \xi_i \geq 0
\end{align*}
\]

(3)

Here, C is the regularization parameter that balances between maximizing the margin and allowing for misclassifications. It controls the trade-off between model complexity and training errors. Higher values of C result in a smaller margin but fewer misclassifications, while lower values of C allow for a larger margin but may lead to more misclassifications.

Solving this optimization problem using techniques such as quadratic programming or gradient descent yields the optimal values for w and b, which define the separating hyperplane. Once the model is trained, we can use it to whether a new patient has heart disease or not by evaluating the decision function f(x) for the patient’s feature vector x. If f(x) > 0, we predict the presence of heart disease; otherwise, the absence of heart disease is predicted.
3. Results

In this section, the performance of KNN and SVM for heart disease detection is evaluated in terms of accuracy, precision, and recall. For the experimental purpose, risk factors such as age, gender, blood pressure, cholesterol levels, and medical history are collected for different patients.

3.1. Accuracy

The percentage of occurrences that are successfully categorized as heart disease patients is known as accuracy. It is determined by dividing the total count of the presence of heart disease who were accurately predicted (true positive) by the total count of the absence of heart disease who were accurately predicted (true negative). It is determined as,

\[
\text{Accuracy} = \frac{\text{True Positive (TP)} + \text{False Negative (FN)}}{\text{TP} + \text{True Negative (TN)} + \text{False Positive (FP)} + \text{FN}} \quad (4)
\]

Table 1 shows the comparison of KNN and SVM in terms of accuracy for heart disease detection.

<table>
<thead>
<tr>
<th>Model</th>
<th>KNN</th>
<th>SVM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>86</td>
<td>89</td>
</tr>
</tbody>
</table>

Figure 1 shows the testing performance between KNN and SVM for heart disease prediction in terms of accuracy. The accuracy of SVM is 3.49% greater than KNN models. From this result, it is proved that the SVM has high accuracy than other models for heart disease prediction.

![Figure 1 Comparison of Accuracy]

3.2. Precision

Precision is the measure to find the capacity of the heart disease prediction model to recognize only the relevant instances in the dataset. It is calculated as,

\[
\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (5)
\]

Table 2 shows the comparison of KNN and SVM in terms of precision for heart disease detection.

<table>
<thead>
<tr>
<th>Model</th>
<th>KNN</th>
<th>SVM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>87</td>
<td>90</td>
</tr>
</tbody>
</table>

Figure 2 shows the testing performance between KNN and SVM for heart disease prediction in terms of precision. The precision of SVM is 3.45% greater than KNN models. From this result, it is proved that the SVM has high precision than other models for heart disease prediction.
3.3. Recall

Recall can measure the heart disease prediction method’s capacity to identify all the data instances of interest in a dataset. It is calculated as,

\[
\text{Recall} = \frac{TP}{TP + FN}
\]

Table 3 shows the comparison of KNN and SVM in terms of recall for heart disease detection.

<table>
<thead>
<tr>
<th>Model</th>
<th>KNN</th>
<th>SVM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>86.5</td>
<td>89.5</td>
</tr>
</tbody>
</table>

Figure 3 shows the testing performance between KNN and SVM for heart disease prediction in terms of recall. The recall of SVM is 3.47% greater than KNN models. From this result, it is proved that the SVM has high recall than other models for heart disease prediction.

4. Conclusions

In this paper, predicting heart disease through online consultation using Support Vector Machines (SVM) is performed. Heart disease prediction through online consultation using Support Vector Machines (SVM) involves developing a system that utilizes SVM algorithms to analyze medical data obtained during an online consultation and predict the likelihood of an individual having heart disease. The system takes into account various factors such as age, gender, blood pressure, cholesterol levels, and other relevant medical indicators. By training the SVM model on a dataset of pre-diagnosed cases, it learns the patterns and relationships between these features and the presence of heart disease. Once trained, the system can accept input from an online consultation, extract the relevant medical information, and feed it into the SVM model to obtain a prediction on the probability of heart disease. This approach enables remote diagnosis and risk assessment, potentially...
facilitating early intervention and improving patient outcomes. In the future, deep learning techniques such as Convolutional Neural Network (CNN) and ResNet50 will be used for heart disease prediction.

**Ethical considerations**

I confirm that I have obtained all consent required by the applicable law to publish any personal details of patients used.

**Conflict of Interest**

The authors declare no conflicts of interest.

**Funding**

This research did not receive any financial support.

**References**


