Heart Attack Prediction With Svm

Hanna Willa Dhany[1]*
Universitas Pembangunan Panca Budi Medan

Keywords:
Heart attacks, Support Vector Machine, age, gender, blood pressure, cholesterol levels, and medical history.

***Correspondence Address:**
hdhany@dosen.pancabudi.ac.id

Abstract: Heart attacks are one of the leading causes of death worldwide, making them an important focus in medical prevention and treatment efforts. Early detection of heart attack risk is crucial to provide timely medical intervention and prevent further complications. In this study, we used Support Vector Machine (SVM), one of the effective machine learning algorithms, to predict heart attacks. SVM works by finding the optimal hyperplane that separates data into different classes. We analyzed various risk factors such as age, gender, blood pressure, cholesterol levels, and medical history to predict the likelihood of a heart attack. The results showed that the developed SVM model had an accuracy rate of 91.80%, indicating that SVM can be a reliable prediction tool. This model is expected to help medical personnel make better decisions and provide more personalized care to high-risk patients, as well as contribute significantly to the treatment and prevention of heart attacks.

INTRODUCTION

Heart attacks are one of the leading causes of death globally, requiring serious attention in their prevention and management. Early detection of heart attack risks is crucial for timely medical intervention and preventing more severe complications. With technological advancements, machine learning methods have shown great potential in predicting various medical conditions, including heart attacks. Support Vector Machine (SVM) is one of the effective machine learning algorithms for classification and regression. SVM works by finding the optimal hyperplane that separates data into different classes. In heart attack prediction, SVM can be used to analyze patient medical data, such as age, gender, blood pressure, cholesterol levels, medical history, and other risk factors to predict the likelihood of a heart attack. This study aims to develop a heart attack prediction model using SVM that is expected to provide accurate and reliable predictions. With an accurate prediction model, it is hoped that medical professionals can make better decisions and provide more personalized care to high-risk patients.

Additionally, this study will discuss model validation and evaluate the performance of the heart attack prediction model using SVM.

RESEARCH METHODS

1. Data Collection

Collect data from medical records, including patient demographics, clinical history, lifestyle information, and laboratory results. Key variables include age, gender, blood pressure, cholesterol levels, smoking status, diabetes status, and family history of heart disease. Utilize hospital databases, electronic health records (EHRs), and publicly available datasets such as the UCI Machine Learning Repository.

2. Data Preprocessing

Remove or impute missing values, handle outliers, and correct any inconsistencies in the data. Normalize or standardize numerical features to ensure they are on a similar scale, which helps improve the performance of the SVM. Convert categorical variables into numerical values using techniques such as one-hot encoding or label encoding.

3. Feature Selection

Identify and select features that have a strong correlation with the target variable (heart attack occurrence). Reduce dimensionality by transforming features into a smaller set of uncorrelated variables while retaining most of the information. Iteratively remove less important features based on their impact on model performance.

4. Model Development

Divide the dataset into training and testing sets, typically using an 80-20 or 70-30 split. Use techniques such as Grid Search or Random Search to find the best hyperparameters for the SVM model, including the kernel type (linear, polynomial, radial basis function), C parameter (regularization), and gamma parameter. Train the SVM model on the training set using the selected features and optimized hyperparameters.

5. Model Evaluation

Evaluate the model's performance using metrics such as accuracy, precision, recall, F1-score, and the area under the receiver operating characteristic curve (AUC-ROC). Perform k-fold cross-validation to ensure the model's robustness and generalizability across different subsets of the data.

6. Model Validation and Testing

Use a validation set separate from the training and testing sets to fine-tune the model and avoid overfitting. Evaluate the final model's performance on the testing set to assess its predictive accuracy and reliability in real-world scenarios.

7. Implementation and Deployment

Integrate the trained SVM model into a clinical decision support system or a standalone application to assist healthcare providers in predicting heart attack risks. Develop a user-friendly interface for clinicians to input patient data and receive predictions and recommendations. Continuously monitor the model's performance and update it with new data to ensure its accuracy and relevance over time.

8. Ethical Considerations

Ensure compliance with data privacy regulations (e.g., HIPAA, GDPR) to protect patient confidentiality and data security. Address potential biases in the data and model to ensure fair and equitable predictions across different patient populations.

RESULTS AND DISCUSSION

The heart attack dataset is a comprehensive collection of medical data related to patients who have experienced heart attacks. This dataset is pivotal for researchers and healthcare professionals aiming to understand and predict heart attack risks. It includes a variety of features such as demographic information (age, gender), clinical data (blood pressure, cholesterol levels), and medical history (smoking status, diabetes, family history of heart disease).

With the advent of machine learning, this dataset provides a valuable resource for developing predictive models that can analyze patterns and identify risk factors associated with heart attacks. The insights gained from such models can facilitate early intervention, personalized treatment plans, and ultimately improve patient outcomes.

By leveraging the heart attack dataset, researchers can harness the power of data to uncover critical trends and correlations that may not be immediately apparent through traditional analysis. This, in turn, supports the advancement of medical knowledge and the enhancement of preventive measures against heart attacks.

Table 1 Heart Attack Dataset

	age	sex	cp	trtbps	chol	fbs	restecg	thalachh	exng	oldpeak	slp	caa	thall	output
0	63	1	3	145	233	1	0	150	0	2.3	0	0	1	1
1	37	1	2	130	250	0	1	187	0	3.5	0	0	2	1
2	41	0	1	130	204	0	0	172	0	1.4	2	0	2	1
3	56	1	1	120	236	0	1	178	0	0.8	2	0	2	1
4	57	0	0	120	354	0	1	163	1	0.6	2	0	2	1

About The Dataset :

age: Age of the patient

sex: Sex of the patient

cp: Chest pain type, 0 = Typical Angina, 1 = Atypical Angina, 2 = Non-anginal Pain, 3 = Asymptomatic

trtbps: Resting blood pressure (in mm Hg)

chol: Cholestoral in mg/dl fetched via BMI sensor

fbs: (fasting blood sugar > 120 mg/dl), 1 = True, 0 = False

restecg: Resting electrocardiographic results, 0 = Normal, 1 = ST-T wave normality, 2 = Left ventricular hypertrophy

thalachh: Maximum heart rate achieved

oldpeak: Previous peak

slp: Slope

caa: Number of major vessels

thall: Thalium Stress Test result ~ (0,3)

exng: Exercise induced angina ~ 1 = Yes, 0 = No

output: Target variable

Pair Plot is a graphical visualization used to see the relationship between several variables in a dataset. Pair Plot creates a scatter plot for each pair of variables and places them in a matrix form, allowing us to see how the two variables interact with each other. On the diagonal of the Pair Plot, there is usually a histogram or distribution of each variable. Pair Plot is very useful for Exploratory Data Analysis because it helps identify patterns, correlations, and potential anomalies in the data. If there are variables that have a strong linear or non-linear relationship, this will be clearly visible in the Pair Plot. Tools such as Seaborn in Python are often used to create Pair Plots easily.

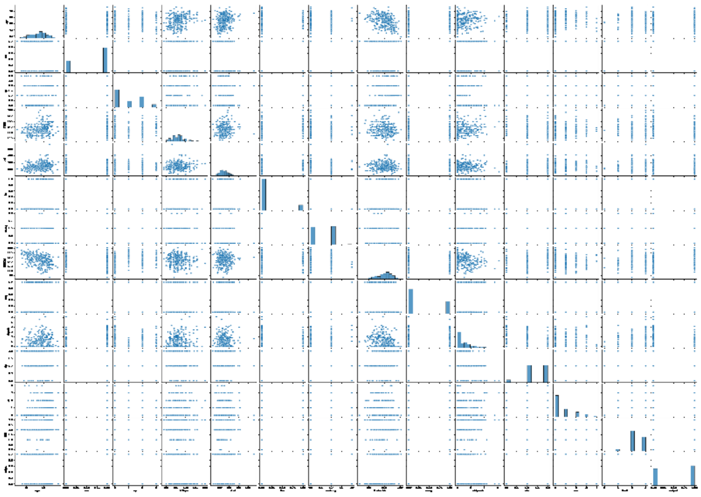

Figure 1 Pair Plot

CONCLUSION

This study shows that Support Vector Machine (SVM) is a very effective algorithm for heart attack prediction, with an accuracy of 91.80%. With a high level of accuracy, this SVM model can be used as a reliable tool for medical personnel to detect the risk of heart attacks in patients. This accurate prediction model allows for earlier and more precise medical interventions, as well as providing more personalized care to high-risk patients. The results of this study also underline the great potential of machine learning technology in improving the quality of diagnosis and treatment in the medical field. Thus, the implementation of the SVM model in clinical practice can contribute significantly to the management and prevention of heart attacks.

REFERENCE

A. Borji, A. Seif, and T. H. Hejazi, "An efcient method for detection of Alzheimer's disease using high-dimensional PET scan images," pp. 729–749, 2023.

Das, S.; Sharma, R.; Gourisaria, M.; Rautaray, S.; Pandey, M. Heart disease detection using core machine learning and deep learning techniques: A comparative study. Int. J. Emerg. Technol. 2020, 11, 531–538.

Deo, R.C. Machine learning in medicine. Circulation 2015, 132, 1920–1930.

Detrano, R.; Janosi, A.; Steinbrunn, W.; Pfisterer, M.; Schmid, J.J.; Sandhu, S.; Guppy, K.H.; Lee, S.; Froelicher, V. International application of a new probability algorithm for the diagnosis of coronary artery disease. Am. J. Cardiol. 1989, 64, 304–310.

Dinar, A.M.; Zain, A.M.; Salehuddin, F. Utilizing of CMOS ISFET sensors in DNA applications detection: A systematic review. J. Adv. Res. Dyn. Control Syst. 2018, 10, 569–583.

Diwakar, M.; Tripathi, A.; Joshi, K.; Memoria, M.; Singh, P. Latest trends on heart disease prediction using machine learning and image fusion. Mater. Today Proc. 2021, 37, 3213–3218.

Elhoseny, M.; Mohammed, M.A.; Mostafa, S.A.; Abdulkareem, K.H.; Maashi, M.S.; Garcia-Zapirain, B.; Mutlag, A.A.; Maashi, M.S. A new multi-agent feature wrapper machine learning approach for heart disease diagnosis. Comput. Mater. Contin 2021, 67, 51–71.

Gennari, J.H.; Langley, P.; Fisher, D. Models of incremental concept formation. Artif. Intell. 1989, 40, 11–61

Hasan, T.T.; Jasim, M.H.; Hashim, I.A. FPGA Design and Hardware Implementation of Heart Disease Diagnosis System Based on NVG-RAM Classifier. In Proceedings of the 2018 Third Scientific Conference of Electrical Engineering (SCEE), Baghdad, Iraq, 19–20 December 2018; pp. 33–38.

Hu, G.; Root, M.M. Building prediction models for coronary heart disease by synthesizing multiple longitudinal research findings. Eur. J. Prev. Cardiol. 2005, 12, 459–464. [Google Scholar] [CrossRef] [PubMed]

J. Ma, D. Lei, Z. Ren, C. Tan, D. Xia, and H. Guo, "Automated machine learning-based landslide susceptibility mapping for the three Gorges reservoir area, China," Mathematical Geo sciences, vol. 2, 2023.

Janosi, A.; Steinbrunn, W.; Pfisterer, M.; Detrano, R. UCI Machine Learning Repository: Heart Disease Dataset [Online]. Available online: https://archive-beta.ics.uci.edu/ml/datasets/heart+disease.

Javid, I.; Khalaf, A.; Ghazali, R. Enhanced accuracy of heart disease prediction using machine learning and recurrent neural networks ensemble majority voting method. Int. J. Adv. Comput. Sci. Appl. 2020, 11, 540–551.

Jawalkar, A. P., Swetcha, P., Manasvi, N., Sreekala, P., Aishwarya, S., Bhavani, K. D., & Anjani, P. (2023). Early prediction of heart disease with data analysis using supervised learning with stochastic gradient boosting. Journal of Engineering and Applied Science, 70, Article number: 122.

M. Rajabi and R. P. R. Hasanzadeh, "A modifed adaptive hysteresis smoothing approach for image denoising based on spatial domain redundancy," Sens Imaging, vol. 22, no. 1, p. 42, 2021.

M. Rajabi, H. Golshan, and R. P. Hasanzadeh, "Non-local adaptive hysteresis despeckling approach for medical ultra sound images," Biomedical Signal Processing and Control, vol. 85, Article ID 105042, 2023.

Mohammed, M.A.; Abdulkareem, K.H.; Al-Waisy, A.S.; Mostafa, S.A.; Al-Fahdawi, S.; Dinar, A.M.; Alhakami, W.; Abdullah, B.A.Z.; Al-Mhiqani, M.N.; Alhakami, H.; et al. Benchmarking methodology for selection of optimal COVID-19 diagnostic model based on entropy and TOPSIS methods. IEEE Access 2020, 8, 99115–99131.

Muhsen, D.K.; Khairi, T.W.A.; Alhamza, N.I.A. Machine Learning System Using Modified Random Forest Algorithm. In Intelligent Systems and Networks, Singapore; Tran, D.-T., Jeon, G., Nguyen, T.D.L., Lu, J., Xuan, T.-D., Eds.; Springer: Singapore, 2021; pp. 508–515.

Mythili, T.; Mukherji, D.; Padalia, N.; Naidu, A. A heart disease prediction model using SVM-decision trees-logistic regression (SDL). Int. J. Comput. Appl. 2013, 68, 0975–8887.

Nasser, A.R.; Hasan, A.M.; Humaidi, A.J.; Alkhayyat, A.; Alzubaidi, L.; Fadhel, M.A.; Santamaría, J.; Duan, Y. IoT and Cloud Computing in Health-Care: A New Wearable Device and Cloud-Based Deep Learning Algorithm for Monitoring of Diabetes. Electronics 2021, 10, 2719. Available online: https://www.mdpi.com/2079-9292/10/21/2719

Rahman, A.U.; Saeed, M.; Mohammed, M.A.; Jaber, M.M.; Garcia-Zapirain, B. A novel fuzzy parameterized fuzzy hypersoft set and riesz summability approach based decision support system for diagnosis of heart diseases.

Rahman, A.U.; Saeed, M.; Mohammed, M.A.; Krishnamoorthy, S.; Kadry, S.; Eid, F. An Integrated Algorithmic MADM Approach for Heart Diseases' Diagnosis Based on Neutrosophic Hypersoft Set with Possibility Degree-Based Setting. Life 2022, 12, 729.

Raniya R. Sarra, Ahmed M. Dinar, Mazin Abed Mohammed, & Karrar Hameed Abdulkareem. (2023). Enhanced Heart Disease Prediction Based on Machine Learning and χ2 Statistical Optimal Feature Selection Model. MDPI Journals, 6(5), 87-99.

Singh Rajpoot, V., Vishwakarma, S., Sahu, S., & Kesharwani, Y. (2024). Enhancing Heart Attack Prediction with Machine Learning. International Journal of Novel Research and Development (IJNRD), 9(6), 1-10.

Soni, M.; Gomathi, S.; Kumar, P.; Churi, P.P.; Mohammed, M.A.; Salman, A.O. Hybridizing Convolutional Neural Network for Classification of Lung Diseases. Int. J. Swarm Intell. Res. (IJSIR) 2022, 13, 1–15.

Wah, T.Y.; Mohammed, M.A.; Iqbal, U.; Kadry, S.; Majumdar, A.; Thinnukool, O. Novel DERMA fusion technique for ECG heartbeat classification. Life 2022, 12, 842.

Y. Long, W. Li, R. Huang, Q. Xu, B. Yu, and G. Liu, "A comparative study of supervised classifcation methods for investigating landslide evolution in the Mianyuan River Basin, China," Journal of Earth Sciences, vol. 34, no. 2, pp. 316 329,2023.

Z. Jia, J. Peng, Q. Lu et al., "A comprehensive method for the risk assessment of ground fssures: case study of the eastern weihe basin," Journal of Earth Sciences, vol. 34, no. 6, pp. 1892–1907, 2023.

Z. Liu, J. Ma, D. Xia et al., "Toward the reliable prediction of reservoir landslide displacement using earthworm optimi zation algorithm-optimized support vector regression (EOA SVR)," Natural Hazards, vol. 14, 2023.

Supiyandi, S., Iqbal, M., Purba, R. B., & Rizal, C. (2023). Development of logic gateway and network learning applications using augmented reality for computer architecture addie method curriculum. Prosiding universitas dharmawangsa, 3(1), 605–613

C. Rizal., Erni Marlina Saari. (2024). Leveraging Artificial Intelligence for Sustainable Software Maintenance: A Case Study Approach. *Proceedings The 2nd Annual Dharmawangsa International Conference*: "Digital Technology and Environmental Awareness in Promoting Sustainable Behavior In Society 5.0. vol. 1, no. 1; pp. 1-12.