Int. J. Advance Soft Compu. Appl, Vol. 18, No. 1, March 2026
Print ISSN: 2710-1274, Online ISSN: 2074-8523

Towards Personalized Lipid Management: Predicting Statins Therapy Eligibility through Machine Learning Models

Amal A. Alzu'bi[1], Eman R. Al Bataineh[1], Rasheed K. Ibdah[2], Rawan M. Shatnawi[3], Zaid F. Nassar[3], Leming Zhou[4]

[1] Department of Computer Information Systems, Jordan University of Science and Technology, Irbid, Jordan
e-mail: aazoubi9@just.edu.jo
[2] Department of Internal Medicine, Jordan University of Science and Technology, Irbid, Jordan
[3] King Abdullah University Hospital, Irbid, Jordan
[4] Department of Health Information Management, University of Pittsburgh, Pittsburgh, PA, USA
Corresponding author: Amal A. Alzu'bi, PhD — aazoubi9@just.edu.jo

Abstract

Cardiovascular diseases are among the leading causes of death worldwide and a major contributor to the deterioration of quality of life. Therefore, it is highly beneficial to follow the clinical guidelines and recommendations for preventing and treating cardiovascular diseases at their early stages. Cholesterol-lowering drugs such as Statins are considered first-line medications for the prevention of atherosclerotic cardiovascular diseases (ASCVD). However, it is not easy to determine patients' eligibility for statin therapy. In this work, we built efficient and accurate prediction models based on several machine learning algorithms for predicting patients' eligibility for Statins using several cardiovascular disease risk factors. The results indicated that the gradient boosting classifier achieved 95.6% accuracy and 99.0% area under the curve in predicting patients' eligibility for statin therapy. Other simpler but more explainable algorithms such as decision tree and logistic regression also demonstrated good performance.

Keywords: *Statin therapy eligibility, Predictive machine learning, Classification.*

1 Introduction

Cardiovascular diseases (CVDs) are among the leading causes of death globally and a major contributor to reduced quality of life [1-3]. Approximately 17.9 million people died because of cardiovascular diseases in 2019, which is around 32% of all global deaths. 85% of those deaths were due to heart attacks and strokes [4].

CVDs represent a group of cardiovascular disorders that include coronary heart disease, cerebrovascular disease, and other associated disorders. In fact, it is estimated that by 2030, cardiovascular diseases will be the top cause of death in the world's poorest countries [4]-[6]. In Jordan, noncommunicable diseases (NCDs) impose a significant health burden, accounting for over 80% of all deaths [4]. Cardiovascular disorders, one of the most

Received 6 November 2025; Accepted 10 February 2026

common NCDs, account for 42% of all deaths. According to the statistics from Jordan's national Stepwise survey for NCD risk factors in 2019, the prevalence of hypertension is 52% of diabetes is 20%, and elevated risk of cardiovascular disease is 25% among persons aged 45-69 years old [4].

Myocardial infarction (commonly called heart attack) is a potentially fatal disease caused by a shortage of blood flow to heart muscle. A lack of blood flow can be caused by a variety of circumstances, but it is most commonly results from a blockage in one or multiple coronary arteries that leads to cardiac muscle death if blood flow is not restored [7]- [9]. Myocardial infarction syndrome is one of the most serious cardiac diseases that affects morbidity and mortality worldwide. Studies showed that more than 3 million people die from acute ST-elevation myocardial infarction (STEMI) while another 4 million die from non-ST-elevation myocardial infarction (NSTEMI) every year [8].

The major behavioral risk factors for myocardial infarction, stroke, and heart failure, include unhealthy foods, physical inactivity, smoking, and alcohol consumption. Consequently, individuals may experience elevated blood pressure, glucose, lipids, and weight that can be assessed in primary care settings [10]- [13]. Identifying people at high risk of cardiovascular diseases and ensuring that they receive adequate therapy can prevent premature deaths and reduce economic burden globally and is particularly beneficial for low- and middle-income countries.

Cholesterol-lowering drugs (Statins) are the first line medications for the prevention of atherosclerotic cardiovascular disease (ASCVD) [14]. Statin therapy is also the cornerstone for controlling high cholesterol levels and has been proven to be able to reduce the risk of cardiovascular diseases [15]- [16]. In 2013 and 2018, the American College of Cardiology (ACC) and the American Heart Association (AHA) published a list of recommendations describing statins eligibility and dosage for managing CVD risk in adults [17]. Recommendations for high- and moderate-intensity statin therapy have been proposed for use in the primary and secondary prevention of CVD [17]. The United States Preventive Services Task Force (USPSTF) recommended statin therapy as the primary prevention of ASCVD in 2016 [15]. These recommendations suggested starting statins therapy for adults aged between 40 to 75 and have one or more risk factors for ASCVD including high blood pressure, tobacco use, diabetes mellitus, dyslipidemia, and calculated 10-year CVD event risk 10% or greater [15]. It seems that these recommendations can be followed in clinical practice. However, the actual situation is far more complicated than that.

2 Related Work

Based on 2013 ACC/AHA recommendations, data from a more recent National Health and Nutrition Examination Surveys (2007–2012) were used to evaluate Statins use among persons aged 21–79. According to the study, 25.5% of survey participants in the aforementioned age group were qualified for statin therapy [18, 19]. However, even if patients received the recommendation from their doctors and started to use Statins, there is a medication adherence issue. After all, patients do not feel any obvious health improvement from statins, on the contrary, some patients may even have side effects such as muscle pain, digestive issues, headaches, and dizziness. Several studies indicated that health outcomes can be even worse if patients choose to stop taking statins. In a multiethnic study of 347,104 eligible adults with ASCVD who had stable statins prescriptions, researchers found that low adherence to Statins treatment was associated with an increased risk of death [20]. De Vera et al. conducted a systematic review to compile the current

evidence on the effects of Statins adherence, discontinuation, and continuation on cardiovascular disease outcomes and mortality [21]. They also found an increased risk of adverse outcomes associated with poor Statins adherence.
Taylor et al. conducted a study of 34,272 participants to evaluate the effects, harms, and benefits of Statins in people without a history of cardiovascular diseases [1]. Only limited evidence has shown that primary prevention with Statins may be cost-effective and can improve the quality of life of patients. They found that several cautions should be taken when prescribing Statins as primary prevention for people at risk of cardiovascular diseases.
Thavendiranathan et al. conducted trials with 42,848 patients to investigate the effect of Statins [22]. In that study, 90% of patients had no history of cardiovascular disease. They found that the treatment with Statins in patients without cardiovascular disease could reduce the incidence of major coronary and cerebrovascular events, and vascular reconstruction, but not coronary heart disease or overall mortality.
Another study has demonstrated that the high intensity Statins, atorvastatin 80 mg and rosuvastatin 20 mg daily, can reduce ASCVD events and low-density lipoprotein (LDL) cholesterol by an average of 50% [23].
According to the new guidelines from ACC/AHA, a population modeling study by Yang et al. showed that up to 12.6% of annual ASCVD deaths could be avoided if the eligible patients for ASCVD primary prevention received Statins [24].
In summary, there are inconsistent findings in the literature related to the use of Statins for cardiovascular disease prevention. Therefore, it is not trivial to determine the statins eligibility even with the availability of detailed guidelines from the ACC/AHA. Physicians need to take into account many different factors and those factors are not equally important. Hence, the clinical decision process becomes highly subjective. Therefore, physicians often have difficulties in deciding statins prescription for prevention purposes. The consequence is that many people who are eligible for statins miss the opportunity of preventing ASCVD from happening or recurring.
In this study, our aim is to identify all eligible patients who should receive statins for secondary prevention. We seek to determine factors that can predict the eligibility for statins in order to prevent cardiovascular diseases recurring. This may provide an objective approach for Statins prescription and provide assistance to physicians in their decision making.

3 The Proposed Method

3.1 Overview of Our Approach

The workflow of our work starts with data collection, data cleaning and preparation, then divide the data into a training set (80%) and a test set (20%), after that we had more data preprocessing steps such as missing data imputation, feature engineering, and data scaling, followed by feature selection. At last, we applied multiple traditional machine learning algorithms on the preprocessed dataset and performed result evaluation. In the following subsections, we provide further details of these steps.

3.2 Data Collection

In total, 1,500 patient records were collected from the King Abdullah University Hospital (KAUH), which is the largest university hospital in Northern Jordan and serves more than

one million patients from Irbid, Jerash, Ajloun, and Mafraq. It is a referral hospital for cardiac cases. Patients admitted from January 2022 to March 2023 were evaluated for inclusion in our study. The dataset was collected and labeled manually by specialists in the Cardiology Department at KAUH in Irbid, Northern Jordan.
Inclusion criteria: 1) adult patients admitted from January 2022 to March 2023 in KAUH; 2) patients had acute myocardial infarction; 3) patients had never used statin therapy in the past.
Exclusion criteria: 1) patients with incomplete medical records or substantial missing clinical or laboratory data.; 2) duplicate patient entries within the dataset; 3) patients with a previous history of statin therapy prior to data collection.
Our study received ethical approval from KAUH Institutional Review Board (Approval # 7/154/2023) on February 19, 2023.

3.3 Data Overview

The collected data consists of 13 factors/features and one class label (eligible/ineligible). The features include age, gender, body mass index (BMI), low-density lipoprotein (LDL), high-density lipoprotein (HDL), hypertension, total cholesterol (TC), diabetes (DM), smoking history, family history of coronary artery disease (CAD), 10-year risk of ASCVD, triglycerides (TG), and acute myocardial infarction (AMI). In the dataset, there is also a feature named "statin therapy", which is an indicator showing whether the patient has received statin therapy in the past. For all patients included in this study, the value for this particular feature is "No" (or 0). It is not informative for machine learning; therefore, it is not included in the machine learning pipeline.

3.4 Data Cleaning and Preprocessing

Data cleaning and preprocessing is a time-consuming but very important step in data analysis and machine learning. We have gone through multiple steps to prepare the data before building the models, most notably:

- Report Integration: We integrated multiple reports related to the same patients into an individual report based on the patient IDs and the date of report.
- Data Cleaning: We cleaned our dataset by removing records with high missingness (> 20% missing), removing noisy data, resolving inconsistencies, and removing outliers.
- Duplicated Records: We removed all the duplicated records to ensure that each patient had only one record in the dataset.
- Statin Therapy: Patients who have already received statin therapy in the past were not eligible for this study. Therefore, we removed all the records for patients who previously received statin therapy.
- Lipid Profile: We removed records that did not have a measured lipid profile.
- 10-Year Risk of ASCVD: We calculated a 10-year risk of ASCVD score using a set of clinical and laboratory variables. To estimate the risk for individuals with value(s) beyond these ranges, the values were modified to be equivalent to the minimum or maximum value of that variable. For example, a cholesterol value of 340 mg/dL was approximated to 320 mg/dL.
- Statin Eligibility and Dosage: According to the 2018 ACC/AHA guidelines, cardiologists at KAUH determined the eligibility of statin therapy. They discussed every single case for which they had different labels during the independent labeling process and reached agreement on all of them. For eligibility, there were

two categories: eligible and ineligible. In the dataset, we used 0 for ineligible and 1 for eligible. These are the class labels for our binary classification.

The following two steps were performed after the data set was split into training and test sets.

- Missing Data Imputation: After data cleaning, for the remaining missing data, we performed imputation. For categorical variables, we used the mode of the variable to replace the missing data. For continuous variables, we used the mean of K-nearest neighbors to calculate the missing values.
- Scaling and Data Transformation: Some machine learning algorithms are sensitive to the scaling and distribute of feature data while other algorithms are not. To make the machine learning results comparable, it is important to make the scaling of features consistent and make the distribution of continuous variables closer to a normal distribution. For this purpose, we first checked the distribution of all continuous variables. For those with highly skewed distributions, we tested multiple mathematical functions (e.g., log, sqrt, square, exp) for data transformation so that the resulting data would be closer to a normal distribution. We then applied the StandardScaler method offered in the Scikit-Learn Python package on features with a wide range of values. All transformed features have a mean of zero and standard deviation of one.

3.5 Feature Selection

We used information gain to prioritize and rank all the features in the dataset. This method can measure the value of each attribute based on the amount of information that we can receive from that attribute in relation to the class label. The features with importance greater than 0 were kept for further analysis.

We calculated Pearson correlation among continuous features. If any pair of features had a correlation greater than 0.90, the one with lower rank was removed from the dataset. The remaining important features identified by these two steps were used in the machine learning models.

3.6 Classification Algorithms and Result Evaluation

We applied traditional machine learning algorithms on the preprocessed dataset with 10-fold cross validation. These algorithms included: K-Nearest Neighbor (KNN), Logistic Regression (LR), Decision Tree (DT), Support Vector Machine (SVM), Random Forest (RF), Naïve Bayes (NB), Ada Boost, Gradient Boosting, and Neural Networks (NN). We evaluated these algorithms based on commonly used performance measurements including accuracy, precision, recall, F-1 score, area under the curve (AUC) of a receiver operating characteristic (ROC) curve. The models were trained on the training dataset and were evaluated on the test set, which has never been used during model training. We also considered the explainability of those algorithms when interpreting the results.

4 Results

4.1 Descriptive Statistics

Our preprocessed dataset includes 800 patients who were admitted to the Cardiology Department at the King Abdullah University Hospital from January 2022 to March 2023. Their descriptive statistics are presented in Table 1 and described below.

Among all cases, there were 564 males and 236 females, representing 70.5% and 29.5% of all included patients, respectively. Out of the 800 patients, 511 (63.9%) were eligible to statin therapy. The remaining 289 (36.1%) were not eligible to statin therapy.

The number of patients with diabetes was 318 and the number without diabetes was 482, representing 39.8% and 60.2%, respectively. The study population had approximately equal proportions of smokers and non-smokers, 413 smokers and 387 non-smokers, or 51.6% and 48.4% of the selected patients, respectively. The number of patients with hypertension was 441 and the number without hypertension was 359, representing 55.1% and 44.9% of the patient population, respectively. All patients in this study had acute myocardial infarction and they were classified into three groups including Unspecified, where they constituted approximately two-thirds of the study population with total number of 540, NSTEMI with a total number of 226, and STEMI with a total number of 34, with percentages of 67.5%, 28.2%, and 4.3%, respectively. Most of the study participants (625, 78.1%) had a first-degree family history of cardiovascular disease, while those who did not have a family history of cardiovascular disease numbered 175 (21.9%).

Table 1. Descriptive statistics of patients included in this study.

N = 800	**Statistics**	
	n	***%***
Sex		
Male	564	70.5%
Female	236	29.5%
Statin Eligibility		
Eligible	511	63.9%
Ineligible	289	36.1%
Diabetes		
Yes	318	39.8%
No	482	60.2%
Smoker		
Yes	413	51.6%
No	387	48.4%
Hypertension		
Yes	441	55.1%
No	359	44.9%
Acute myocardial infarction		
Unspecified	540	67.5%
NSTEMI	226	28.2%
STEMI	34	4.3%
Family History of CVD		
Yes	625	78.1%
No	175	21.9%

4.2 Feature Importance

Figure 1 shows the rank of each feature in the dataset using the method of information gain. It clearly indicates that the 10-years risk of ASCVD is the most important feature, followed by diabetes, age, HDL, hypertension, AMI, TC, TG, and smoking.

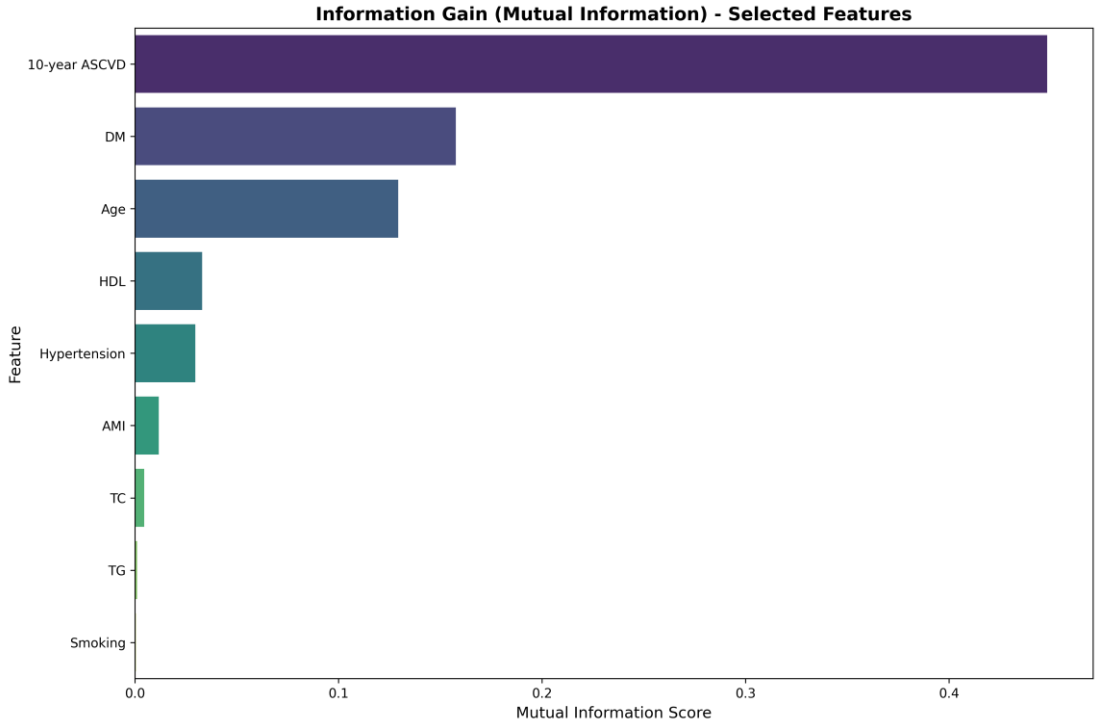

Fig. 1. Feature importance from information gain. Longer bars mean more information gain, also higher feature importance.

4.3 Results of the Statin Therapy Eligibility Experiment

There were 640 patients in the training set (409 eligible and 231 ineligibles for statin therapy) and 160 patients in the test set (102 eligible and 58 ineligible).
The classification results for predicting statin therapy eligibility on the test set using traditional machine learning algorithms are presented in Table 2. Among the nine algorithms tested in this experiment, several algorithms achieved great results based on the evaluation metrics in the 10-fold cross-validation. The Gradient Boosting algorithm achieved the highest accuracy rate of 95.6%, with precision, recall, and F1 score of 96.1%, 97.1%, and 96.6%, respectively. A few other models, such as random forest, Ada Boost, and SVM achieved comparable prediction performance on the test set as well. The prediction performance of logistic regression and neural networks is just slightly worse than that of those aforementioned models.

Table 2. Results of statin eligibility prediction on test set from traditional machine learning algorithms.

Model	**Accuracy**	**Precision**	**Recall**	**F1 Score**	**ROC AUC**
Random Forest	94.4%	95.1%	96.1%	95.6%	99.0%
Gradient Boosting	95.6%	96.1%	97.1%	96.6%	99.0%
Ada Boost	95.0%	97.0%	95.1%	96.0%	98.9%

Decision Tree	95.6%	95.2%	98.0%	96.6%	97.1%
Logistic Regression	92.5%	92.5%	96.1%	94.2%	97.0%
Neural Network	92.5%	94.1%	94.1%	94.1%	96.5%
Naive Bayes	86.9%	94.5%	84.3%	89.1%	94.8%
KNN	90.6%	91.4%	94.1%	92.8%	94.8%
SVM	93.8%	94.2%	96.1%	95.1%	93.8%

Fig. 2 demonstrates the confusion matrices for decision tree and logistic regression. In the test set, there were 102 patients eligible and 58 patients ineligible for statin therapy. The decision tree model made two incorrect predictions on eligible patients and five incorrect predictions on ineligible patients. The logistic regression model made four incorrect predictions on eligible patients and eight incorrect predictions on ineligible patients. In other words, the predictive performance of these models is relatively worse when applied to ineligible patients.

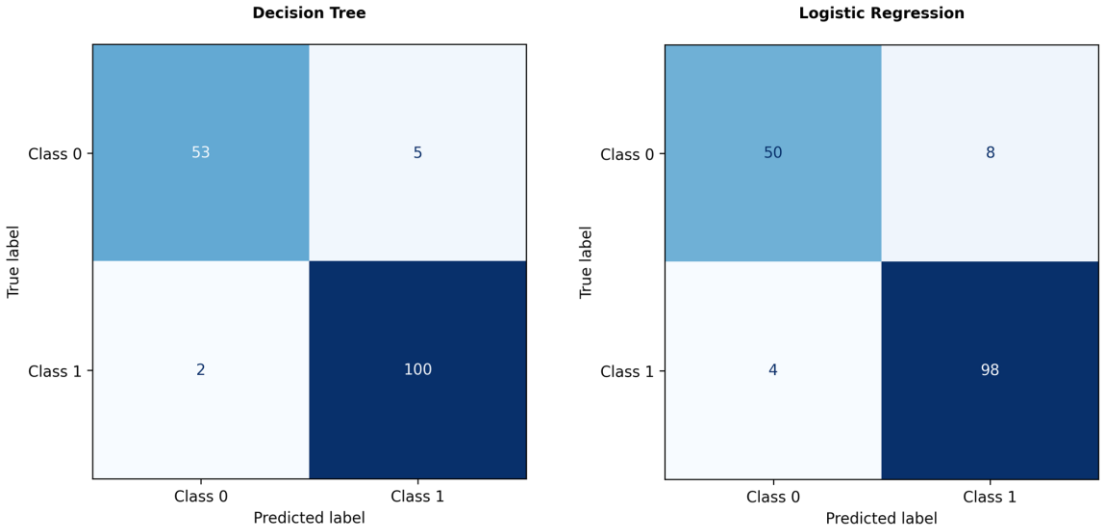

Fig. 2. Confusion matrices from test set for decision tree and logistic regression models.
Models such as logistic regression and decision tree had very good performance and excellent explainability. In clinical applications, models with high explainability are preferred.

5 Discussion

In this study, we developed machine learning models to predict statin therapy eligibility for secondary prevention in patients who had experienced acute myocardial infarction. Our focus on secondary prevention—preventing recurrent cardiovascular events in patients with established cardiovascular disease—differs from primary prevention efforts aimed at preventing initial events in at-risk individuals. Our approach achieved strong predictive performance, with Gradient Boosting reaching 95.6% accuracy, 96.1% precision, and 97.1% recall for eligibility prediction. While these results demonstrate the potential of machine learning for clinical decision support in cardiovascular disease prevention, they require careful interpretation within the context of existing literature and clinical practice.

5.1 Performance in Context of Existing Literature

Our 95.6% accuracy and 99.0% AUC for predicting statin therapy eligibility in secondary prevention exceed performance metrics reported in similar studies, though with important differences in study design and objectives.

Sarraju et al. developed machine learning models to estimate 5-year CVD event risk in multiethnic patients with established cardiovascular disease, achieving AUC values of 0.70-0.71 [26]. The study included 32,192 patients from a large health system with highly diverse clinical characteristics, which are known to reduce the prediction power of machine learning models.

For station-related predictions, recent studies have focused on different but related outcomes. Xiong et al. developed machine learning models to predict statin efficacy and safety, achieving AUC values of 0.883 for efficacy prediction, AUC values of 0.964 for liver enzyme abnormalities, and AUC values of 0.981 for muscle pain/creatine kinase abnormalities [27]. All patients in the study were hospitalized at two Chinese hospitals, and all had a history of statin use. A study by Han et al. predicted low density lipoprotein cholesterol (LDL-C) target attainment in patients with coronary artery disease receiving moderate-dose statins, achieving an average AUC of 0.695 across multiple machine learning models [28]. These studies focused on treatment response prediction rather than eligibility determination, making direct comparison challenging.

A review by Wang et al. found that while machine learning models showed promise for cardiovascular risk assessment, most achieved AUC values between 0.75-0.85 when externally validated, with performance degradation common when models were applied to new populations [29].

The higher performance in our study compared to these recent benchmarks likely reflects several factors. First, our models predict guideline-concordant prescribing decisions for secondary prevention rather than actual clinical outcomes or primary prevention decisions. Secondary prevention decisions tend to be more standardized than primary prevention decisions, as most patients with acute myocardial infarction have clear indications for statin therapy based on current guidelines. Second, the single-center nature of our data may indicate highly consistent guideline implementation and prescribing patterns within our institution, making predictions more deterministic. Third, our feature set includes the 10-year ASCVD risk score, which is itself a composite risk calculation that already incorporates multiple cardiovascular risk factors and is central to guideline-based decision-making. The strong performance may therefore reflect our model's ability to learn institution-specific interpretations of guidelines rather than discovering novel risk relationships.

Importantly, our feature importance analysis aligns well with established clinical guidelines for secondary prevention. The prominence of 10-year ASCVD risk, diabetes, hypertension, and age as key predictive features corresponds directly to the 2018 ACC/AHA cholesterol management guidelines [30]. This concordance provides face validity for our approach and suggests that our models are learning clinically meaningful patterns rather than spurious correlations.

5.2 Clinical Utility Beyond Workflow Efficiency

While reducing physician time and effort represents one benefit of machine learning for automated decision support, the clinical value of this work for secondary prevention extends to several other important domains.

First, this work might improve guideline adherence in secondary prevention, which remains suboptimal globally. Automated risk assessment and eligibility determination could help identify patients who would benefit from therapy initiation.

Second, work similar to ours can help to reduce practice variation by standardizing risk factor assessment and ensuring comprehensive evaluation of all relevant variables. Even in secondary prevention, where indications are generally clearer than in primary prevention, variability exists in determining appropriate statin intensity. Manual calculation of risk scores and consideration of multiple factors can be time-consuming and subject to individual clinician interpretation. Automated systems could ensure that clinical decisions are made systematically based on all available risk information.

Third, such systems may facilitate shared decision-making between patients and physicians. Following acute myocardial infarction, patients face decisions about long-term medical therapy, with trade-offs between risk reduction and medication burden, cost, and potential side effects. Decision support tools can help structure these conversations by providing clear, consistent risk estimates and treatment recommendations tailored to individual patient profiles.

Fourth, these tools could be particularly valuable in resource-limited settings or healthcare systems with limited access to cardiology specialists. Standardized, guideline-based decision support could help ensure that patients with acute myocardial infarction receive appropriate secondary prevention therapy regardless of where they receive care.

However, the clinical utility of our models depends critically on their ability to generalize beyond our training environment. Models that simply replicate existing institutional prescribing patterns offer limited value beyond automation. True clinical utility requires that models either improve upon current decision-making or successfully transfer their performance to new settings where guideline adherence may be lower or where patient populations differ.

5.3 Model Explainability and Clinical Adoption

Our finding that simpler algorithms such as decision trees and logistic regression achieved strong performance while maintaining greater explainability is particularly relevant for clinical implementation. The "black box" nature of more complex machine learning models remains a significant barrier to clinical adoption, as physicians are understandably reluctant to rely on recommendations they cannot understand or verify. Logistic regression, in particular, provides interpretable coefficients that directly correspond to clinical reasoning about risk factors and can be easily communicated to patients.

The ability to explain model predictions is not merely a matter of physician comfort—it has important implications for patient safety, regulatory compliance, and medical-legal considerations. Explainable models allow clinicians to evaluate whether recommendations make clinical sense for individual patients and to override automated suggestions when clinical circumstances warrant deviation from standard guidelines. This human-in-the-loop approach is essential for safe implementation of clinical AI systems, particularly in secondary prevention where patient-specific factors such as frailty, comorbidities, life expectancy, and treatment preferences may justify alternative approaches.

5.4 Limitations and Considerations

Several important limitations must be acknowledged. First and most significantly, our dataset of 800 patients from a single hospital in Northern Jordan limits generalizability. Regional differences in patient populations, healthcare systems, guideline implementation, and prescribing practices mean our model may not transfer to other settings. Single-center studies are particularly prone to capturing institution-specific patterns that do not represent broader clinical practice. The homogeneity of prescribing patterns within a single institution may also contribute to our high-performance metrics. Studies have shown substantial international variation in secondary prevention practices, with the PURE study documenting marked differences in statin use for secondary prevention by socioeconomic status and geographic region [31].

Second, we lack external validation on data from different hospitals, regions, or countries. Internal cross-validation provides optimistic estimates of model performance, and true generalizability can only be assessed through testing on independent datasets. This is particularly important given that our models predict guideline-based decisions; if guidelines are interpreted or applied differently in other settings, model performance may degrade substantially. Recent studies have highlighted the importance of external validation, with many models showing significant performance degradation when applied to new populations or healthcare systems.

Third, our relatively small sample size of 800 patients may limit the models' ability to capture rare but clinically important scenarios. While this sample size is sufficient for initial model development, larger datasets would be needed to ensure robust performance across diverse patient subgroups.

Fourth, our study does not address several clinically relevant scenarios that complicate real-world statin prescribing decisions in secondary prevention. These include: patients with contraindications to statins (such as active liver disease or pregnancy), patients who experience statin-associated muscle symptoms requiring alternative approaches, patients who decline therapy despite being eligible based on informed preference, patients with multiple comorbidities that may alter risk-benefit calculations, and considerations of frailty or limited life expectancy that may modify treatment intensity decisions. Our models assume guideline-based decisions represent optimal care, but clinical practice appropriately involves individualized decision-making that may deviate from guidelines in justified circumstances.

Fifth, our dataset included only patients admitted to a cardiology department with acute myocardial infarction, representing a specific population with acute cardiovascular events requiring secondary prevention. This differs from the broader population of patients with established cardiovascular disease who may be managed in primary care or other settings. The acute care context, with readily available comprehensive cardiovascular risk assessment, may make decision-making more straightforward than in outpatient settings where information may be incomplete.

5.5 Future Directions

To address these limitations and advance toward clinical implementation, several steps are needed. First, we plan to collect multi-institutional datasets from diverse geographic regions and healthcare settings for external validation. This will provide a realistic assessment of model generalizability and identify factors that affect model performance

across different environments. Collaboration with other hospitals in Jordan and the broader Middle East region would be particularly valuable.

Second, prospective evaluation comparing model recommendations to actual clinical decisions with detailed chart review of discordant cases would help identify scenarios where models perform well versus situations requiring human judgment. This could inform appropriate use cases and help define the scope of automated decision support for secondary prevention.

Third, we will evaluate model performance in outpatient settings where most long-term secondary prevention management occurs. While our study focused on acute inpatient decisions, the majority of secondary prevention care happens in follow-up visits where decision support could have substantial impact on medication adherence, intensity titration, and management of side effects.

Fourth, incorporation of patient-reported outcomes, preferences, and values into the decision framework could move beyond simple eligibility prediction toward more nuanced shared decision-making support. This would acknowledge that optimal secondary prevention involves more than guideline concordance—it requires alignment with patient goals, tolerance of medications, and individual risk-benefit assessments.

Fifth, longitudinal studies examining actual clinical outcomes (recurrent myocardial infarction, stroke, cardiovascular mortality) in relation to model recommendations would provide crucial validation of clinical utility. This would require following patients over time and comparing outcomes between those whose treatment followed model recommendations versus those whose treatment differed.

Finally, implementation studies examining how such tools affect physician decision-making, guideline adherence, patient outcomes, and healthcare efficiency in real clinical workflows are essential. The value of clinical AI systems ultimately depends on their impact when deployed in actual practice, not just their performance in controlled validation studies. Studies should also examine potential unintended consequences, such as automation bias or reduced clinical reasoning, that may arise from over-reliance on automated systems.

6 Conclusion

Our machine learning models demonstrate strong performance in predicting statin eligibility for secondary prevention in patients with acute myocardial infarction, with Gradient Boosting achieving 95.6% accuracy. The identification of key predictive features aligns well with established clinical guidelines for secondary prevention, supporting the clinical validity of our approach. Simpler, more explainable models such as logistic regression also achieved strong performance, which may facilitate clinical adoption by providing transparent, interpretable predictions. However, the high-performance metrics must be interpreted cautiously given our single-center dataset, relatively small sample size, and lack of external validation.

The distinction between our secondary prevention focusses and most literature on primary prevention is important, as secondary prevention decisions tend to be more standardized based on current guidelines. The true value of this work lies not merely in replicating existing decisions efficiently, but in improving clinical outcomes through better risk stratification, enhanced guideline adherence, support for personalized treatment decisions, and facilitation of shared decision-making between patients and providers. Rigorous external validation across diverse healthcare settings, prospective evaluation with long-

term outcome data, and careful implementation studies are essential next steps before clinical deployment in secondary prevention practice.

Data Availability

The data used in this study are available upon reasonable request, as the dataset is private

References

[1] F. Taylor, K. Ward, T. H. M. Moore, M. Burke, G. D. Smith, J. P. Casas, et al., "Statins for the primary prevention of cardiovascular disease," Cochrane Database of Systematic Reviews, 2011(1).

[2] Y. Zhao, E. P. Wood, N. Mirin, S. H. Cook, R. Chunara, "Social Determinants in Machine Learning Cardiovascular Disease Prediction Models: A Systematic Review," American Journal of Preventive Medicine, vol. 61, no. 4, pp. 596-605, 2021, doi: 10.1016/j.amepre.2021.04.016.

[3] K. Tsarapatsani, A. I. Sakellarios, V. C. Pezoulas, V. D. Tsakanikas, M. E. Kleber, W. Marz, et al., "Machine Learning Models for Cardiovascular Disease Events Prediction," in Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2022, pp. 1066-1069.

[4] World Health Organization (WHO), "Fact sheets on cardiovascular diseases," [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds).

[5] G. D. Flora, M. K. Nayak, "A brief review of cardiovascular diseases, associated risk factors and current treatment regimes," Current Pharmaceutical Design, vol. 25, no. 38, pp. 4063-4084, 2019.

[6] S. Robinson, Cardiovascular Disease. Priorities for Health Promotion and Public Health: Routledge, 2021, pp. 355-393.

[7] M. Saleh, J. A. Ambrose, "Understanding myocardial infarction," F1000Research, vol. 7, 2018.

[8] G. W. Reed, J. E. Rossi, C. P. Cannon, "Acute myocardial infarction," The Lancet, vol. 389, no. 10065, pp. 197-210, 2017.

[9] D. Gabriel-Costa, "The pathophysiology of myocardial infarction-induced heart failure," Pathophysiology, vol. 25, no. 4, pp. 277-284, 2018.

[10] J. C. Brown, T. E. Gerhardt, E. Kwon, "Risk factors for coronary artery disease," 2020.

[11] M. Jokela, L. Pulkki-Råback, M. Elovainio, M. Kivimäki, "Personality traits as risk factors for stroke and coronary heart disease mortality: pooled analysis of three cohort studies," Journal of Behavioral Medicine, vol. 37, pp. 881-889, 2014.

[12] R. A. Hahn, G. W. Heath, M.-H. Chang, "Cardiovascular disease risk factors and preventive practices among adults—United States, 1994: a behavioral risk factor atlas," Morbidity and Mortality Weekly Report: CDC Surveillance Summaries, pp. 35-69, 1998.

[13] M. Vassilaki, M. Linardakis, D. M. Polk, A. Philalithis, "The burden of behavioral risk factors for cardiovascular disease in Europe. A significant prevention deficit," Preventive Medicine, vol. 81, pp. 326-332, 2015.

[14] D. S. Kazi, J. M. Penko, K. Bibbins-Domingo, "Statins for Primary Prevention of Cardiovascular Disease: Review of Evidence and Recommendations for Clinical Practice," The Medical Clinics of North America, vol. 101, no. 4, pp. 689-699, 2017, doi: 10.1016/j.mcna.2017.03.001.

[15] R. Chou, A. Cantor, T. Dana, J. Wagner, A. Y. Ahmed, R. Fu, et al., "Statin Use for the Primary Prevention of Cardiovascular Disease in Adults: Updated Evidence

Report and Systematic Review for the US Preventive Services Task Force," JAMA, vol. 328, no. 8, pp. 754-771, 2022, doi: 10.1001/jama.2022.12138.
[16] A. Sitar-tăut, D. Zdrenghea, D. Pop, D. Sitar-tăut, "Using machine learning algorithms in cardiovascular disease risk evaluation," Age, vol. 1, no. 4, p. 4, 2009.
[17] D. K. Arnett, R. S. Blumenthal, M. A. Albert, A. B. Buroker, Z. D. Goldberger, E. J. Hahn, et al., "2019 ACC/AHA guideline on the primary prevention of cardiovascular disease: a report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines," Circulation, vol. 140, no. 11, pp. e596-e646, 2019.
[18] M. J. Pencina, A. M. Navar-Boggan, R. B. D'Agostino Sr, K. Williams, B. Neely, A. D. Sniderman, et al., "Application of new cholesterol guidelines to a population-based sample," N Engl J Med, vol. 370, pp. 1422-1431, 2014.
[19] E. S. Istvan, J. Deisenhofer, "Structural mechanism for statin inhibition of HMG-CoA reductase," Science, vol. 292, no. 5519, pp. 1160-1164, 2001.
[20] T. N. Harrison, R. D. Scott, T. C. Cheetham, S.-C. Chang, J.-W. Y. Hsu, R. Wei, et al., "Trends in statin use 2009–2015 in a large integrated health system: pre-and post-2013 ACC/AHA guideline on treatment of blood cholesterol," Cardiovascular Drugs and Therapy, vol. 32, pp. 397-404, 2018.
[21] M. A. De Vera, V. Bhole, L. C. Burns, D. Lacaille, "Impact of statin adherence on cardiovascular disease and mortality outcomes: a systematic review," British Journal of Clinical Pharmacology, vol. 78, no. 4, pp. 684-698, 2014, doi: 10.1111/bcp.12339.
[22] P. Thavendiranathan, A. Bagai, M. A. Brookhart, N. K. Choudhry, "Primary Prevention of Cardiovascular Diseases With Statin Therapy: A Meta-analysis of Randomized Controlled Trials," Archives of Internal Medicine, vol. 166, no. 21, pp. 2307-2313, 2006, doi: 10.1001/archinte.166.21.2307.
[23] C. P. Cannon, E. Braunwald, C. H. McCabe, D. J. Rader, J. L. Rouleau, R. Belder, et al., "Intensive versus moderate lipid lowering with statins after acute coronary syndromes," New England Journal of Medicine, vol. 350, no. 15, pp. 1495-1504, 2004.
[24] T. R. Pedersen, O. Faergeman, J. J. P. Kastelein, A. G. Olsson, M. J. Tikkanen, I. Holme, et al., "High-dose atorvastatin vs usual-dose simvastatin for secondary prevention after myocardial infarction: the IDEAL study: a randomized controlled trial," JAMA, vol. 294, no. 19, pp. 2437-2445, 2005.
[25] A. Davoudi, M. Ahmadi, A. Sharifi, R. Hassantabar, N. Najafi, A. Tayebi, et al., "Studying the effect of taking statins before infection in the severity reduction of COVID-19 with machine learning," BioMed Research International, 2021.
[26] A. Sarraju, A. Ward, S. Chung, J. Li, D. Scheinker, and F. Rodríguez, "Machine learning approaches improve risk stratification for secondary cardiovascular disease prevention in multiethnic patients," Open Heart, vol. 8, no. 2, p. e001802, 2021, doi: 10.1136/openhrt-2021-001802.
[27] Y. Xiong et al., "Machine learning-based prediction model for the efficacy and safety of statins," Frontiers in Pharmacology, Original Research vol. Volume 15 - 2024, 2024-July-29 2024, doi: 10.3389/fphar.2024.1334929.
[28] J. Han et al., "Predicting low density lipoprotein cholesterol target attainment using machine learning in patients with coronary artery disease receiving moderate-dose statin therapy," Scientific Reports, vol. 15, no. 1, p. 5346, 2025/02/13 2025, doi: 10.1038/s41598-025-88693-y.
[29] Y. Wang et al., "Machine learning in cardiovascular risk assessment: Towards a precision medicine approach," European Journal of Clinical Investigation, vol. 55, no. S1, p. e70017, 2025, doi: https://doi.org/10.1111/eci.70017.

[30] S. M. Grundy et al., "2018 AHA/ACC/AACVPR/AAPA/ABC/ACPM/ADA/AGS/APhA/ASPC/NLA/PCNA Guideline on the Management of Blood Cholesterol: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines," Circulation, vol. 139, no. 25, pp. e1082-e1143, 2019, doi: doi:10.1161/CIR.0000000000000625.

[31] A. Murphy et al., "Inequalities in the use of secondary prevention of cardiovascular disease by socioeconomic status: evidence from the PURE observational study," The Lancet Global Health, vol. 6, no. 3, pp. e292-e301, 2018, doi: 10.1016/S2214-109X(18)30031-7.