IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 11, November 2026

DOI: 10.17148/IARJSET.2026.131144

VitalScore: Heart Disease Risk Classification Using Bayesian Logistic Regression and Support Vector Machines

S. Lakshman[1], Banshika Chowdhury B[2], S. Poornima[3]

Department of Computer Science and Engineering (Emerging Technologies),

SRM Institute of Science and Technology, Vadapalani Campus, Chennai, India[1,2]

Department of Computer Science and Engineering (Emerging Technologies)

SRM Institute of Science and Technology, Vadapalani Campus, Chennai, India[3]

Abstract: Heart disease risk models commonly return a point estimate without indicating how uncertain that estimate is for an individual. This paper presents VitalScore, a web-based decision-support prototype that combines Bayesian logistic regression with an RBF-kernel support vector machine (SVM) baseline. The models were evaluated on 918 records from the public Heart Failure Prediction dataset using a stratified 80/20 split. Preprocessing treated zero-coded cholesterol and resting blood-pressure values as missing, encoded nominal variables using one-hot indicators, and standardized features using training data only. On the held-out set of 184 patients, Bayesian logistic regression achieved 90.2% accuracy, 90.4% precision, 92.2% recall, an F1-score of 0.913, and an ROC-AUC of 0.935. The SVM achieved 87.0% accuracy and an ROC-AUC of 0.933. The Bayesian model additionally produced patient-level 95% credible intervals and additive feature contributions. When ST slope was unrecorded, predictions were averaged across its categories. These results show the feasibility of presenting risk, uncertainty, and interpretable contributions in one application; they do not establish clinical effectiveness.

Index Terms: Heart disease, Bayesian logistic regression, support vector machine, uncertainty quantification, clinical decision support.

I. INTRODUCTION

Cardiovascular disease remains a major cause of mortality worldwide. The World Health Organization estimates that cardiovascular diseases account for approximately 17.9 million deaths annually [1]. Conventional risk scores, including the Framingham risk score and Duke treadmill score, provide structured ways to combine clinical measurements [2], [3]. Machine-learning models have also been studied for heart-disease prediction, but their reported probabilities do not always communicate how reliable a particular estimate is.

VitalScore addresses this communication gap by pairing a Bayesian logistic-regression model with a calibrated RBF- kernel SVM comparison model. The Bayesian model represents uncertainty in its parameters and reports a posterior risk distribution for each patient. The application presents this estimate alongside feature contributions and a rule-based triage indicator. The study focuses on technical feasibility using a public dataset; it is not a prospective clinical validation.

The main contributions are: (1) an uncertainty-aware prediction workflow for ten routine clinical inputs; (2) a comparison with SVM and conventional logistic-regression baselines on a held-out test set; and (3) a web application that presents risk, uncertainty, and interpretable feature contributions together.

The work is motivated by four limitations of existing prediction workflows:

• Point estimates do not show reliability, so typical and atypical patient profiles may appear equally certain.

• Probability calibration is not always assessed, and an 80% predicted risk may not match an 80% observed event rate.

• Limited interpretability can leave clinicians without an account of the factors behind an estimate.

• Data-quality problems can bias models silently: 172 patients have a cholesterol value of 0 (not measured), and chest-pain type is often treated as an ordered number.

The objectives are to assess Bayesian logistic regression for risk estimation from routine clinical data; to quantify patient-level uncertainty with interpretable explanations; to compare it with SVM and conventional logistic-regression baselines on held-out data; to support review of borderline, elevated-risk and uncertain cases through transparent triage indicators; and to evaluate the feasibility of a web application for research and educational use. The application accepts ages from 18 to 100 and excludes imaging, ECG waveforms and genetic data.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 11, November 2026

DOI: 10.17148/IARJSET.2026.131144

II. RELATED WORK

Prior work on heart-disease prediction spans ensemble learning, classical classifiers, feature selection, systematic reviews and deep-learning methods. Fitriyani et al. proposed HDPM for clinical decision support, combining DBSCAN outlier removal, SMOTE-ENN resampling and XGBoost [5]. Comparative studies evaluate logistic regression, SVM, naive Bayes, k-nearest neighbours, decision trees and random forests, including work on imbalanced heart-failure data and prognostic features [6], [7]. Other studies use SVM, logistic regression and decision trees with ensemble feature selection for cardiovascular-disease detection [16], while a systematic review surveys cardiac-disease prediction methods [17].

More recent work examines attention-based cross-modal transfer learning [8], machine learning and optimization for heart-attack treatment in women [18], and traditional machine learning and AutoML for heart-data analysis [19]. ECG digitization and classification combines SVM, logistic regression, XGBoost, k-nearest neighbours and deep models on signal imagery [20]. These ECG and multimodal approaches use richer inputs than VitalScore, which operates on ten routine clinical variables.

Across these studies, evaluation generally reports a point prediction per patient. Bayesian classifiers, patient-level uncertainty intervals, probability calibration, treatment of zero-coded values and rules that convert risk estimates into review actions are less prominent. VitalScore targets these areas. Table I summarizes the ten studies described in the project report.

TABLE I. SELECTED RELATED STUDIES

Ref.	Journal, year	Methods relevant to this work
[5]	IEEE Access, 2020	DBSCAN, SMOTE-ENN and XGBoost for clinical decision support
[6]	Scientific Programming, 2022	LR, SVM, naive Bayes, KNN, trees and forest; SMOTE-ENN; tuning
[7]	ACM, 2021	SVM, LR, decision tree, naive Bayes and KNN; prognostic features
[8]	Computers in Biology and Medicine, 2024	Attention-based cross-modal transfer learning
[15]	IEEE Access, 2020	Machine-learning classification for heart-disease identification in e-healthcare
[16]	International Journal of Ambient Computing and Intelligence, 2022	Ensemble feature selection; SVM, LR and decision tree
[17]	International Journal of Information Technologies and Systems Approach, 2022	Systematic review of cardiac-disease prediction
[18]	Computers in Biology and Medicine, 2025	Machine learning and optimization for heart-attack treatment in women
[19]	Computers in Biology and Medicine, 2025	Traditional ML and AutoML, feature engineering and balancing
[20]	2025 International Conference on Ambient Intelligence in Health Care	ECG digitization; ML and deep-learning classification

Bayesian logistic regression places priors on model weights and yields a posterior distribution. The Laplace approximation fits a Gaussian around the most probable weights [10], [11], making posterior intervals computationally practical. Standard SVMs do not directly produce calibrated probabilities; Platt scaling maps decision scores to probability estimates [12]. Prior work compares calibration methods [13], and the Brier score measures probabilistic accuracy [14].

© IARJSET This work is licensed under a Creative Commons Attribution 4.0 International License

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 11, November 2026

DOI: 10.17148/IARJSET.2026.131144

TABLE II. COMPARISON OF EXISTING TECHNIQUES

Technique	Strength	Limitation for this problem
Framingham / Duke scores	Clinically validated, transparent	Few inputs; no per-patient uncertainty; not refitted
Logistic regression	Simple, fast, exact explanations	One number per patient; no weight uncertainty
SVM (RBF kernel)	Strong on small tabular data	Needs Platt scaling for probabilities; hard to explain
Random forest / boosting	High accuracy, nonlinear	Black box; needs calibration and post-hoc explanation
Neural networks	Flexible on signals and images	Data-hungry; can be overconfident on small tabular data
Bayesian LR (MCMC)	Strong uncertainty characterization	Slow; requires sampling and tuning
Bayesian LR (Laplace), this work	Fast; credible intervals; exact contributions	Gaussian posterior assumed; linear decision boundary

III. DATA AND METHODS

A. Dataset

The study used the public Heart Failure Prediction dataset published by fedesoriano [4]. It contains 918 patient records compiled from five UCI heart-disease collections. The target is binary: 508 records (55.3%) indicate heart disease and 410 (44.7%) do not. Ten clinical inputs were used: age, sex, chest- pain type, resting blood pressure, cholesterol, fasting blood sugar, maximum heart rate, exercise-induced angina, ST depression, and ST-segment slope.

The class distribution did not require resampling (Fig. 1). ST depression and age correlate positively with disease and maximum heart rate negatively; the raw cholesterol correlation is distorted by the 172 zero-coded values (Fig. 2).

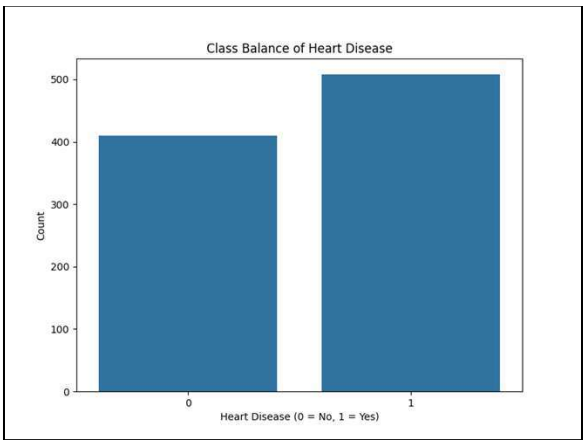


Fig. 1. Class balance of the target variable.

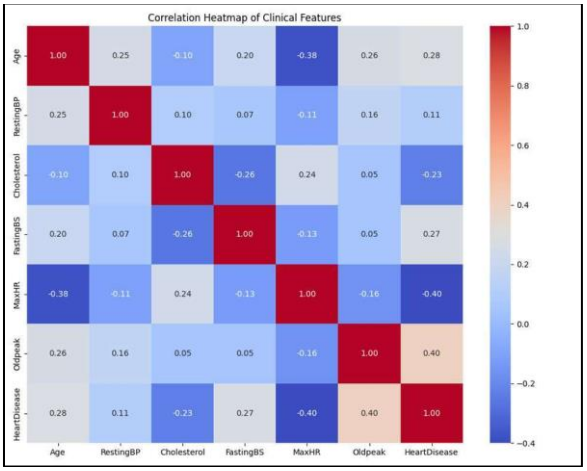


Fig. 2. Correlation heatmap of the numeric features.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 11, November 2026

DOI: 10.17148/IARJSET.2026.131144

Table III lists the ten inputs and how each was handled;

TABLE III. CLINICAL INPUTS AND HANDLING

Input	Type	Handling
Age	Numeric (years)	Standardized
Sex	Binary	Used as coded
Chest-pain type	Nominal	One-hot indicators
Resting blood pressure	Numeric (mm Hg)	Zero treated as missing; median imputation
Cholesterol	Numeric (mg/dl)	Zero treated as missing (172 values); median imputation
Fasting blood sugar	Binary (above threshold)	Used as coded
Maximum heart rate	Numeric (bpm)	Standardized
Exercise-induced angina	Binary	Used as coded
ST depression	Numeric (mm)	Standardized
ST-segment slope	Nominal (up, flat, down)	One-hot; pooled over categories if unrecorded

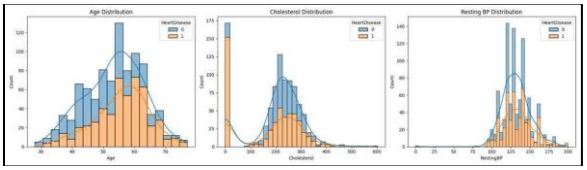


Fig. 3. Feature distributions by heart-disease class (age, cholesterol, resting blood pressure).

B. Preprocessing and Feature Representation

A stratified split assigned 734 records to training and 184 to testing, using random seed 42. One hundred seventy-two cholesterol values and one resting blood-pressure value were coded as zero, representing unrecorded measurements. These values were treated as missing and imputed with medians calculated from the training partition only. Chest-pain type and ST slope were represented as nominal variables using one- hot indicators; the remaining inputs retained their numeric or binary representation. A standard scaler was fitted to the training data and applied unchanged to the test data.

The ten clinical inputs yielded 13 model features after categorical encoding. Resting ECG was not included. For an unrecorded ST slope at inference time, the system calculates a prediction for each possible slope and pools the predictions using category frequencies from the training data. This avoids assigning an unobserved category and allows the resulting interval to reflect the missing input.

C. Prediction Models

The primary model is Bayesian logistic regression with a weakly informative normal prior. Let x denote the standardized feature vector and θ the intercept and regression coefficients. The probability of disease is given by the logistic function, $p(y = 1 \mid x, \theta) = 1 / (1 + \exp(-x^T\theta))$. The posterior is approximated by a Gaussian centered at the maximum a posteriori estimate, with covariance given by the inverse Hessian [10], [11]. For each prediction, 2,000 parameter vectors are drawn from this approximation. Their predicted probabilities yield the posterior mean and 95% credible interval.

$$\theta \mid D \approx N(\theta MAP, \Sigma), \Sigma = H^{-1} \quad (1)$$

$$\hat{p}(x) = (1/S) \Sigma_s 1 / (1 + \exp(-x^T\theta_s)) \quad (2)$$

The prior is $N(0, 10^2)$. The MAP estimate is found with L- BFGS-B using the analytic gradient (up to 1,000 iterations, tolerance 10^{-9}); the Hessian at the optimum is inverted to obtain Σ. The S = 2,000 draws use a fixed seed (42), so identical inputs always give identical outputs. Each feature contributes $\beta_j x_j$ to the log-odds, so the intercept plus all contributions reproduces the model log-odds exactly.

The comparison model is an SVM with a radial basis function kernel [9]. Hyperparameters C and γ were selected by five-fold cross-validation on the training partition using ROC- AUC. Platt scaling was used to convert decision scores into probabilities [12]. A conventional logistic-regression model was also evaluated as a reference.

The selected SVM used C = 1.0 and γ = scale (1/13 ≈ 0.077), with 320 support vectors. Its five-fold cross-validated ROC-AUC was 0.916 ± 0.029, against 0.923 ± 0.034 for plain logistic regression on the same folds. Platt scaling maps the decision value f(x) to a probability, as in (3); the fitted parameters were A ≈ −1.92 and B ≈ −0.004. Table IV summarizes the three candidates.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 11, November 2026

DOI: 10.17148/IARJSET.2026.131144

$$p(y = 1 \mid x) = 1 / (1 + \exp(A \cdot f(x) + B)) \quad (3)$$

TABLE IV. CANDIDATE MODELS AND SELECTION RATIONALE

Model	Role	Why it was chosen
Bayesian LR (Laplace)	Main	Per-patient posterior and credible interval; exact feature contributions; fast enough for real-time serving
RBF SVM, Platt scaling	Baseline	Strong classical model on small tabular data; captures nonlinear structure
Plain logistic regression	Reference	Same linear form without priors; isolates what Bayesian treatment adds

D. Evaluation and Triage

Final evaluation used the held-out test set and a 0.5 classification threshold. Metrics included accuracy, precision, recall, F1-score, ROC-AUC, and Brier score [14]. Triage indicators were calibrated using training data only. The borderline-risk interval was defined as 30% to below 70%. A separate uncertainty alert was set at the 85th percentile of training-set credible-interval widths, corresponding to 31.9 percentage points. These indicators are intended to organize review, not to determine care.

Labels follow a fixed priority (Table V). When ST slope is unrecorded, the three slope predictions are pooled using training frequencies of 42.6% (up), 51.0% (flat) and 6.4% (down).

TABLE V. TRIAGE RULES IN PRIORITY ORDER

Label	Condition
Dual Escalation	Risk ≥ 70% with a wide credible interval
Elevated Risk	Risk ≥ 70%
Uncertainty Alert	95% interval width > 31.9 percentage points
Borderline Risk	Risk from 30% to below 70%
Calibrated Normal	None of the above

Only the 734-patient training split was used for fitting, tuning and threshold selection. The interval-width alarm depends on a patient's inputs and the posterior, not on the outcome, so it does not leak labels. Seeds are fixed (42) for the split, the SVM and the posterior draws, library versions are pinned, and the deployed model artifact (v1.3.0) records its hyperparameters, thresholds and metrics.

The Brier score is the mean squared difference between predicted probability and outcome, as in (4); lower values indicate better probabilistic accuracy. The credible interval is the 2.5th to 97.5th percentile of the 2,000 predicted probabilities, and its width is the difference between the two.

$$\text{Brier} = (1/N) \Sigma_i (\hat{p}_i - y_i)^2 \quad (4)$$

IV. SYSTEM IMPLEMENTATION

VitalScore consists of a React single-page application, a FastAPI prediction service, and an offline model-training pipeline. The application and API are deployed under the same domain on Vercel. The browser submits assessment inputs to the API and displays the posterior mean, credible interval, SVM probability, feature contributions, and triage label.

Patient records are stored in browser storage; the server does not retain patient records. Model artifacts include preprocessing parameters, fitted model data, posterior covariance, and triage thresholds.

Inputs are validated by the form and again by the API before model evaluation.

The browser loads the static application from a content delivery network and sends assessment data as JSON to /api/predict. The serverless Python function validates the request, evaluates both models from the bundled artifact, and returns the results. The artifact holds the scaler, fitted models, imputation values, slope frequencies and triage thresholds.

The API applies per-client rate limiting (120 requests per minute) and a 64 KB request limit, and logs no patient values.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 11, November 2026

DOI: 10.17148/IARJSET.2026.131144

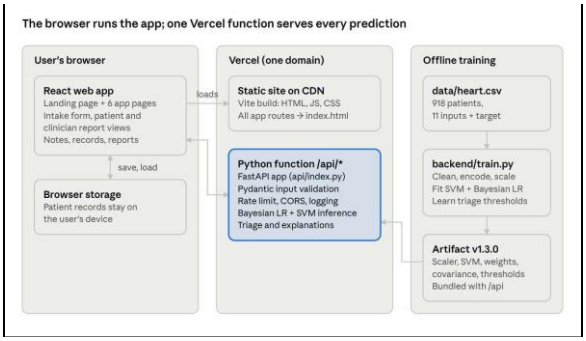


Fig. 4. VitalScore system architecture.

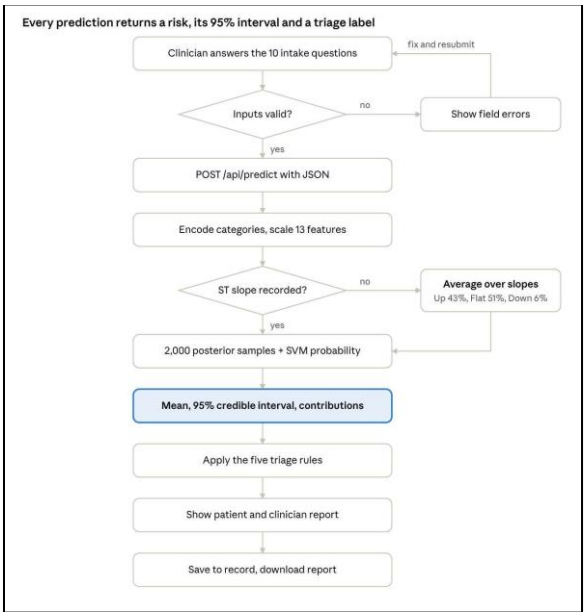


Fig. 5. Prediction request flow.

The training and serving workflows use a shared preprocessing module to maintain feature consistency. The inference service loads the versioned model artifact on the first request and reuses it for subsequent predictions. The project reports 50 backend tests and 25 frontend tests covering validation, prediction invariants, preprocessing consistency, triage conditions, API behavior, and report rendering.

The web application provides an intake form with plain- language questions, separate patient and clinician views of the report, a patient registry with risk history, clinical notes, laboratory results with reference ranges, medications and encounters, and a downloadable HTML clinical report.

Records are autosaved to the browser, and the server stores no personal data.

Table VI lists the software modules and Table VII the main tools.

TABLE VI. PRINCIPAL SOFTWARE MODULES

Module	Responsibility
Preprocessing	Zero-to-missing rules, median imputation, one-hot encoding and standardization; shared by training and serving
Training and export	Stratified split, SVM grid search, Laplace fit, threshold calibration and versioned artifact
Inference engine	Loads artifact, draws 2,000 posterior samples and returns risk, interval, SVM probability, contributions and triage
API layer	/predict, /model-info, /health; validation, CORS, rate limiting and privacy-safe logging
Web application	Intake form, reports, patient registry, notes and HTML report export
Tests	50 backend and 25 frontend tests

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 11, November 2026

DOI: 10.17148/IARJSET.2026.131144

TABLE VII. SOFTWARE STACK

Tool	Version	Used for
Python	3.12	Data processing, training and API
scikit-learn	1.9.1	Scaler, SVC with Platt scaling, grid search and metrics
SciPy	1.18.1	L-BFGS-B optimizer for MAP fit
FastAPI, Pydantic	0.154.0, 2.14.1	REST API and input validation
React, Vite	18, 5	User interface and build
pytest, Vitest	9.1.1, 3	Backend and frontend tests

V. RESULTS AND DISCUSSION

Table VIII summarizes performance on the 184-patient held-out test set. Bayesian logistic regression achieved 90.2% accuracy and an ROC-AUC of 0.935. The RBF SVM achieved 87.0% accuracy and an ROC-AUC of 0.933. The conventional logistic-regression reference produced the same point- prediction metrics as the Bayesian model on this split. Brier scores were 0.097 for the Bayesian model and 0.096 for both comparison models.

TABLE VIII. HELD-OUT TEST PERFORMANCE

Metric	Bayesian LR	RBF SVM	Plain LR
Accuracy	0.902	0.870	0.902
Precision	0.904	0.868	0.904
Recall	0.922	0.902	0.922
F1-score	0.913	0.885	0.913
ROC-AUC	0.935	0.933	0.935
Brier score	0.097	0.096	0.096

The two main models showed similar discrimination, while the Bayesian model classified six more test cases correctly at the 0.5 threshold. It produced 10 false positives and 8 false negatives, compared with 14 and 10 for the SVM.

These differences are descriptive and should be interpreted cautiously because performance was measured on one test split.

Table IX and Fig. 6 give the confusion matrices. The Bayesian model correctly classified 166 test patients against 160 for the SVM, with specificity of 87.8% (72/82) versus 82.9%.

TABLE IX. CONFUSION MATRICES ON THE TEST SET

Model	TN	FP	FN	TP
Bayesian LR	72	10	8	94
RBF SVM	68	14	10	92

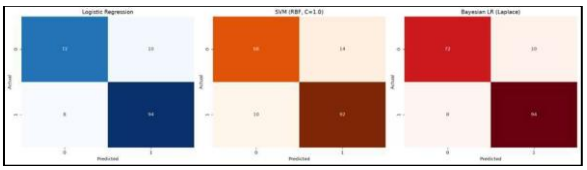


Fig. 6. Confusion matrices on the test set (plain LR, SVM, Bayesian LR).

Table X compares training and test results. The SVM had the higher training ROC-AUC (0.959 versus 0.928), which did not carry over to the test set and may indicate a larger train- test gap; a single split cannot establish generalization.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 11, November 2026

DOI: 10.17148/IARJSET.2026.131144

TABLE X. TRAINING VERSUS TEST PERFORMANCE

Model	Train acc.	Test acc.	Train AUC	Test AUC
Bayesian LR	0.862	0.902	0.928	0.935
RBF SVM	0.895	0.870	0.959	0.933

The ROC curves of the three models overlap closely, with AUC from 0.933 to 0.935 (Fig. 7). The reliability diagram (Fig. 8) shows predicted probabilities broadly tracking observed frequencies for both models; with 184 test patients and one split, this should be read cautiously.

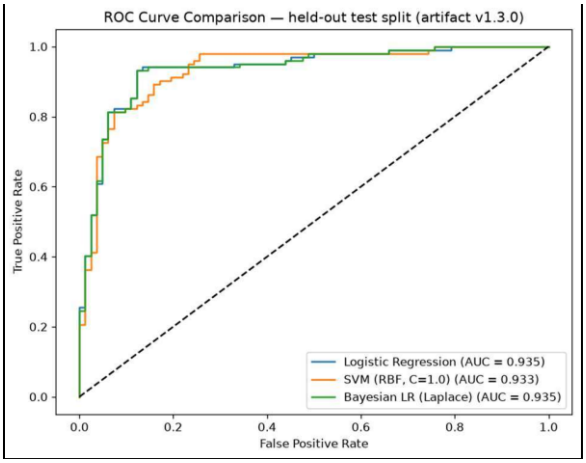


Fig. 7. ROC curve comparison on the held-out test set.

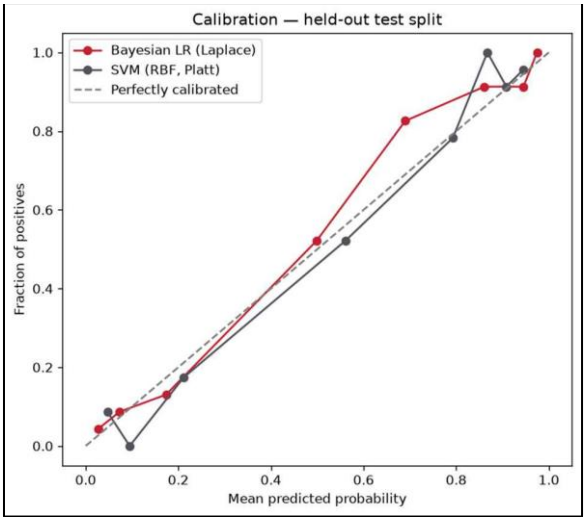


Fig. 8. Calibration (reliability) diagram, Bayesian LR versus SVM.

The Bayesian model's mean 95% credible-interval width was 18.1 percentage points. Widths ranged from 1.8 to 61.4 points, indicating substantial variation in posterior uncertainty across patients. The uncertainty threshold flagged 16.3% of test cases. The 30–70% borderline band contained 19.6% of test patients and 39% of test errors. Interval width and borderline risk are therefore reported as separate signals: the former indicates uncertainty in the model estimate, while the latter identifies estimates near the classification threshold.

TABLE XI. UNCERTAINTY SUMMARY, BAYESIAN MODEL (TEST SET)

Quantity	Value
Mean 95% interval width	18.1 points
Mean posterior standard deviation	4.7 points

© IARJSET This work is licensed under a Creative Commons Attribution 4.0 International License

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 11, November 2026

DOI: 10.17148/IARJSET.2026.131144

Interval width range	1.8 to 61.4 points
Flagged by width alarm (31.9 points)	16.3% of patients
Share in 30–70% band	19.6%
Share of test errors in 30–70% band	39%
Accuracy inside / outside band	80.6% / 92.6%

Table XI summarizes the uncertainty results. Fig. 9 shows the distribution of interval widths.

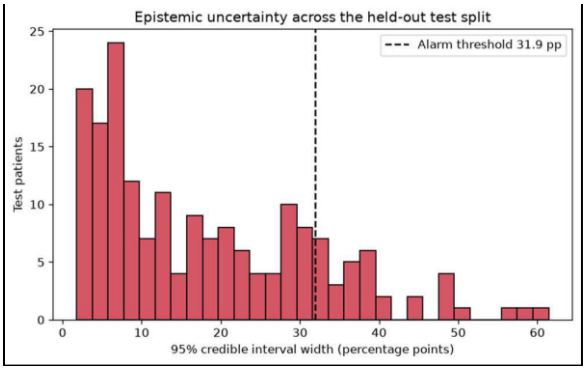


Fig. 9. Distribution of 95% credible-interval widths; dashed line marks the 31.9-point alert threshold.

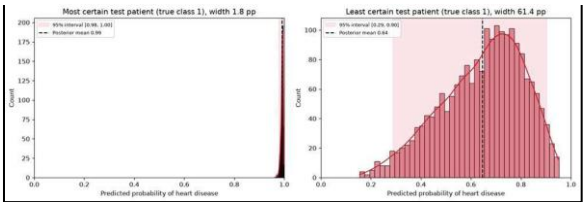


Fig. 10. Posterior predictive distributions for the most and least certain test patients.

For the Bayesian model, the largest positive standardized coefficients were associated with flat ST slope (+1.23), asymptomatic chest pain (+0.73), male sex (+0.59), elevated fasting blood sugar (+0.54), exercise-induced angina (+0.40), and downsloping ST slope (+0.39). Maximum heart rate had a negative coefficient (−0.26). These are model associations within this dataset and should not be interpreted as causal effects. The 95% intervals for age, cholesterol, resting blood pressure, and several chest-pain categories included zero.

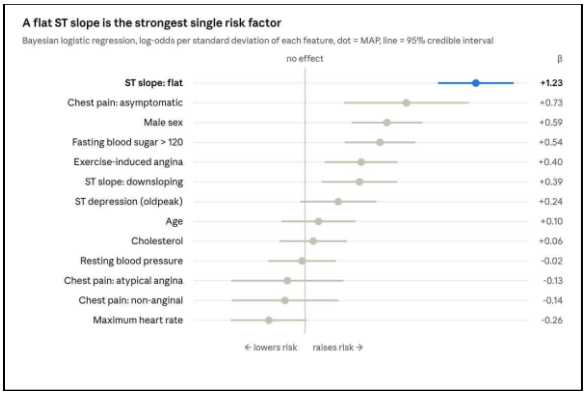


Fig. 11. Bayesian LR coefficients (log-odds per standard deviation) with 95% credible intervals.

In this dataset, 79% of asymptomatic chest-pain records were positive versus 13.9% of atypical-angina records.

Because the data come from referral cohorts, these rates may reflect patient selection and are not population prevalence.

Table XII shows the application's sample patient: a 55- year-old man with atypical angina, resting blood pressure 142, cholesterol 236, maximum heart rate 148, exercise angina, and ST depression 1.6 mm. Estimated risk changes with the entered ST slope, and the interval widens when the slope is unrecorded, reflecting the missing measurement.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 11, November 2026

DOI: 10.17148/IARJSET.2026.131144

TABLE XII. SAMPLE-PATIENT RISK BY ST-SLOPE ENTRY

ST slope	Risk	95% CI	SVM	Triage
Rising	19.4%	9%–34%	21.5%	Calibrated Normal
Flat	71.9%	54%–85%	78.9%	Elevated Risk
Dipping	52.4%	24%–79%	57.2%	Uncertainty Alert
Not recorded	48.3%	11%–84%	53.0%	Uncertainty Alert

Preprocessing and feature representation materially affected the results. The initial version retained zero-coded cholesterol values and treated chest-pain categories as ordinal;

it achieved 88.0% accuracy but estimated a negative cholesterol coefficient. Treating those values as missing and encoding nominal categories corrected the coefficient direction. Adding ST slope increased accuracy from 84.8% to 90.2% (Table XIII).

TABLE XIII. EFFECT OF PREPROCESSING AND ST SLOPE

Model version	Accuracy	ROC-AUC
Zeros kept, ordinal chest pain	0.880	0.916
Zeros missing, one-hot chest pain	0.848	0.916
Final model, ST slope added	0.902	0.935

The main findings on this split are:

• The Bayesian model matched or exceeded the SVM (90.2% versus 87.0% accuracy; ROC-AUC 0.935 versus 0.933) while also giving each patient a 95% credible interval and exact feature contributions.

• Data quality mattered more than model complexity:

treating zero cholesterol as missing, encoding chest pain as nominal, and adding ST slope changed the result more than any change of algorithm.

• Errors concentrate in the 30-70% risk band (16% of training patients but 47% of training errors), which motivated the Borderline Risk flag; the width alarm is kept separate because it marks atypical patients rather than likely errors.

• An unrecorded ST slope is handled by averaging over its three categories, which keeps a sensible estimate and widens the interval.

VI. LIMITATIONS AND FUTURE WORK

The study has the following limitations:

• The dataset has 918 patients collected in the 1980s from five referral centres; practice and risk patterns may have changed, and referral populations may differ from the general population.

• Men account for 79% of records, so estimates for women rest on fewer patients (193).

• Metrics come from one stratified 80/20 split. With 184 test patients, one patient is about 0.5 percentage points of accuracy, so small differences between models are not conclusive.

• No independent hospital or cohort was used; external validation is necessary before clinical use can be considered.

• The linear decision boundary cannot capture interactions unless they are added explicitly, and the Laplace approximation can understate uncertainty in the tails compared with full MCMC.

• The interval reflects uncertainty about model weights only, not measurement error in the inputs. The prototype has not undergone prospective evaluation, regulatory review or integration with clinical record systems.

Future work should evaluate the model on independent and more recent cohorts, use repeated validation with uncertainty intervals for performance metrics, assess calibration and performance across demographic groups, and compare alternative posterior-inference methods. Clinical usability studies would be needed to determine whether uncertainty intervals and triage indicators improve review decisions.

© IARJSET This work is licensed under a Creative Commons Attribution 4.0 International License

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 11, November 2026

DOI: 10.17148/IARJSET.2026.131144

VII. CONCLUSION

This paper presented VitalScore, a web-based heart- disease risk assessment prototype that pairs Bayesian logistic regression with an RBF SVM baseline. On one held-out test split, the Bayesian model achieved 90.2% accuracy and an ROC-AUC of 0.935, while providing patient-level credible intervals and additive feature contributions. The results support the technical feasibility of integrating risk and uncertainty in a single application. Independent validation and clinical evaluation remain necessary before the approach can be considered for clinical use.

REFERENCES

[1] World Health Organization, "Cardiovascular diseases (CVDs)," Fact Sheet, 2021.

[2] P. W. F. Wilson et al., "Prediction of coronary heart disease using risk factor categories," Circulation, vol. 97, no. 18, pp. 1837–1847, 1998.

[3] D. B. Mark et al., "Prognostic value of a treadmill exercise score in outpatients with suspected coronary artery disease," New England Journal of Medicine, vol. 325, no. 12, pp. 849–853, 1991.

[4] fedesoriano, "Heart Failure Prediction Dataset," Kaggle, 2021.

[5] N. L. P. W. Fitriyani et al., "HDPM: An Effective Heart Disease Prediction Model for a Clinical Decision Support System," IEEE Access, vol. 8, 2020, doi: 10.1109/ACCESS.2020.3010511.

[6] M. M. Nishat et al., "A Comprehensive Investigation of the Performances of Different Machine Learning Classifiers with SMOTE- ENN Oversampling Technique and Hyperparameter Optimization for Imbalanced Heart Failure Dataset," Scientific Programming, 2022, Art. no. 3649406, doi: 10.1155/2022/3649406.

[7] Y. Srizon et al., "Identifying Prognostic Features for Predicting Heart Failure by Using Machine Learning Algorithm," ACM, 2021, doi: 10.1145/3460238.3460245.

[8] V. Jothi Prakash et al., "A Novel Attention-Based Cross-Modal Transfer Learning Framework for Predicting Cardiovascular Disease," Computers in Biology and Medicine, vol. 170, 2024, Art. no. 107977, doi: 10.1016/j.compbiomed.2024.107977.

[9] C. Cortes and V. Vapnik, "Support-vector networks," Machine Learning, vol. 20, no. 3, pp. 273–297, 1995.

[10] D. J. C. MacKay, "The evidence framework applied to classification networks," Neural Computation, vol. 4, no. 5, pp. 720–736, 1992.

[11] C. M. Bishop, Pattern Recognition and Machine Learning. Springer, 2006.

[12] J. Platt, "Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods," in Advances in Large Margin Classifiers, pp. 61–74, 1999.

[13] A. Niculescu-Mizil and R. Caruana, "Predicting good probabilities with supervised learning," in Proc. 22nd Int. Conf. Machine Learning, pp. 625–632, 2005.

[14] G. W. Brier, "Verification of forecasts expressed in terms of probability," Monthly Weather Review, vol. 78, no. 1, pp. 1–3, 1950.

[15] J. P. Li et al., "Heart disease identification method using machine learning classification in e-healthcare," IEEE Access, vol. 8, pp. 107562–107582, 2020, doi: 10.1109/ACCESS.2020.3001149.

[16] D. GhoshRoy, P. A. Alvi, and J. M. R. S. Tavares, "Detection of cardiovascular disease using ensemble feature engineering with decision tree," International Journal of Ambient Computing and Intelligence, vol. 13, no. 1, pp. 1–16, 2022, doi: 10.4018/IJACI.300795.

[17] S. Wadhawan and R. Maini, "A systematic review on prediction techniques for cardiac disease," International Journal of Information Technologies and Systems Approach, vol. 15, no. 1, pp. 1–33, 2022, doi: 10.4018/IJITSA.290001.

[18] D. S. Khafaga et al., "Empowering heart attack treatment for women through machine learning and optimization techniques," Computers in Biology and Medicine, vol. 195, Art. no. 110597, 2025, doi: 10.1016/j.compbiomed.2025.110597.

[19] S. U. Değer, "A study on heart data analysis and prediction using advanced machine learning methods," Computers in Biology and Medicine, vol. 192, Part B, Art. no. 110308, 2025.

[20] "Digitization and classification of electrocardiograms for automated cardiovascular disease diagnosis using machine learning and deep learning models," in Proc. 2025 Int. Conf. Ambient Intelligence in Health Care, 2025, doi: 10.1109/ICAIHC64101.2025.10956735.

© IARJSET This work is licensed under a Creative Commons Attribution 4.0 International License