SCOPE & COVERAGE

- **Sections Covered:** 34 specialized sections spanning 143 topics across Medicine, Biology, Environmental Sciences, and General Science
- Ensures broad interdisciplinary visibility for high-impact research.

PUBLICATION FEATURES

- **Review Process:** Double-blind peer review ensuring transparency and quality
- **Time to Publication:** Rapid 21-day review-to-publication cycle
- **Frequency:** Published monthly
- **Plagiarism Screening:** All submissions checked with iThenticate

INDEXING & RECOGNITION

- **Indexed in:** Google Scholar, IndexCopernicus (**ICV 2022: 88.03**)
- **DOI:** Registered with CrossRef (**10.37871**) for long-term discoverability
- **Visibility:** Articles accessible worldwide across universities, research institutions, and libraries

OPEN ACCESS POLICY

- Fully Open Access journal under Creative Commons Attribution 4.0 License (CC BY 4.0)
- Free, unrestricted access to all articles globally

GLOBAL ENGAGEMENT

- **Research Reach:** Welcomes contributions worldwide
- **Managing Entity:** SciRes Literature LLC, USA
- **Language of Publication:** English

SUBMISSION DETAILS

- Manuscripts in Word (.doc/.docx) format accepted

SUBMISSION OPTIONS

- **Online:** https://www.jelsciences.com/submit-your-paper.php
- **Email:** support@jelsciences.com, support@jbresonline.com

HOME · ABOUT · ARCHIVE · SUBMIT MANUSCRIPT · APC

Vision: The Journal of Biomedical Research & Environmental Sciences (JBRES) is dedicated to advancing science and technology by providing a global platform for innovation, knowledge exchange, and collaboration. Our vision is to empower researchers and scientists worldwide, offering equal opportunities to share ideas, expand careers, and contribute to discoveries that shape a healthier, sustainable future for humanity.

RESEARCH ARTICLE

Gut Microbiome-Augmented Models for Heart Failure Survival in FINRISK 2002

Hojin Moon*, Nhat Hoang Nguyen, Jaehee Park, Sohyul Ahn and Gyumin Kim

Department of Mathematics and Statistics, California State University, Long Beach, 1250 Bellflower Blvd., Long Beach, CA 90840, USA

Abstract

Background: Conventional Heart-Failure (HF) risk scores use limited clinical variables and show moderate accuracy. We tested whether gut-microbiome signatures improve long-term HF survival prediction in a population-based Finnish cohort.

Methods: We analyzed FINRISK 2002 participants free of HF at baseline (n = 5,212; training 3,471; test 1,741; median follow-up 13.8 years). Microbiome profiles were prevalence-abundance filtered (≥ 30% prevalence; ≥ 0.01% relative abundance), centered log-ratio transformed, and reduced to 125 core taxa. We trained penalized Cox (elastic net), Random Survival Forests (RSF), and DeepSurv neural network under two feature sets: nine-clinical covariates alone versus clinical + microbiome. Feature selection was performed using elastic net and RSF. Discrimination was assessed with Harrell's C-index on a test set, and model comparisons were tested for significance.

Results: Microbiome-enhanced models achieved test-set C-indices of 0.7225 (elastic net), 0.7231 (RSF), and 0.7211 (DeepSurv), modestly exceeding the Cox baseline model (0.7110). The elastic-net model selected 14 predictors, including age, prevalent coronary heart disease, and multiple microbial taxa. Absolute gains were modest yet statistically significant across algorithms.

Conclusion: In a population-based cohort with long follow-up, incorporating gut microbiome taxa with routine clinical variables produced modest but consistent improvements in HF survival prediction. Microbiome-informed models may enhance early risk stratification and support personalized prevention in cardiovascular care.

*Corresponding author(s)

Hojin Moon, Department of Mathematics and Statistics, California State, University, Long Beach, 1250 N Bellflower Blvd., Long Beach, CA 90840-1001, USA

Tel: +949-880-7819
Fax: +562-985-8227
Email: hojin.moon@csulb.edu

Introduction

Cardiovascular Disease (CVD) continues to be a formidable global health challenge, remaining the leading cause of mortality worldwide [1]. Heart Failure (HF) is a complex and multifactorial cardiovascular condition that imposes a significant burden on patients and healthcare systems. It is characterized by the heart's inability to pump sufficient blood to meet the body's needs, leading to symptoms such as shortness of breath, fatigue, and fluid retention.

HF has reached pandemic proportions, affecting an estimated 64 million people worldwide and steadily rising in parallel with ageing populations, obesity, and diabetes [2]. In the United States, prevalence is projected to rise from 6.7 million today to approximately 8.5 million

by 2030 as populations age and cardiometabolic comorbidities become more common [3]. Although therapeutic options have expanded, long-term outcomes remain dismal: Contemporary cohorts report a five-year all-cause mortality of roughly 50–60% and a median survival of only three years after diagnosis [4]. The economic toll is enormous-an up-to-date systematic review places the annual global cost of HF at approximately US $284 billion, almost half of which is direct medical expenditure [5].

This trend demonstrates the pressing need for accurate risk prediction models that can enhance clinical decision-making and improve patient outcomes in HF management. These figures also highlight an unmet clinical need for more accurate, personalized prognostic tools. Current risk scores incorporate a limited set of demographic and haemodynamic variables, yielding only moderate discrimination and sub-optimal calibration when applied to diverse patient groups.

Integrating novel data types-such as fecal microbiota composition and host phenotypic traits-with traditional clinical factors offers a promising path to improve predictive accuracy by providing a more comprehensive assessment of patient risk. This goal of this study is to develop an innovative clinical support system using advanced Machine Learning (ML) algorithms to integrate such heterogeneous data. By combining gut microbiota profiles with clinical variables, we aim to establish a robust, multifaceted framework for HF survival prediction.

Mounting evidence implicates the gut-heart axis as a key modulator of HF pathophysiology. Profiling studies have revealed a distinctive gut microbial signature in chronic HF, characterized by reduced diversity and depletion of butyrate-producing taxa [6]. Similarly, Cui X, et al. [7] found evidence of gut dysbiosis in patients with ischemic cardiomyopathy, suggesting that imbalances in the gut microbiota may influence HF pathogenesis.

Several pathophysiological mechanisms have been proposed to explain this gut-heart connection. Diminished cardiac output and venous congestion impair intestinal perfusion, increase epithelial permeability, and facilitate the translocation of endotoxins and microbial metabolites. These processes exacerbate systemic inflammation and drive maladaptive cardiac remodeling [8]. These observations position the gut microbiome as both a biomarker and a potential therapeutic target, yet its predictive value has not been systematically incorporated into large-scale survival models.

A growing body of evidence supports bidirectional gut–heart cross talk in CVD. In HF, reduced cardiac output and splanchnic venous congestion impair mucosal perfusion, disrupt epithelial barrier integrity, and facilitate translocation of endotoxins and microbe derived metabolites, thereby amplifying systemic inflammation and adverse cardiac remodeling-the classic "gut hypothesis" of HF [6-8]. Profiling studies across independent cohorts report dysbiosis in HF-reduced diversity and depletion of butyrate producing taxa-implicating perturbed microbial ecology in disease progression [6,7,9].

Microbial co-metabolism generates bioactive compounds with direct cardiovascular effects. Trimethylamine–N–oxide (TMAO), formed from dietary choline/carnitine, shows prospective associations with CVD, augments platelet reactivity and thrombosis, and exacerbates pressure-overload–induced HF in experimental models [10-12]. In contrast, short-chain fatty acids (SCFAs) such as acetate, propionate, and butyrate generally stabilize the intestinal barrier and modulate blood pressure, immune tone, and inflammation through host receptors and epigenetic pathways; depletion of SCFA-producing taxa has been linked to adverse cardiometabolic profiles [9,13,14]. Bile-acid transformations by gut microbes further intersect with lipid and inflammatory signaling relevant to CVD [13,15].

Population data support these mechanistic links, with gut-microbiome composition associating with lifetime CVD risk in community cohorts [16]. Collectively, these insights position the microbiome as both a biomarker source and a potential therapeutic target and motivate our evaluation of whether microbiome features add incremental prognostic value to clinical predictors for long-term HF survival.

In parallel with these biological findings, advanced ML algorithms have emerged as powerful tools for assimilating heterogeneous, high-dimensional data and uncovering non-linear interactions in healthcare for risk prediction. Utilizing ML has the potential to improve the precision, reliability, and clinical applicability of HF prognostic models, ultimately enhancing patient stratification and outcomes [17].

In this study, we employ three distinct ML approaches to develop a comprehensive HF survival

prediction model. Specifically, we utilize penalized Cox Proportional Hazards (PH) regression [18], Random Survival Forests (RSF) [19], and a deep neural network survival model (DeepSurv) [20]. By systematically evaluating these methodologies in parallel, we seek to identify the most effective strategy for integrating gut microbiota data with clinical variables in HF risk prediction.

Accordingly, the present study leverages the population-based FINRISK 2002 cohort [21] released through the DREAM Challenge 2022 to develop microbiome-informed HF survival models. We compare three state-of-the-art ML approaches-penalized Cox regression, RSF, and the DeepSurv neural-network-to create a robust, generalizable decision-support tool that addresses the pressing clinical gap in HF risk stratification. Early identification of individuals at risk enables timely and targeted interventions to prevent HF or mitigate its severity. By fostering interdisciplinary collaboration between clinical cardiology and data science, our approach aligns with the push forward personalized and precision medicine in HF management. Ultimately, we aim to improve patient outcomes and contribute to public health by elevating the standard of HF risk assessment.

Methods

Patient data description

We utilized from the FINRISK 2002 cohort, a comprehensive population-based study of Finnish adults, to develop our predictive models. Our analysis focused on participants free of Heart Failure (HF) at baseline, with follow-up data available for incident HF outcomes. The median follow-up time was 15.8 years. The dataset includes a rich array of demographic, clinical, and microbiome-related variables, such as age, sex, Body Mass Index (BMI), smoking status, medical history (e.g. hypertension treatment, diabetes, Coronary Heart Disease (CHD), HF), systolic BP, non-HDL cholesterol, and survival outcomes (HF events and time to event). Microbiome profiles were derived from fecal samples, yielding 5,748 microbial features per participant. Each participant's record included 10 clinical covariates; only 1.5% of records (n-81) had missing HF outcome data. Baseline demographic and clinical features for the full dataset and its training and testing partitions are summarized in table 1.

To protect patient privacy, the DREAM Challenge released a synthetic version of the FINRISK data,

Table 1: Baseline characteristics of the FINRISK 2002 development dataset by data partition (DREAM provided development dataset: Overall n = 5,424; Training n = 3,615; Test n = 1,809).

Characteristic	Overall	Training	Test	SMD*	P value†
Demographics & anthropometrics					
Age, years (mean ± SD)	49.4 ± 14.8	49.5 ± 14.8	49.3 ± 14.8	0.01	-
Body mass index, kg/m² (mean ± SD)	27.0 ± 4.7	27.1 ± 4.7	26.9 ± 4.7	0.04	0.337
Cardiometabolic risk factors					
Current smoking, *n/N* (%)	1.266/5.397 (23.5)	838/3.599 (23.3)	428/1.798 (23.8)	0.01	0.253
Antihypertensive treatment, *n/N* (%)	842/5.424 (15.5)	562/3.615 (15.5)	280/1.809 (15.5)	0.00	0.379
Prevalent diabetes, *n/N* (%)	339/5.343 (6.3)	220/3.564 (6.2)	119/1.779 (6.7)	0.02	0.560
Prevalent CHD, *n/N* (%)	163/5.343 (3.1)	117/3.564 (3.3)	46/1.779 (2.6)	0.04	0.0163
Systolic blood pressure, mm Hg (mean ± SD)	136 ± 22	136 ± 22	136 ± 22	0.00	0.597
Non-HDL cholesterol, mmol/L (mean ± SD)	4.08 ± 1.09	4.09 ± 1.09	4.08 ± 1.09	0.01	0.153
Male sex, *n/N* (%)	2.395/5.424 (44.2)	1.604/3.615 (44.4)	791/1.809 (43.7)	0.01	0.772
Follow-up and outcomes					
Incident HF during follow-up, *n/N* (%)	445/5.343 (8.3)	300/3.564 (8.4)	145/1.779 (8.2)	0.01	-
Follow-up time, years (mean ± SD)	13.8 ± 5.6	13.8 ± 5.7	13.9 ± 5.3	0.02	-

*SMD (Standardized Mean Difference) is the absolute difference between test and training groups divided by the pooled SD (continuous) or by the pooled binomial SD (categorical). SMD < 0.10 indicates negligible imbalance.

†P value is from Cox regression model and reflects *univariate Cox* associations with HF, not train−test balance. We retained significant value for prevalent CHD (p = 0.0163).

Notes: Per-variable denominators reflect available data: BMI (training missing = 1), Smoking (test = 11; training = 16), Diabetes (test = 30; training = 51), CHD (test = 30; training = 51), Incident HF and Follow-up time (test = 30; training = 51), Systolic BP (test = 1), Non-HDL cholesterol (test = 3; training = 7). Units: mm Hg = millimeters of mercury; mmol/L = millimoles per liter. CHD = Coronary Heart Disease; HF = Heart Failure.

divided into training, testing, and scoring datasets. The combined dataset comprised 7,231 individuals [22]: 5,424 were provided for model development (3,615 in the training set and 1,809 in the test set), and 1,807 were reserved as a hidden evaluation set by the challenge organizers. We used the training set to develop and tune our models (hyperparameter optimization) and the test set for internal performance validation; the held-out scoring set was not accessible and was used only by the organizers for final evaluation. Within the data provided, 131 already had HF at baseline, and 445 experienced an HF event during the follow-up period. We applied the challenge's exclusion criteria to both the training and test datasets, excluding 212 individuals-131 with a history of HF at baseline and 81 with missing HF status-resulting in a final cohort of 5,212 participants (3,471 training; 1,741 testing).

Recognizing the emerging importance of microbiomes in cardiovascular health, we hypothesized that integrating microbiome data could enhance HF risk prediction. The gut microbiome can influence systemic inflammation, immune responses, and metabolic pathways that are relevant to HF prognosis. For example, an imbalance in gut microbes (dysbiosis) is associated with increased inflammation and metabolic disturbances such as insulin resistance and obesity, which are known risk factors for HF. By integrating microbiome-derived features into our models, we aimed to capture these additional dimensions of risk, potentially improving predictive accuracy and enabling more personalized therapeutic strategies.

Modeling framework and data preprocessing

Our modeling framework combined classical survival analysis methods with modern advanced ML techniques to predict HF outcomes. We integrated microbiome features with traditional risk factors in all models. The overall approach consisted of training with three different ML approaches (penalized Cox, RSF, and DeepSurv).

Initially, the datasets were preprocessed. The raw microbiome sequencing data were imported into R (version 4.3.2) using *phyloseq* package (version 1.44.0). This package provides a comprehensive pipeline for microbiome census data integration and processing.

The initial phase involves the identification of core taxa that are microbial taxa consistently observed across most samples, based on detection frequency

and prevalence thresholds. To reduce dimensionality from the original 5,748 microbial features, we applied a prevalence-abundance filtering strategy for identifying core taxa. Specifically, we defined core taxa as those microbial taxa that were present in at least 30% of samples and had a minimum relative abundance threshold of 0.01%. These thresholds were chosen to retain biologically relevant and informative microbial taxa while excluding rare or extremely sparse features that might introduce noise in modeling.

Following this filtering step, the microbiome data underwent compositional transformation using the Centered Log-Ratio (CLR) transformation to approximately account for the compositional nature of microbiome data and compositional bias. The CLR transformation was carried out using the *compositions* R package (version 2.0-6). Following this transformation, the dataset was retained 125 core microbial taxa for model development. These 125 core microbiomes, along with clinical variables, were used as the predictors in our survival models.

Advanced machine learning algorithms: We evaluated three complementary survival-modeling approaches-Penalized Cox regression with the elastic net, Random Survival Forests (RSF), and DeepSurv-to balance interpretability, assumption strength, and flexibility. The elastic net Cox model extends the standard proportional-hazards framework by adding L_1/L_2 penalty terms, yielding sparse, readily interpretable hazard ratios and controlling overfitting in high-dimensional settings. RSF is a fully non-parametric, ensemble-tree method that makes no proportional-hazards or linearity assumptions and can automatically capture complex interactions and nonlinear effects; its variable-importance measures facilitate moderate interpretability. DeepSurv employs a feedforward neural network to learn nonlinear representations of the log-hazard function, maximizing predictive power in the presence of highly complex covariate patterns. Together, these three models span a spectrum from classical, easy-to-explain hazard modeling (penalized Cox) through flexible, interaction-driven ensembles (RSF) to highly expressive deep learning (DeepSurv), allowing us both to benchmark against standard practice and to explore gains in predictive accuracy under increasingly relaxed assumptions.

Baseline clinical models: To establish a purely clinical benchmark against which the added value

of gut-microbiome data could be judged, we first fit three "baseline" survival models-elastic net Cox baseline, RSF baseline, DeepSurv baseline-using only the nine routinely collected clinical predictors available at enrollment (age, sex, body-mass index, current smoking, hypertension treatment, diabetes, prior coronary heart disease, systolic blood pressure, and non-HDL cholesterol). All three baseline models were evaluated on the held-out test cohort using Harrell's C-index. These metrics serve as reference points for the extended models that incorporate the 125 core microbiome features (Sections 2.3.1–3); any statistically significant improvement beyond the baselines can therefore be attributed to the added microbial information rather than differences in model architecture or training procedure.

Penalized Cox regression with the elastic net: We employed a Cox Proportional Hazards (PH) model [23] regularized with the elastic net penalty [18] to handle high-dimensional predictors (the 125 core microbiomes and the selected clinical and demographic variables identified in Section 2.2.1) and avoid overfitting. The elastic net combines L_1 (lasso) and L_2 (ridge) regularization, shrinking coefficient estimates and setting some coefficients exactly to zero. This hybrid penalty retains lasso's ability to perform feature selection while incorporating a small ridge component to stabilize estimates in the presence of highly correlated features. In other words, the elastic net penalty simultaneously encourages parsimony and mitigates multicollinearity by including grouped correlated predictors together rather than arbitrarily selecting only one. This approach is well-suited for our setting of numerous candidate variables, as it can identify a relevant subset of prognostic features with improved robust compared to using lasso or ridge alone.

The elastic net regularization introduces a penalty term into the partial log-likelihood function used in Cox PH regression model, facilitating model sparsity and multicollinearity control. The penalized log-likelihood function is expressed as:

$$l(\lambda, \boldsymbol{\beta}) = -\log\left(\prod_{i=1}^{n}\left\{\frac{\exp(\boldsymbol{\beta}' \boldsymbol{Z}_i)}{\sum_{j \in R(t_i)} \exp(\boldsymbol{\beta}' \boldsymbol{Z}_j)}\right\}^{\delta_i}\right) +$$

$$\alpha\lambda\sum_{j=1}^{p}|\boldsymbol{\beta}_j| + (1-\alpha)\lambda\sum_{j=1}^{p}\boldsymbol{\beta}_j^2,$$

where α ($0 \le \alpha \le 1$) dictates the balance between the

L_1 (lasso, when $\alpha = 1$) and L_2 (ridge, when $\alpha = 0$) regularization, while λ serves as the tuning parameter controlling the overall magnitude of penalization. Higher values of λ impose stronger regularization, leading to more coefficients being shrunk toward zero, whereas lower values result in less penalization, preserving more predictor information.

Prior to model training, all continuous predictor variables were standardized to mean 0 and unit variance. Standardization ensures that the regularization penalty is applied uniformly, preventing features with larger scales from dominating the model purely due to unit differences. The penalized Cox model was implemented using the *glmnet* package in R.

We conducted a 10-fold Cross-Validation (CV) to select the optimal elastic net hyperparameters. Specifically, we evaluated a grid of values for the mixing parameter α (ranging from 0 to 1) and the regularization parameter λ, typically on a logarithmic scale from 10^{-4} to 10^{-2} to cover a broad spectrum of regularization strengths. For each combination (α, λ), the model was trained on 90% of the training data and validated on the remaining 10%, rotating so that each fold served as validation once. Model performance in each fold was assessed via the partial likelihood deviance ($D = -2\log L(\hat{\beta})$), a measure of error based on how well the model's risk predictions concord with observed outcomes. The optimal α and λ were selected by minimizing the average deviance across folds, ensuring the best trade-off between model complexity and fit.

Using these optimal values of α and λ, we refit the penalized Cox model on the entire training set to obtain the final model. In the final model, many coefficients were shrunk toward zero, leaving a subset of non-zero coefficients that represent the most influential predictors of HF survival. This approach results in a more interpretable model that focuses on the key factors contributing to heart failure. The magnitude and sign of these coefficients (expressed as hazard ratios when exponentiated) provide insight into each predictor's impact on risk. A positive coefficient implies an increased hazard (higher risk), while a negative coefficient suggests a reduced hazard (lower risk).

Random survival forests: Random Survival Forests (RSF) is an ensemble tree-based method for survival analysis that extends the Random Forest algorithm

to handle censored time-to-event data [19,24]. An RSF builds many survival trees on bootstrap samples of the training data, where each tree is constructed to predict survival outcomes. At each split in a tree, a random subset of candidate variables is considered (introducing randomness like standard random forests), and the optimal split is chosen based on maximizing the difference in survival between the resulting subgroups (often evaluated by log-rank test statistics). Each survival tree yields an estimate of the Cumulative Hazard Function (CHF) for Heart Failure (HF) for observations falling in its terminal nodes. The CHF provides a cumulative measure of the hazard at each time point, representing the accumulated risk up to that point.

We used the Nelson–Aalen estimator [25,26] within each terminal node to compute the CHF, properly accounting for censored observations. To obtain an ensemble CHF, the estimated CHF from each tree in the RSF, denoted as $H_b^*(t|\mathbf{x}_i)$, is averaged across all trees, yielding the ensemble CHF:

$$H_e^*(t|\mathbf{x}_i) = \frac{1}{n_{tree}} \sum_{b=1}^{ntree} H_b^*(t|\mathbf{x}_i),$$

where n_{tree} represents the total number of trees in the forest. From the ensemble CHF, the ensemble survival function $S_e(t|\mathbf{x}_i)$ is derived as:

$$S_e(t|\mathbf{x}_i) = \exp\left(-H_e^*(t|\mathbf{x}_i)\right).$$

This provides a comprehensive estimation of the patient's survival probability based on the aggregated CHF. The RSF framework, by averaging predictions from multiple de-correlated survival trees, captures complex interactions among variables while mitigating overfitting. By differentiating the ensemble CHF, one can derive the ensemble hazard function, which describes the instantaneous risk of the event at any point in time.

Figure 1 illustrates the RSF method. The left panel shows that the training dataset is repeatedly sampled with replacement (bootstrapped) to build multiple survival trees. Each survival tree is grown using random feature selection at each split and a survival-based splitting criterion (e.g., the log-rank test), producing terminal nodes that include patients with similar outcomes. In practice, hundreds of such bootstrap samples (trees) are generated (often 500 or more) to form the forest for robust prediction. The right panel illustrates that each tree's leaves yield an estimated Cumulative Hazard Function (CHF) for the risk of heart failure over time based on the patients falling in that leaf. The ensemble combines the predictions of all trees by averaging their CHFs, equivalently averaging the corresponding survival probabilities, across the forest, which produces a final ensemble CHF for a given new patient. This ensemble CHF represents the model's overall estimated risk accumulation over time for that individual, integrating information from all trees in the forest. The model was implemented in *R* using the *randomForestSRC* package.

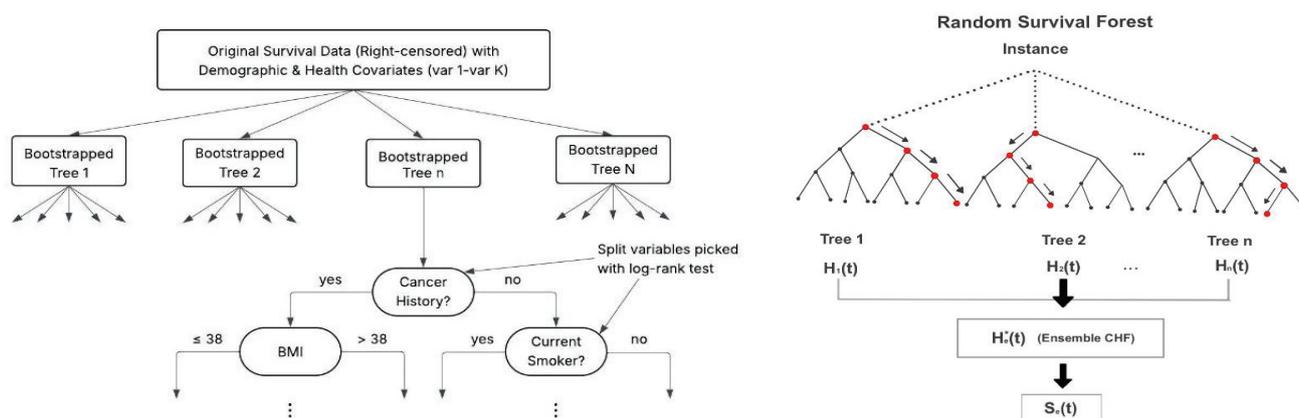RSF also provides measures of variable importance.

Figure 1 Schematic of the RSF pipeline used in this study.
Left panel: bootstrap sampling of the training partition to grow many survival trees; at each node a random subset of predictors is considered, and the split is chosen by a survival-based criterion (log-rank). Terminal nodes contain patients with similar outcomes.
Right panel: for a new individual, each tree yields a terminal-node Cumulative Hazard Function (CHF) estimated by the Nelson–Aalen method; the ensemble CHF is the average of all trees' CHFs, and survival is recovered as $S(t) = \exp\left(-H(t)\right)$.

We estimated variable importance by permuting each predictor's values in the out-of-bag samples (the portion of data not used in building a particular tree) and observing the impact on prediction errors. Predictors that, when permuted, lead to a large increase in error are considered highly important for the model. This feature is useful for interpreting which variables (clinical or microbial) contribute most to survival prediction in the RSF model. RSF's ability to handle high-dimensional data and automatically model non-linear interactions makes it a valuable approach for our problem.

DeepSurv – Deep neural network for survival analysis

DeepSurv is a deep learning framework for survival analysis that uses a neural network to model the relationship between covariates and survival time [20]. It can be seen as a non-linear extension of the Cox model: instead of assuming a linear effect of covariates on the log-hazard, DeepSurv uses a multi-layer neural network to learn a complex function mapping patient features to a risk score. This allows the model to capture interactions and non-linearities in the data that traditional models might miss. DeepSurv has demonstrated improved performance over traditional Cox models in certain domains, such as cancer prognosis [27].

In our DeepSurv implementation, the network architecture comprised an input layer containing all predictor features (the same features used in other models), multiple fully connected hidden layers, and an output layer producing a single risk score for each participant. We used Scaled Exponential Linear Unit (SeLU) activation functions in the hidden layers due to their beneficial self-normalizing property, which stabilizes the network's outputs, especially useful for handling highly non-linear survival data. The network's output represents the log-risk (log-hazard) for each patient.

The network parameter (θ, including weights and biases) were optimized by maximizing the Cox partial likelihood-equivalently, by minimizing the negative log-partial-likelihood loss function:

$$l(\boldsymbol{\theta}) := -\frac{1}{N_{E=1}} \sum_{i:E_i=1}$$

$$\left(\hat{h}_\theta(x_i) - \log \sum_{j \in \mathfrak{R}(T_i)} e^{\hat{h}_\theta(x_j)} \right) + \lambda_1 \boldsymbol{\theta}_1 + \lambda_2 \boldsymbol{\theta}_2^2.$$

Here, $N_{E=1}$ denotes the number of individuals experiencing the event, $\hat{h}_\theta(x_i)$ is the neural network output representing the risk score (log hazard) for the for the i-th individual, and $\mathfrak{R}(T_i)$ is the set of individuals still at risk at time T_i. The regularization terms ($\lambda_1 \boldsymbol{\theta}_1$ and $\lambda_2 \boldsymbol{\theta}_2^2$) apply L_1 and L_2 penalties, respectively, on the network weights during training. These penalties promote sparsity and weight shrinkage in the network, which helps reduce overfitting and aligns with the regularization approach used in penalized Cox regression. This loss function specifically enables the model to optimize rankings of survival times rather than predicting exact survival durations, consistent with the Cox proportional hazards approach.

We implemented DeepSurv using TensorFlow (Python) and optimized the network parameters with the Adaptive Moment Estimation (Adam) algorithm [28], a first-order gradient-based optimization algorithm that adaptively adjusts the learning rate for each parameter to enable efficient training. Hyperparameter tuning was critical for DeepSurv due to the flexibility and complexity of neural networks.

To address this, we employed Bayesian optimization [29] to efficiently explore the hyperparameter space, including the number of hidden layers, number of neurons per layer, learning rate schedule, dropout rate (used for regularization by randomly dropping units during training), and the strengths of L_1 and L_2 regularization. Bayesian optimization used a Tree-structured Parzen Estimator (TPE) to model the objective function (validation C-index) and identify promising hyperparameter configurations without exhaustively searching the entire space. The final network architecture and hyperparameters were selected based on those that maximized the validation C-index in cross-validation, and this final model was retrained on the full training dataset before evaluation.

Figure 2 illustrates the DeepSurv neural network architecture used for survival analysis. Hyperparameter tuning is performed prior to training using Bayesian optimization, followed by network parameter optimization using the Cox log-partial-likelihood with L_1 and L_2 regularization using the Adam optimizer.

Model Evaluation

We evaluated our models using Harrell's Concordance index (C-index) as the primary measure of discriminative performance [30]. The C-index assesses the model's ability to correctly rank-order patients by risk. It is calculated as the proportion
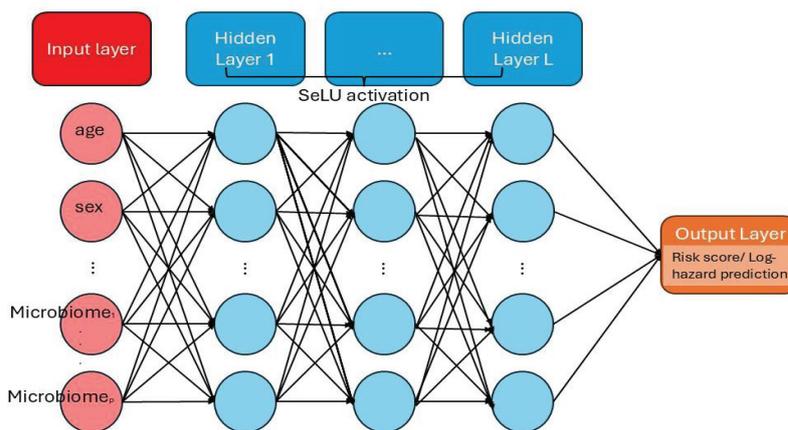
Figure 2 Conceptual architecture of the DeepSurv model. Inputs comprise the nine clinical variables and the selected microbiome features. Hidden layers (SeLU activation) map inputs to a risk score; the network is trained by minimizing the negative Cox partial-likelihood (i.e., maximizing concordance of risk rankings). Regularization (L_1/L_2 penalties and dropout) controls overfitting; parameters are optimized with Adam. Boxes depict layers; arrows depict forward propagation; the single output is a log-risk value. The model's discrimination is evaluated with Harrell's C-index on the held-out test set.

of all comparable patient pairs in which the patient who survived longer had a lower predicted risk score (or equivalently, the model predicts a higher risk for the patient who experienced the event earlier). A C-index of 0.5 indicates random prediction (no better than chance), while 1.0 indicates perfect concordance between predictions and outcomes. In practice, a higher C-index (closer to 1) means better discrimination. We computed the C-index for each model on the external validation set (the DREAM Challenge test dataset) to compare model performances. This external validation helped ensure that our models generalize beyond the training data. All survival analyses and C-index calculations were performed using standard libraries in R (ensuring consistency across models and taking advantage of well-tested routines for concordance calculation).

Results

Feature selection

The FINRISK dataset provided two broad categories of features for analysis: microbiome-derived features and traditional clinical/demographic variables. From the original 5,748 microbiome variables, we identified 125 core features as described in Section 2.2.

Among the nine clinical and demographic variables, univariate Cox regression on the training set ($n - 3471$) revealed seven variables significantly associated with Heart Failure (HF) outcomes ($p < 0.05$): age, BMI, smoking status, blood pressure treatment,

prevalent CHD, systolic blood pressure, and non-HDL cholesterol. These seven were then included in a multivariate Cox model to account for overlapping effects and potential interactions. Using backward elimination (equivalently, retention threshold of $p < .3$), four variables remained as significant independent predictors: age, smoking, prevalent CHD, and non-HDL cholesterol. Multicollinearity among these four variables was low, with all Variance Inflation Factors (VIFs) below 1.13. As a result, these four clinical variables, along with the 125 core microbiome features, were selected as the predictor set for the subsequent advanced machine learning models.

Penalized cox model performance

Using the 129 candidate predictors identified in Section 3.3.1, we trained a Cox proportional hazards model with the elastic net regularization. Cross-validated hyperparameter tuning identified an optimal penalty combination of penalty mix of $\hat{\alpha} = .97$ and $\hat{\lambda} = .00676$, which we then used to refit the model on the full training cohort. The regularized Cox model (elastic net) achieved a Harrell's C-index of 0.7347 on the training set and 0.7225 on the separate test set, demonstrating robust discrimination with minimal overfitting. By comparison, a baseline penalized Cox model incorporating only 9 clinical variables attained C-indices of 0.7187 (training) and 0.7168 (test). Similarly, a baseline Cox model test result provided by DREAM challenge with all 9 covariates was 0.7110.

As a result of the shrinkage and variable selection

process, 14 predictors with non-zero coefficients were retained. These included 2 clinical variables- age and prevalent CHD-and 12 microbiome features corresponding to specific bacterial taxa. The microbial taxa (identified by their index in the feature list and taxonomy) were: Senegalimassilia anaerobia (#9), Adlercreutzia equolifaciens (#10), Bacteroides coprocola (#15), Bacteroides oleiciplenus (#20), Bacteroides salanitronis (#23), Bacteroides uniformis (#27), Paraprevotella xylaniphila (#34), Streptococcus salivarius (#53), Clostridium sp. L2-50 (#58), Roseburia intestinalis (#87), Ruminococcus callidus (#102), and Ruminococcus lactaris (#105). These 14 features represent the most influential predictors of HF survival in our analysis, combining established clinical risk factors with gut microbial species that may play mechanistic roles in HF through inflammation or metabolic pathways.

Accumulating evidence supports the role of gut microbiota in the pathophysiology of Heart Failure (HF) through inflammation, gut barrier function, metabolic dysregularion, and bioactive metabolite production. Senegalimassilia anaerobia taxon belongs to the Coriobacteriaceae family, involved in producing secondary bile acids and influencing cholesterol metabolism, potentially affecting cardiovascular risk profiles [15,31]. Adlercreutzia equolifaciens converts dietary isoflavones to equol, a potent antioxidant. Reduced antioxidant capacity and elevated oxidative stress are implicated in HF progression [13,15,31]. Bacteroides species (coprocola, oleiciplenus, salanitronis, uniformis) produce Short-Chain Fatty Acids (SCFAs) essential for maintaining gut barrier integrity and reducing systemic inflammation, which are crucial in mitigating HF development and exacerbation [13,14,16,32]. Specifically, Bacteroides uniformis has demonstrated anti-inflammatory properties [32]. Reduced abundance could facilitate inflammation, worsening cardiovascular outcomes.

Paraprevotella xylaniphila ferments carbohydrates, producing SCFAs such as butyrate, an important anti-inflammatory metabolite. Lower butyrate levels lead to increased inflammation and potential cardiovascular complications, including HF [9]. Typically, an oral commensal bacterium, abnormal gut colonization by oral bacteria like Streptococcus salivarius indicates dysbiosis and has been associated with systemic inflammation and cardiovascular diseases, including HF [15,16,31]. Clostridium species are known producers of Trimethylamine-N-Oxide (TMAO), a metabolite

strongly associated with increased cardiovascular risk, endothelial dysfunction, and exacerbation of HF [10-12]. Roseburia intestinalis, known for its significant butyrate production, supports gut integrity and modulates inflammation. Its reduced abundance heightens systemic inflammation and cardiovascular risk factors linked to HF progression [9]. Ruminococcus species (callidus, lactaris) ferment dietary fibers, producing beneficial SCFAs, essential for gut homeostasis and reducing inflammation. Decreased abundance may result in increased intestinal permeability, systemic inflammation, and worsening cardiovascular stress [9,13,14].

Random survival forests performance

Next, we used the 9 clinical variables and 12 core microbiomes identified in Section 3.2– to train a Random Survival Forests (RSF) model. During the learning phase, hyperparameters were tuned via a random grid search across several settings: the number of trees (1000, 2000, 3000); the number of features to consider at each split("mtry" = $\sqrt{p}$ and $p/2$), where p is the total number of features (e.g. 21); the minimum terminal node size ("nodesize" = 200:300, by 10) patients per terminal node, and the number random split points per node ("nsplit" = 10:50 by 5). Among these 594 total combinations, the optimal settings were found to be 2,000 trees, $p/2$ - [10.5] features per split, terminal node size of 250, and 25 random split points per node.

Under these settings, the RSF model achieved a C-index of 0.7730 on the training set, and 0.7140 on the test set. The drop in performance from training to test might suggest that, despite the use of out-of-bag error estimation, the RSF model's greater flexibility may have led to some degree of overfitting. On the other hand, the baseline RSF with 9 clinical variables only had C-indices of 0.7371 (train) and 0.7058 (test) with $p/2$ - [4.5], "nodesize" = 250, "ntree" = 1000, and "nsplit" = 10.

The RSF achieved respectable test-set performance with the penalized Cox model by a small margin. To interpret the RSF, we inspected its variable-importance profile (Figure 3): the leading predictors matched those chosen by the Cox model, whereas the bottom five contributed little. To enhance our model, we therefore removed these low-value variables and retrained the RSF using the top 16 features. Updated hyperparameters were $p/2$ - [8], "nodesize" = 260, "ntree" = 2500, and "nsplit" =

1. The final training C-index was 0.7422 and the test C-index was 0.7231.

DeepSurv neural network performance

Our third modeling approach was the DeepSurv neural network. Using the 9 clinical variables and 12 core microbiomes as input, we conducted Bayesian hyperparameter optimization to configure the network architecture and training parameters. The final DeepSurv model included two hidden layers with 64 neurons each, a learning rate of 0.01 (with a decay factor of 0.1 applied during training), an L_1 regularization coefficient of 0.000106, an L_2 regularization coefficient of 0.000217, a dropout rate of 0.3 in the hidden layers, and a momentum of 0.8 for the optimizer. The model was trained until convergence was achieved on the training data (epoch = 205).

The DeepSurv model achieved a C-index of 0.7217 on the training set and 0.7211 on the test set. This nearly identical performance suggests good generalization and minimal overfitting. However, its overall accuracy was slightly lower than that of the elastic net Cox model. The model was implemented in Python using TensorFlow, along with several supporting Python packages to facilitate network construction, survival data processing, and model evaluation.

Model comparison

All three microbiome-enhanced models exceeded the DREAM challenge baseline Cox model with 9 clinical variables only (test C-index = 0.7110), confirming that gut-microbial information adds measurable prognostic value (Table 2). On the held-out test set, RSF achieved the highest C-index (0.7231), followed very closely by the elastic net Cox model (0.7225) and DeepSurv (0.7211). The RSF and elastic net margin was only 0.0006, which is well within typical uncertainty bounds and indicates practical equivalence in discrimination.

The penalized Cox and RSF models showed modest optimism (training - test ΔC-index = 0.0122 and 0.0191, respectively), whereas DeepSurv generalized almost perfectly (training − test ΔC-index = 0.0006). Early stopping and dropout of DeepSurv and elastic
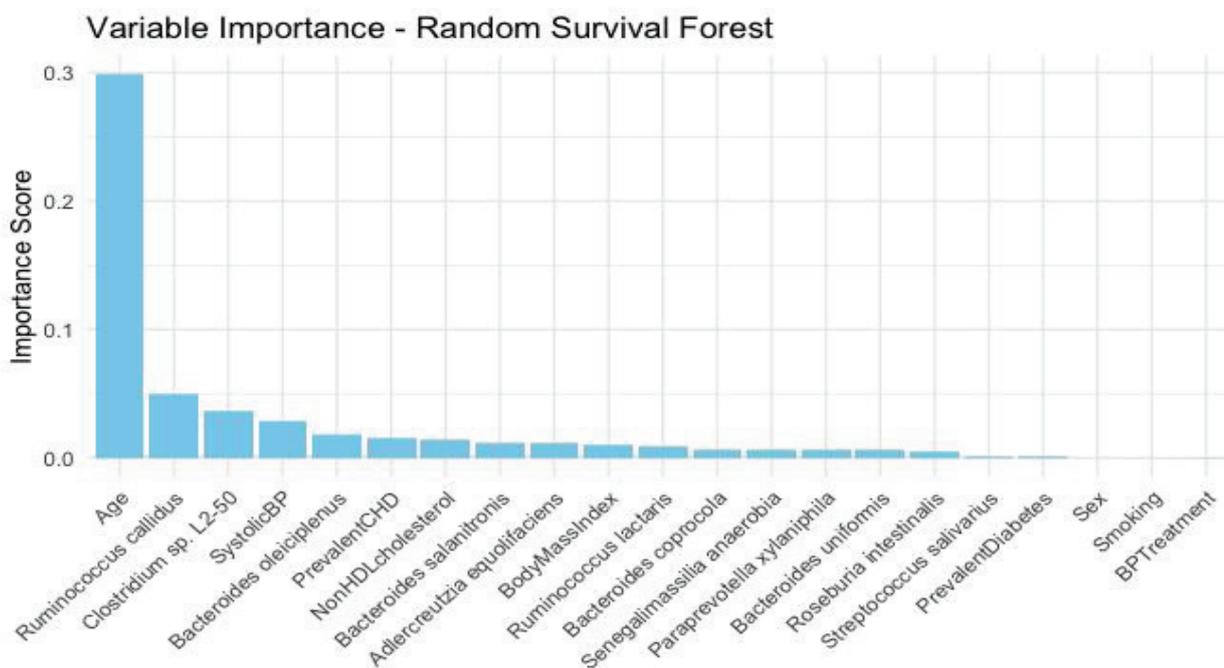

Figure 3 Bar plot of permutation Variable Importance (VIMP) for the final RSF fitted on the development training data and evaluated with Out-Of-Bag (OOB) predictions. For each predictor, values are permuted in OOB samples and the increase in prediction error (relative to the unpermuted model) is averaged across trees to yield VIMP; larger positive bars indicate greater importance. Bars are ordered by VIMP (highest to lowest). Clinical and microbial predictors are distinguished by labels, leading signals. Importance values are relative and not directly interpretable as hazard ratios; they rank predictors' contribution to RSF predictive accuracy. The RSF used in this figure corresponds to the final model reported in Results.

Table 2: Model performance and improvement vs the clinical-only baseline on the held-out test set.

Model	No. predictors	Train C-index	Test C-index	Δ vs baseline	Relative gain, %	P value vs baseline
Random Survival Forests (clinical + microbiome)	16	0.7422	0.7231	+0.0121	1.70	0.003
Elastic-net Cox (clinical + microbiome)	14	0.7347	0.7225	+0.0115	1.62	0.004
DeepSurv (clinical + microbiome)	21	0.7217	0.7211	+0.0101	1.42	0.012
Baseline Cox (clinical-only)	9	–	0.7110	–	–	–

Notes: Harrell's C-index assessed on the DREAM Challenge test partition; Δ and relative gain are calculated against the baseline clinical-only model. *p* values are for differences in concordance versus baseline as reported in the study; top test performance is **bolded.**

net regularization (Cox) effectively constrained over-fitting, while the slightly wider gap for RSF reflects the flexibility of tree ensembles.

Given its near-top discrimination, simpler functional form, and explicit hazard ratios, the relatively strong performance of the elastic net Cox model suggests that it was sufficient to capture the key signals for clinical translation. This result highlights the value of rigorous feature selection and regularization in building predictive models for HF survival. Additionally, the Cox model's interpretability (with hazard ratios for each predictor) is a practical advantage, as it allows clinicians to understand the contribution of each risk factor.

RSF offers not only a strong non-parametric benchmark that corroborates the selected feature set but also can reveal potential interactions and non-linear effects via variable importance measures. The performance of RSF also shows slightly better than the other two models. DeepSurv demonstrates feasibility of complex non-linear neural survival modeling, but additional network complexity does not yield further performance gains in this cohort.

To contextualize our results, it is useful to compare them against the baseline model from the DREAM Challenge. According to the challenge report [22], a Cox model that included all available covariates (clinical and microbiome) without feature selection achieved a C-index of 0.6592 on the same synthetic test dataset. On the other hand, our models substantially outperformed these baseline models. We emphasize that feature selection and regularization materially improved discrimination over the clinical-only baseline on the test set.

On the real-world (hidden) scoring dataset used for the final challenge rankings, the same baseline Cox model achieved a C-index of 0.8236. Given the consistent performance gap between our model and the baseline on the synthetic dataset (~0.012

improvement), it is reasonable to expect that our model would also outperform the baseline on the hidden dataset. Extrapolating from this relative improvement, our approach may be capable of achieving a C-index exceeding 0.83 on real-world data-potentially placing it on par with or above current state-of-the-art models. However, further validation using external patient cohorts is necessary to confirm this hypothesis in a clinical setting.

Discussion

In this study, we developed a comprehensive predictive framework for Heart Failure (HF) survival by integrating gut microbiome data with traditional clinical and demographic risk factors. Using the FINRISK 2002 cohort, we benchmarked three modeling approaches-the elastic net Cox regression, Random Survival Forests (RSF), and DeepSurv neural network. On the held-out test set, RSF achieved the highest C-index (0.7231), closely followed by elastic net Cox (0.7225) and DeepSurv (0.7211). All three were outperformed by the DREAM clinical-only baseline (0.7110).

RSF and the elastic net Cox model showed near-identical discrimination. While nonparametric tree-based RSF achieved the highest C-index, the elastic net Cox model was relatively parsimonious and interpretable and could be highly effective in modeling HF risk. In this context, the more complex machine learning models did not provide a substantial advantage in predictive accuracy, suggesting that parsimonious models with appropriate feature selection may be more suitable for this type of analysis.

An important finding of this work is the identification of a combined set of clinical and microbial predictors for HF survival. Notably, the model that performed best incorporated 14 key features, including conventional factors such as

age and prevalent coronary heart disease, alongside specific gut microbial species (*Senegalimassilia anaerobia, Bacteroides coprocola, Roseburia intestinalis, Streptococcus salivarius*, among others). This reinforces evidence for the gut–heart axis as a relevant component in cardiovascular health. The inclusion of microbiome-derived variables suggests that the composition of an individual's gut flora may influence processes like systemic inflammation and metabolic regulation, which in turn affect cardiovascular outcomes. By refining risk models with these microbial markers, our approach provides a more nuanced risk stratification. Over the long term, these insights could support the development of novel preventive or therapeutic strategies. For instance, if causal relationships are established, interventions aimed at modulating the gut microbiota-such as dietary changes, probiotic, or other approaches-may offer new avenues for improving cardiovascular outcomes.

Limitations/Future Work

Despite the encouraging results, several limitations must be acknowledged. First, our model development and validation were performed on the DREAM-Challenge "synthetic" version of FINRISK 2002. Although that dataset preserves many marginal distributions and correlation structures of the original cohort, it is not a substitute for external validation on independent real-world cohorts. Synthetic datasets lack several sources of real-world variability—batch effects in sequencing, center-specific treatment patterns, missing-value mechanisms, and unmeasured confounders—that can influence both microbiome and clinical variables. The generalizability of our findings needs to be tested on actual patient data from different populations and clinical settings.

Second, some of the advanced models we explored (particularly the DeepSurv) are complex and can be considered "black boxes," making their predictions difficult to interpret in a clinical context. For clinical adoption of any predictive model, interpretability and transparency are crucial so that healthcare providers can understand and trust the risk estimates. This challenge highlights an ongoing tension between model complexity and usability in healthcare AI.

Future research should therefore prioritize two avenues: (1) external validation and calibration of these models using real-world datasets (for instance,

applying the models to other cohort studies or hospital patient data to see if similar performance and key features are observed), and (2) enhancing model interpretability. For the latter, approaches could include using explainable AI techniques for the neural network (such as feature attribution methods), simplifying the deep learning model without greatly sacrificing accuracy, or integrating mechanistic insights (e.g., known pathways linking gut microbes to cardiac function) to constrain and inform the models. Additionally, further investigation into the biological relationships suggested by our model is warranted. For example, exploring why certain microbiome features are associated with HF outcomes could yield insights into disease mechanisms or potential therapeutic targets. Understanding these pathways might involve experimental work or advanced analysis linking microbiome metabolic functions to host physiology in HF patients.

In summary, our study demonstrates the feasibility and value of integrating gut microbiome data into HF survival prediction models. We show that this integration improves predictive performance compared to models based solely on traditional risk factors. The inclusion of microbiome-derived features provides a more comprehensive view of HF risk, reflecting the growing importance of interdisciplinary collaboration between data science and biomedical research.

As precision medicine advances, it becomes increasingly important to bridge sophisticated computational models with real-world clinical practice. Ensuring these models are accurate, interpretable, and clinically validated is essential for their successful adoption. Incorporating diverse biological data-from genomes to microbiomes-into predictive tools has the potential to enable more accurate risk prediction, earlier identification of individuals at risk, and more personalized treatment approaches. This framework can contribute to better outcomes for patients with heart failure and support the broader goal of data-informed, individualized care.

Conclusion

This study presents a robust and interpretable predictive framework for heart failure survival by integrating gut microbiome profiles with conventional clinical and demographic risk factors. Among the evaluated models, the elastic net Cox regression demonstrated superior predictive performance,

reinforcing the effectiveness of parsimonious modeling approaches in high-dimensional, biological complex data settings.

The identification of key microbial taxa as independent predictors, in combination with established clinical variables, demonstrates the potential clinical value of incorporating microbiome data into cardiovascular risk assessment. These findings indicate the relevance of the gut—heart axis and point to promising avenues for future research in heart failure prevention and management.

By demonstrating the feasibility of microbiome-informed predictive modeling in a large population-based cohort, this work contributes to the development of more comprehensive and individualized approaches in cardiovascular risk assessment. Future validation in real-world clinical settings will be necessary to assess its broader applicability and clinical impact.

Author Contributions

H.M. and N.H.N. had complete access to all study data and assumed responsibility for data integrity and analytical accuracy. Conceptualization: H.M.; Methodology: H.M. and N.H.N.; Software: N.H.N.; Validation: H.M., N.H.N., J.P., S.A.; Formal statistical analysis: N.H.N., J.P., S.A., G.A.; Investigation: N.H.N., J.P., S.A., G.A.; Resources: H.M., N.H.N.; Data curation (Reference management): N.H.N.; Writing–Original draft preparation: H.M.

Section attributions

Introduction-H.M., J.P., S.A., G.A.; Methodology-H.M., N.H.N.; Results-H.M. N.H.N., J.P., S.A., G.A.; Discussion and Conclusion-H.M.; Writing–Review and editing, all authors; Visualization, N.H.N., J.P. (Table 1, Figure 3); S.A. (Figure 1); G.A. (Figure 2); N.H.N. S.A. (Table 2); Supervision, H.M.; Project administration, H.M.; Funding acquisition, H.M.

All authors have read and agreed to the published version of the manuscript.

Funding

This research received no external funding.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Data Availability Statement

Data sharing will be available upon reasonable request to the first or second author.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Roth GA, Mensah GA, Johnson CO, Addolorato G, Ammirati E, Baddour LM, Barengo NC, Beaton AZ, Benjamin EJ, Benziger CP, Bonny A, Brauer M, Brodmann M, Cahill TJ, Carapetis J, Catapano AL, Chugh SS, Cooper LT, Coresh J, Criqui M, DeCleene N, Eagle KA, Emmons-Bell S, Feigin VL, Fernández-Solà J, Fowkes G, Gakidou E, Grundy SM, He FJ, Howard G, Hu F, Inker L, Karthikeyan G, Kassebaum N, Koroshetz W, Lavie C, Lloyd-Jones D, Lu HS, Mirijello A, Temesgen AM, Mokdad A, Moran AE, Muntner P, Narula J, Neal B, Ntsekhe M, Moraes de Oliveira G, Otto C, Owolabi M, Pratt M, Rajagopalan S, Reitsma M, Ribeiro ALP, Rigotti N, Rodgers A, Sable C, Shakil S, Sliwa-Hahnle K, Stark B, Sundström J, Timpel P, Tleyjeh IM, Valgimigli M, Vos T, Whelton PK, Yacoub M, Zuhlke L, Murray C, Fuster V; GBD-NHLBI-JACC Global Burden of Cardiovascular Diseases Writing Group. Global Burden of Cardiovascular Diseases and Risk Factors, 1990-2019: Update From the GBD 2019 Study. J Am Coll Cardiol. 2020 Dec 22;76(25):2982-3021. doi: 10.1016/j.jacc.2020.11.010. Erratum in: J Am Coll Cardiol. 2021 Apr 20;77(15):1958-1959. doi: 10.1016/j.jacc.2021.02.039. PMID: 33309175; PMCID: PMC7755038.

2. Shahim B, Kapelios CJ, Savarese G, Lund LH. Global Public Health Burden of Heart Failure: An Updated Review. Card Fail Rev. 2023 Jul 27;9:e11. doi: 10.15420/cfr.2023.05. PMID: 37547123; PMCID: PMC10398425.

3. Bozkurt B, Ahmad T, Alexander KM, Baker WL, Bosak K, Breathett K, Fonarow GC, Heidenreich P, Ho JE, Hsich E, Ibrahim NE, Jones LM, Khan SS, Khazanie P, Koelling T, Krumholz HM, Khush KK, Lee C, Morris AA, Page RL 2nd, Pandey A, Piano MR,

TOPIC(S): CARDIOVASCULAR DISEASES ● CLINICAL CARDIOLOGY ● PUBLIC HEALTH ● EPIDEMIOLOGY

Stehlik J, Stevenson LW, Teerlink JR, Vaduganathan M, Ziaeian B; Writing Committee Members. Heart Failure Epidemiology and Outcomes Statistics: A Report of the Heart Failure Society of America. J Card Fail. 2023 Oct;29(10):1412-1451. doi: 10.1016/j.cardfail.2023.07.006. Epub 2023 Sep 26. PMID: 37797885; PMCID: PMC10864030.

4. Taylor CJ, Ordóñez-Mena JM, Roalfe AK, Lay-Flurrie S, Jones NR, Marshall T, Hobbs FDR. Trends in survival after a diagnosis of heart failure in the United Kingdom 2000-2017: population based cohort study. BMJ. 2019 Feb 13;364:l223. doi: 10.1136/bmj.l223. Erratum in: BMJ. 2019 Oct 8;367:l5840. doi: 10.1136/bmj.l5840. PMID: 30760447; PMCID: PMC6372921.

5. Darvish M, Shakoor A, Feyz L, Schaap J, van Mieghem NM, de Boer RA, Brugts JJ, van der Boon RMA. Heart failure: assessment of the global economic burden. Eur Heart J. 2025 Aug 14;46(31):3069-3078. doi: 10.1093/eurheartj/ehaf323. PMID: 40444781; PMCID: PMC12349942.

6. Kummen M, Mayerhofer CCK, Vestad B, Broch K, Awoyemi A, Storm-Larsen C, Ueland T, Yndestad A, Hov JR, Trøseid M. Gut Microbiota Signature in Heart Failure Defined From Profiling of 2 Independent Cohorts. J Am Coll Cardiol. 2018 Mar 13;71(10):1184-1186. doi: 10.1016/j.jacc.2017.12.057. PMID: 29519360.

7. Cui X, Ye L, Li J, Jin L, Wang W, Li S, Bao M, Wu S, Li L, Geng B, Zhou X, Zhang J, Cai J. Metagenomic and metabolomic analyses unveil dysbiosis of gut microbiota in chronic heart failure patients. Sci Rep. 2018 Jan 12;8(1):635. doi: 10.1038/s41598-017-18756-2. PMID: 29330424; PMCID: PMC5766622.

8. Branchereau M, Burcelin R, Heymes C. The gut microbiome and heart failure: A better gut for a better heart. Rev Endocr Metab Disord. 2019 Dec;20(4):407-414. doi: 10.1007/s11154-019-09519-7. PMID: 31705258.

9. Louis P, Flint HJ. Diversity, metabolism and microbial ecology of butyrate-producing bacteria from the human large intestine. FEMS Microbiol Lett. 2009 May;294(1):1-8. doi: 10.1111/j.1574-6968.2009.01514.x. Epub 2009 Feb 13. PMID: 19222573.

10. Wang Z, Klipfell E, Bennett BJ, Koeth R, Levison BS, Dugar B, Feldstein AE, Britt EB, Fu X, Chung YM, Wu Y, Schauer P, Smith JD, Allayee H, Tang WH, DiDonato JA, Lusis AJ, Hazen SL. Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. Nature. 2011 Apr 7;472(7341):57-63. doi: 10.1038/nature09922. PMID: 21475195; PMCID: PMC3086762.

11. Organ CL, Otsuka H, Bhushan S, Wang Z, Bradley J, Trivedi R, Polhemus DJ, Tang WH, Wu Y, Hazen SL, Lefer DJ. Choline Diet and Its Gut Microbe-Derived Metabolite, Trimethylamine N-Oxide, Exacerbate Pressure Overload-Induced Heart Failure. Circ Heart Fail. 2016 Jan;9(1):e002314. doi: 10.1161/CIRCHEARTFAILURE.115.002314. Epub 2015 Dec 23. PMID: 26699388; PMCID: PMC4943035.

12. Zhu W, Gregory JC, Org E, Buffa JA, Gupta N, Wang Z, Li L, Fu X, Wu Y, Mehrabian M, Sartor RB, McIntyre TM, Silverstein RL, Tang WHW, DiDonato JA, Brown JM, Lusis AJ, Hazen SL. Gut Microbial Metabolite TMAO Enhances Platelet Hyperreactivity and Thrombosis Risk. Cell. 2016 Mar 24;165(1):111-124. doi: 10.1016/j.cell.2016.02.011. Epub 2016 Mar 10. PMID: 26972052; PMCID: PMC4862743.

13. Marques FZ, Mackay CR, Kaye DM. Beyond gut feelings: how the gut microbiota regulates blood pressure. Nat Rev Cardiol. 2018 Jan;15(1):20-32. doi: 10.1038/nrcardio.2017.120. Epub 2017 Aug 24. PMID: 28836619.

14. Witkowski M, Weeks TL, Hazen SL. Gut Microbiota and Cardiovascular Disease. Circ Res. 2020 Jul 31;127(4):553-570. doi: 10.1161/CIRCRESAHA.120.316242. Epub 2020 Jul 30. PMID: 32762536; PMCID: PMC7416843.

15. Brown JM, Hazen SL. Microbial modulation of cardiovascular disease. Nat Rev Microbiol. 2018 Mar;16(3):171-181. doi: 10.1038/nrmicro.2017.149. Epub 2018 Jan 8. PMID: 29307889; PMCID: PMC5885760.

16. Kelly TN, Bazzano LA, Ajami NJ, He H, Zhao J, Petrosino JF, Correa A, He J. Gut Microbiome Associates With Lifetime Cardiovascular Disease Risk Profile Among Bogalusa Heart Study Participants. Circ Res. 2016 Sep 30;119(8):956-64. doi: 10.1161/CIRCRESAHA.116.309219. Epub 2016 Aug 9. PMID: 27507222; PMCID: PMC5045790.

17. Rajkomar A, Dean J, Kohane I. Machine Learning in Medicine. N Engl J Med. 2019 Apr 4;380(14):1347-1358. doi: 10.1056/NEJMra1814259. PMID: 30943338.

18. Zou H, Hastie T. Regularization and variable selection via the elastic net. J R Stat Soc Series B Stat Methodol. 2005;67:301-20. doi: 10.1111/j.1467-9868.2005.00503.x.

19. Ishwaran H, Kogalur UB, Blackstone EH, Lauer MS. Random survival forests. Ann Appl Stat. 2008;2(3):841-60. doi: 10.1214/08-AOAS169.

20. Katzman JL, Shaham U, Cloninger A, Bates J, Jiang T, Kluger Y. DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network. BMC Med Res Methodol. 2018 Feb 26;18(1):24. doi: 10.1186/s12874-018-0482-1. PMID: 29482517; PMCID: PMC5828433.

21. Borodulin K, Tolonen H, Jousilahti P, Jula A, Juolevi A, Koskinen S, Kuulasmaa K, Laatikainen T, Männistö S, Peltonen M, Perola M, Puska P, Salomaa V, Sundvall J, Virtanen SM, Vartiainen E. Cohort Profile: The National FINRISK Study. Int J Epidemiol. 2018 Jun 1;47(3):696-696i. doi: 10.1093/ije/dyx239. PMID: 29165699.

22. Erawijantari PP, Kartal E, Liñares-Blanco J, Laajala TD, Feldman LE; FINRISK Microbiome DREAM Challenge and ML4Microbiome Communities; Carmona-Saez P, Shigdel R, Claesson MJ, Bertelsen RJ, Gomez-Cabrero D, Minot S, Albrecht J, Chung V, Inouye M, Jousilahti P, Schultz JH, Friederich HC, Knight R, Salomaa V, Niiranen T, Havulinna AS, Saez-Rodriguez J, Levinson RT, Lahti L. Microbiome-based risk prediction in incident heart failure: a community challenge. medRxiv [Preprint]. 2023 Oct 12:2023.10.12.23296829. doi: 10.1101/2023.10.12.23296829. PMID: 37873403; PMCID: PMC10593042.

23. Cox DR. Regression models and life-tables. J R Stat Soc. 1972;34(2):187-220. doi: 10.1111/j.2517-6161.1972.tb00899.x

24. Breiman L. Random forests. Mach Learn. 2001;45(1):5-32. doi: 10.1023/A:1010933404324.

25. Nelson W. Theory and applications of hazard plotting for censored failure data. Technometrics. 1969;11(3):945-65.

26. Aalen OO. Nonparametric inference for a family of counting processes. Ann Stat. 1978;6(4):701-26.

27. Ren Z, Zhu J, Wang Z, Li X, Hu W. Deep learning-based survival analysis for cancer prognosis. IEEE Trans Neural Netw Learn Syst. 2019;30(9):2739-48. doi: 10.1186/s12920-020-0686-1.

28. Kingma DP, Ba J. Adam.A method for stochastic optimization. arXiv. 2014. doi: 10.48550/arXiv.1412.6980.

29. Snoek J, Larochelle H, Adams RP. Practical Bayesian optimization of machine learning algorithms. Adv Neural Inf Process Syst. 2012;25:2951-2959. doi: 10.48550/arXiv.1206.2944.

30. Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med. 1996 Feb 28;15(4):361-87. doi: 10.1002/(SICI)1097-0258(19960229)15:4<361::AID-SIM168>3.0.CO;2-4. PMID: 8668867.

31. Tang WH, Kitai T, Hazen SL. Gut Microbiota in Cardiovascular Health and Disease. Circ Res. 2017 Mar 31;120(7):1183-1196. doi: 10.1161/CIRCRESAHA.117.309715. PMID: 28360349; PMCID: PMC5390330.

32. Din AU, Hassan A, Zhu Y. Inhibitory effect of Bacteroides uniformis on inflammation and metabolic disorder. Front Immunol. 2021;12:653948.