
Predicting Post-Discharge Foot Complications for People with Diabetes in Ontario using Interpretable Deep Survival Model for Competing Risks

Dhanesh Ramachandram¹

Anne Löffler^{2,3}

Surain Roberts^{2,3}

Maia Norman¹

Charles de Mestral^{2,5}

Conrad Pow⁵

Fahad Razak^{2,3,4}

Amol Verma^{2,3,4}

¹Vector Institute ²Li Ka Shing Knowledge Institute, Unity Health Toronto – St. Michael's Hospital
³GEMINI

⁴Dept. of Medicine, University of Toronto ⁵Diabetes Action Canada

Abstract

Model interpretability is crucial for establishing AI safety and clinician trust in medical applications for example, in survival modelling with competing risks. Recent deep learning models have attained very good predictive performance but their limited transparency, being black-box models, hinders their integration into clinical practice. To address this gap, we propose an intrinsically interpretable survival model called CRISPNAM-FG. Leveraging the structure of Neural Additive Models (NAMs) with separate projection vectors for each risk, our approach predicts the Cumulative Incidence Function using the Fine-Gray formulation, achieving high predictive power with intrinsically transparent and auditable predictions. We validated the model on several benchmark datasets and applied our model to predict future foot complications in diabetic patients across 29 Ontario hospitals (2016-2023). Our method achieves competitive performance compared to other deep survival models while providing transparency through shape functions and feature importance plots. Code is available at <https://github.com/VectorInstitute/crispnam-FG>

1 Introduction

While traditional survival analysis focused on predicting risk for a single event of interest, many real-world problems, particularly in the healthcare domain, involve more than one “competing” events of interest. For example, patients who are at risk of death due to cancer may die following other unrelated complications. In this case, death due to cancer is the primary risk while death due to other causes is the competing risk that must be taken into account when predicting the primary risk. Traditionally, competing risks have been addressed using statistical methods such as the Fine-Gray model [5] and cause-specific hazard models, which rely on strong parametric assumptions about the relationship between covariates and risk. While these classical approaches offer interpretability, they are limited in their ability to capture complex, non-linear patterns in the data. Recent advances in deep learning have demonstrated improved predictive performance for competing risks scenarios [10], but these gains come at the cost of interpretability, resulting in black-box models that are difficult for clinicians to trust and validate in practice. In this work, we present an intrinsically interpretable

survival model called CRISPNAM-FG (Competing Risks Survival Prediction using Neural Additive Models: Fine Gray) that handles competing risks using the Fine-Gray formulation while maintaining the predictive power of deep learning and the interpretability of classical statistical model

The remainder of this paper proceeds as follows. We begin by reviewing related work on deep neural network-based competing risks survival models, outlining key research gaps. Next, Section 3 provides the mathematical background and details our interpretable model architecture and its design principles. Section 4 describes the benchmark and GEMINI datasets, along with the procedures for model training, optimization, and evaluation. Finally, Section 5 presents our results and discussion.

2 Related Work

Conventional survival analysis approaches, including the Kaplan-Meier estimator and Cox proportional hazards model [3], are inappropriate for competing risk scenarios because they erroneously handle alternative events as censored observations, resulting in inflated estimates of event probabilities.

Several deep learning methodologies have been developed to handle competing risks in survival analysis. DeepHit [10] presents a joint modeling framework that employs a shared representation network coupled with cause-specific sub-networks to model the joint distribution of event time and type. The model optimization combines negative log-likelihood with ranking loss to ensure concordance between predicted and observed outcomes. Neural Fine Gray [8] extends the classical Fine-Gray sub-distribution model through neural networks, enabling non-linear covariate relationships while preserving direct cumulative incidence function estimation capabilities.

Despite these methodological advances, existing deep survival approaches for competing risks [8, 10] lacks inherent interpretability, particularly at the feature level. This opacity makes it challenging to understand individual feature contributions to risk predictions across different competing events. Researchers thus have relied upon post-hoc explanations such as SHAP [11]. Another example is SurvNAM [15], a post-hoc explanation method that approximates the predictions of a pre-trained black-box survival model (such as Random Survival Forest) by fitting a GAM-extended Cox PH model using NAMs as surrogate learners. Post hoc explanations, generally suffer from low fidelity, leading to unfaithful or misleading explanations.

Recently, glass box models or intrinsically interpretable models for survival analysis have been proposed. CoxNAM [17] embeds Neural Additive Models [1] (NAMs) directly within the Cox framework, enabling end-to-end learning of nonlinear feature effects from survival data. Each feature is processed by its own neural network, and the results are summed to obtain the hazard function value. Interpretability is provided through shape functions and feature importance plots, however, this model is limited to single risk predictions. Inspired by CoxNAM, an intrinsically interpretable model called CRISPNAM [13] was introduced to support competing risks through the *cause-specific* framework. This model uses separate projection functions for each risk-feature pair to jointly learn all cause-specific hazards in a single framework, enabling features to have different effects across competing events while avoiding the need to treat competing events as censoring or fit separate models. However, cause-specific approaches impose independence assumptions and can produce biased cumulative incidence estimates when competing events are dependent. To address these limitations, we propose CRISPNAM-FG, which adapts the CRISPNAM architecture to implement the Fine-Gray formulation [5], enabling direct modeling of cumulative incidence functions while preserving the interpretability and computational efficiency of the original framework.

3 Methodology

3.1 Competing Risks in Survival Analysis

There are two main methodologies for estimation in competing risks survival analysis, where subjects can experience one of K mutually exclusive event types, with $E \in \{0, 1, 2, \dots, K\}$ where $E = 0$ denotes censoring.

3.1.1 The Cause-Specific Hazard Method

Proposed by Prentice et al. [12], this method models competing events separately, treating each outcome as a distinct hazard function. Subjects experiencing competing events are removed from the risk set (treated as censored). For each cause k , the cause-specific hazard $\lambda_k(t|\mathbf{x})$ represents the instantaneous rate of event type k at time t :

$$\lambda_k(t|\mathbf{x}) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t, E = k \mid T \geq t, \mathbf{x})}{\Delta t} \quad (1)$$

where $\mathbf{x}$ represents covariates, T the event time, and the conditioning on $T \geq t$ indicates the subject has survived event-free until time t .

3.1.2 Fine-Gray Subdistribution Method

This method directly models the subdistribution hazard, maintaining subjects who experience competing events in the risk set. The subdistribution risk set for event k at time t is:

$$\mathcal{R}_k^{\text{sub}}(t) = \{j : T_j \geq t\} \cup \{j : T_j < t \text{ and } E_j \neq k \text{ and } E_j \neq 0\} \quad (2)$$

This includes both subjects event-free at time t and those who experienced a competing event before time t .

The subdistribution hazard is defined as:

$$\lambda_k^{\text{sub}}(t|\mathbf{x}) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t, E = k \mid T \geq t \text{ or } (T < t \text{ and } E \neq k), \mathbf{x})}{\Delta t} \quad (3)$$

The relationship between cause-specific and subdistribution hazards is:

$$\lambda_k^{\text{sub}}(t|\mathbf{x}) = \lambda_k(t|\mathbf{x}) \cdot r_k(t|\mathbf{x}) \quad (4)$$

where $r_k(t|\mathbf{x}) = P(T \geq t \mid T \geq t \text{ or } E \neq k) / P(T \geq t)$ is the reduction factor.

3.2 Model Architecture

3.2.1 Neural Additive Model (FeatureNet)

In line with the Neural Additive Model framework, each input feature x_i is processed by its own dedicated neural network $f_i(\cdot)$, referred to here as a *FeatureNet*. These feature-specific sub-networks are designed to learn the non-linear contribution of each individual feature to the overall risk score, while preserving interpretability by isolating feature effects.

$$\mathbf{h}_i = f_i(x_i) \in \mathbb{R}^d \quad (5)$$

where d is the dimension of the hidden representation. Each FeatureNet is a fully-connected feedforward neural network with L layers, taking the scalar input x_i and producing a hidden representation $\mathbf{h}_i \in \mathbb{R}^d$. The activations are computed recursively using the hyperbolic tangent function:

$$\mathbf{z}_i^{(l)} = \tanh(W_i^{(l)} \mathbf{z}_i^{(l-1)} + b_i^{(l)}), \quad \text{for } l = 1, \dots, L, \quad (6)$$

where $\mathbf{z}_i^{(0)} = x_i$, and $\mathbf{h}_i = \mathbf{z}_i^{(L)}$ denotes the output of the final layer. Each layer l has weights $W_i^{(l)} \in \mathbb{R}^{d_l \times d_{l-1}}$ and biases $b_i^{(l)} \in \mathbb{R}^{d_l}$, with $d_0 = 1$ and $d_L = d$.

In this implementation, Dropout is used with rate p_{dropout} after each hidden layer, Feature dropout with rate p_{feature} during training to increase robustness and an optional batch normalization layer after each linear transformation to stabilize learning especially for deeper FeatureNets.

3.3 Risk-Specific Projections

For each feature i and competing risk k , a separate linear projection transforms the feature representation to its contribution to the log-hazard ratio. To address scale ambiguities across different competing risks and ensure fair comparison of feature contributions, we constrain the projection vectors to have unit L2 norm.

The risk-specific projection is defined as:

$$g_{i,k}(\mathbf{h}_i) = \tilde{\mathbf{w}}_{i,k}^T \mathbf{h}_i \quad (7)$$

where $\tilde{\mathbf{w}}_{i,k}$ is the L2-normalized projection vector:

$$\tilde{\mathbf{w}}_{i,k} = \frac{\mathbf{w}_{i,k}}{\|\mathbf{w}_{i,k}\|_2 + \epsilon} \quad (8)$$

with $\mathbf{w}_{i,k} \in \mathbb{R}^d$ being the learnable weight vector for projection $i \in \{1, 2, \dots, p\}$ and risk $k \in \{1, 2, \dots, K\}$, and $\mathbf{h}_i \in \mathbb{R}^d$ being the feature representation.

This normalization constraint ensures that $\|\tilde{\mathbf{w}}_{i,k}\|_2 = 1$ for all feature-risk pairs, which constraints all projection vectors to operate on the same scale and enabling direct comparison of feature importance across different competing risks.

3.3.1 Additive Risk Aggregation

The log-hazard ratio for risk k given input features $\mathbf{x} = [x_1, x_2, \dots, x_p]$ is computed as the sum of individual feature contributions:

$$\eta_k(\mathbf{x}) = \sum_{i=1}^p g_{i,k}(f_i(x_i)) \quad (9)$$

This preserves the additive nature of the model while allowing for complex non-linear feature effects.

3.3.2 Fine-Gray Subdistribution Hazards

The Fine-Gray approach directly models the subdistribution hazard, which has a more direct relationship with the cumulative incidence function. For a subject with covariates $\mathbf{x}$, the subdistribution hazard for event type k is parameterized as:

$$\lambda_k^{\text{sub}}(t|\mathbf{x}) = \lambda_{0k}^{\text{sub}}(t) \exp(\eta_k(\mathbf{x})) \quad (10)$$

where $\lambda_{0k}^{\text{sub}}(t)$ is the baseline subdistribution hazard.

3.3.3 Fine-Gray Partial Likelihood Loss

The Fine-Gray negative log partial likelihood differs significantly from the cause-specific approach in its risk set construction. For event type k , the loss function is:

$$\mathcal{L}_k^{\text{FG}} = - \sum_{i: E_i=k} \left[\eta_k(\mathbf{x}_i) - \log \left(\sum_{j \in \mathcal{R}_k^{\text{sub}}(T_i)} w_j(T_i) \exp(\eta_k(\mathbf{x}_j)) \right) \right] \quad (11)$$

where $w_j(T_i)$ are inverse probability of censoring weights:

$$w_j(T_i) = \frac{I(C_j \geq T_i)}{\hat{G}(T_i)} \quad (12)$$

with $\hat{G}(t)$ being the Kaplan-Meier estimate of the censoring distribution.

For numerical stability and to handle class imbalance, we implement a weighted version:

$$\mathcal{L}^{\text{FG}} = \sum_{k=1}^K \omega_k \mathcal{L}_k^{\text{FG}} + \gamma \|\Theta\|_2^2 \quad (13)$$

where $\omega_k = n/(K \cdot n_k)$ for class balancing, and $\gamma \geq 0$ is the L2 regularization coefficient controlling the strength of weight decay.

3.3.4 Cumulative Incidence Function Estimation

Following Fine-Gray methodology, the cumulative incidence function for event k is:

$$F_k(t|\mathbf{x}) = 1 - \exp\left(-\int_0^t \lambda_k^{\text{sub}}(s|\mathbf{x})ds\right) \quad (14)$$

In practice, using the Breslow-type estimator:

$$\hat{F}_k(t|\mathbf{x}) = 1 - \left[1 - \hat{F}_{0k}(t)\right]^{\exp(\eta_k(\mathbf{x}))} \quad (15)$$

where $\hat{F}_{0k}(t)$ is the baseline cumulative incidence function estimated from:

$$\hat{F}_{0k}(t) = \sum_{i: T_i \leq t, E_i = k} \frac{d_i}{\sum_{j \in \mathcal{R}_k^{\text{sub}}(T_i)} \exp(\eta_k(\mathbf{x}_j))} \quad (16)$$

with d_i being the number of events at time T_i .

3.4 Interpretability Mechanisms

To facilitate interpretability of the model, *shape functions plots* can be generated to visualize the contribution of each feature to the prediction. Specifically, for each feature i and risk k , we can compute a shape function that describes how the feature affects the log-hazard ratio.

$$s_{i,k}(x_i) = g_{i,k}(f_i(x_i)) \quad (17)$$

The importance of feature i for risk k is quantified by the mean absolute value of its contribution across the dataset.

$$\mathcal{I}_{i,k} = \frac{1}{N} \sum_{j=1}^N |s_{i,k}(x_{ij})| \quad (18)$$

Note that the feature importance metric $\mathcal{I}_{i,k}$ measures only the **magnitude** of a feature's influence on the competing risk, not its **direction**. Since $\mathcal{I}_{i,k}$ is computed using absolute values of the shape functions, a high importance score indicates that feature i has a strong effect on risk k , but does not reveal whether this effect is protective (decreasing risk) or harmful (increasing risk). To understand the actual risk relationship, one must examine the shape function plots $s_{i,k}(x_i)$ directly, where positive values indicate risk-increasing effects and negative values indicate risk-reducing effects. This separation of effect magnitude (captured by $\mathcal{I}_{i,k}$) from effect direction (shown in $s_{i,k}$) enables practitioners to first identify which features matter most for risk stratification, and then understand how these influential features actually modulate patient risk.

4 Experiments

We first validated our architecture by comparing the performance of our proposed model against two other black box models for competing risks prediction namely Neural Fine-Gray [8] and DEEP-HIT [10] using 3 benchmark medical datasets, namely Primary Biliary Cholangitis (PBC) dataset [14], Framingham Heart Study dataset [9] and SUPPORT2 dataset [7].

One of this paper's key contribution is applying the interpretable CRISPNAM-FG survival model to predict the probability of future foot complications in diabetes patients discharged from General Internal Medicine services at 29 Ontario hospitals. Among the many complications that diabetic patients experience, diabetic foot complications represent a major concern, with individuals facing more than a 30% lifetime risk of developing such complications and a 40% recurrence rate within one year following wound healing [16]. Research demonstrates that annual foot screening can reduce both treatment costs and amputation rates [2]. While primary care data provides the optimal source for patient monitoring, hospitalization records from General Internal Medicine departments serve as strong predictors for identifying diabetic patients at risk of developing foot complications.

In this work, a foot complication was defined as the first hospitalization for a foot ulcer, infection, gangrene, or (toe,foot,leg) amputation. Candidate predictors were specified a priori and on the basis of literature [4] describing the risk of foot complications in patients with diabetes. The dataset used for this work was obtained from GEMINI [6], a non-profit organization that systematically collects administrative and clinical data from hospital information systems across Ontario. During the study period (April 2015-March 2023), GEMINI comprised 32 hospitals representing 50 – 60% of medical and intensive care unit beds across Ontario, Canada. The network incorporates administrative data from the Canadian Institute for Health Information’s Discharge Abstract Database and National Ambulatory Care Reporting System, alongside clinical data extracted from hospital-specific electronic medical records and mapped by clinical experts.

Of the 32 GEMINI hospitals, three hospitals were excluded: one lacked hemoglobin A1C data (a key model predictor), and two specialized cancer and rehabilitation facilities did not align with our target population. The cohort of index hospitalizations was comprised of all adult patients with diabetes discharged from general internal medicine between April 2016 and March 2023. Patients with diabetes were identified based on the presence of any ICD-10-CA code for diabetes (E10 – E14) in emergency department or inpatient records. Only patients discharged home were included to ensure eligibility for ambulatory foot screening.

Patients with recorded age > 105 years and those without a valid Ontario Health Insurance Plan (OHIP) number were excluded (the latter being required for linkage of different encounters to the same patient). Any encounters with a current foot complication (defined in Risk Calculator Outcome section below) were also excluded from the cohort of index admissions. Given that GEMINI data provided a 1-year lookback period, hospitalizations preceded by a foot complication in the prior year were excluded. Encounters from time periods with systematically missing laboratory data and those discharged during the final 30 days of the study period were removed since GEMINI data could not capture potential outcome encounters for these patients. When multiple eligible encounters existed for a single patient, the first hospitalization over the study period was selected as the index hospitalization.

4.0.1 Outcome and Feature Selection

Table 1 summarizes the study methods and features that were selected to train the risk prediction model.

Table 1: Study Methods Summary

Component	Description
Primary Outcome	
Definition	Time to first foot complication
Complications	Ulceration, infection, gangrene, amputation
Identification	Validated ICD-10-CA diagnosis and procedure codes
Setting	29 GEMINI hospitals
Timing	Index hospitalization to first complication admission
Censoring	Last available date for patients without complications
Competing risk	In-hospital death
Predictors	
Demographics	Age (continuous years), sex (female vs. non-female)
Clinical	Admission urgency (elective vs. urgent), homelessness
Laboratory	HbA1C, creatinine, albumin (last recorded values)
Comorbidities	Peripheral artery disease, coronary artery disease, stroke, congestive heart failure, hypertension, chronic obstructive pulmonary disease, chronic kidney disease, malignancy, mental illness
Coding	Binary variables from ICD-10-CA codes (index hospitalization)
Exclusions	Hospital-level factors excluded

Three laboratory variables contained missing values: albumin, creatinine, and HbA1C. Laboratory test ordering depends on physicians’ clinical judgment regarding the potential utility of abnormal results and the availability of recent outpatient values, creating informative missingness patterns. To address this issue, we implemented the missing indicator method (MIM) whereby missing laboratory

values were replaced with -1, and corresponding binary missingness indicators were created for each affected variable.

4.1 Experimental Setup

Our CRISPNAM-FG model is implemented using PyTorch. For benchmark datasets, we implemented nested K-fold cross validation with hyperparameter tuning. However, for the large GEMINI dataset, we held out 10% of each dataset for hyperparameter optimization (learning rate, L_2 regularization strength, dropout rates, network architecture (1-3 hidden layers with 8-128 units), and batch normalization) using validation loss as the objective. We then performed 5-fold cross validation to obtain robust and unbiased estimate of the model’s performance on the remaining 90% of each datasets. Continuous features were normalized using standard scaling ($\mu = 0, \sigma = 1$) and categorical features were one-hot encoded. For the benchmark datasets, missing categorical values were imputed using mode imputation. For continuous variables, mean imputation was used. For missing lab values from the GEMINI dataset, we implemented the missing indicator method. Training employed the *AdamW* optimizer to minimize the negative log-likelihood loss with a batch size of 256 and early stopping (patience=10) to prevent over-fitting.

Model performance was assessed using complementary discrimination and calibration metrics summarized in Table 2.

Table 2: Performance metrics for model discrimination and calibration assessment

Metric	Description	Interpretation
TD-AUC	Discrimination ability: quantifies how well the model ranks subjects by risk	Values range from 0.5 (no better than chance) to 1.0 (perfect discrimination)
TD-CI	Discriminative ability that accounts for temporal changes in model performance	Values range from 0.5 to 1.0, with higher values indicating better discriminative ability; considers that model performance can evolve as time progresses
Brier Score	Accuracy of probabilistic predictions; penalizes errors in both discrimination and calibration	Lower values indicate better performance

5 Results and Discussion

Benchmark performance results are presented in Table 3. We compared our CRISPNAM-FG model performance against two other models namely the Neural Fine-Gray [8] (NFG) and DEEPHIT [10]. The comparative analysis reveals distinct performance patterns between CRISPNAM-FG, DEEPHIT, and NFG models across the three competing risks datasets.

CRISPNAM-FG demonstrates superior discriminative performance, consistently achieving the highest TD-AUC and TD-CI values across most dataset-risk combinations, particularly excelling in the FHS dataset where it outperforms alternatives across all quantiles for both discrimination metrics. However, this enhanced discriminative ability comes at the cost of calibration performance, as evidenced by systematically higher Brier scores compared to DEEPHIT. DEEPHIT exhibits a contrasting profile, consistently producing the lowest Brier scores across all datasets and risks, indicating superior probabilistic calibration, while maintaining competitive but generally lower discriminative performance than CRISPNAM-FG. The interpretability of CRISPNAM-FG is evidenced through the model generated shape functions and feature importance plots for each of these benchmark datasets, which are provided in the Appendix section.

5.1 GEMINI Dataset Results

As shown in Table 4, for the GEMINI dataset, the CRISPNAM-FG model shows impressive results for both discriminatory metrics (TD-AUC, TD-CI). While the Brier score for the competing risk is poorer compared to other models, its calibration metric is very good for the primary risk of interest.

Figure 1 illustrates the shape functions for 9 significant features contributing to model predictions. The presence of comorbidities, specifically coronary artery disease (*coronary_ad*), shows increased

Table 3: Comparative 5-fold performance metrics for various models across multiple competing risks datasets.

Dataset	Model	Risk	TD-AUC			TD-CI			Brier Score		
			$q_{.25}$	$q_{.50}$	$q_{.75}$	$q_{.25}$	$q_{.50}$	$q_{.75}$	$q_{.25}$	$q_{.50}$	$q_{.75}$
FHS	CRISPNAM-FG	1	0.856±0.025	0.832±0.007	0.812±0.010	0.770±0.026	0.755±0.008	0.745±0.007	0.122±0.102	0.202±0.141	0.249±0.153
		2	0.812±0.049	0.789±0.016	0.776±0.019	0.727±0.054	0.716±0.019	0.714±0.021	0.044±0.011	0.092±0.027	0.143±0.038
	NFG	1	0.678±.160	0.673±.149	0.666±.133	0.632±.150	0.629±.141	0.628±.134	0.058±.005	0.102±.005	0.134±.007
		2	0.617±.126	0.629±.117	0.620±.136	0.612±.156	0.611±.148	0.610±.152	0.041±.003	0.077±.006	0.113±.009
	DEEPHIT	1	0.854±.019	0.831±.012	0.807±.009	0.738±.030	0.729±.018	0.724±.017	0.056±.006	0.101±.004	0.134±.004
		2	0.796±.053	0.779±.028	0.776±.031	0.737±.030	0.728±.018	0.724±.017	0.038±.003	0.072±.004	0.109±.005
SUP	CRISPNAM-FG	1	0.880±0.031	0.848±0.027	0.848±0.031	0.857±0.026	0.830±0.017	0.833±0.018	0.193±0.120	0.307±0.161	0.329±0.162
		2	0.947±0.043	0.877±0.091	0.832±0.112	0.894±0.007	0.829±0.005	0.793±0.006	0.144±0.035	0.251±0.075	0.312±0.100
	NFG	1	0.790±.036	0.777±.023	0.797±.020	0.620±.038	0.513±.025	0.438±.025	0.044±.003	0.100±.005	0.115±.005
		2	0.902±.004	0.840±.003	0.812±.005	0.847±.008	0.741±.007	0.701±.010	0.092±.004	0.151±.004	0.176±.002
	DEEPHIT	1	0.779±.205	0.640±.222	0.665±.188	0.746±.013	0.540±.010	0.497±.011	0.039±.003	0.089±.002	0.101±.003
		2	0.955±.025	0.893±.069	0.860±.083	0.863±.008	0.745±.005	0.697±.006	0.087±.005	0.144±.006	0.171±.003
PBC	CRISPNAM-FG	1	0.992±0.006	0.968±0.023	0.956±0.028	0.838±0.029	0.813±0.016	0.796±0.015	0.199±0.087	0.249±0.089	0.270±0.080
		2	0.980±0.029	0.986±0.012	0.988±0.010	0.834±0.029	0.857±0.027	0.868±0.024	0.121±0.054	0.179±0.072	0.190±0.066
	NFG	1	0.853±.023	0.835±.011	0.824±.027	0.782±.017	0.765±.013	0.756±.016	0.143±.020	0.152±.008	0.170±.013
		2	0.491±.055	0.491±.055	0.537±.056	0.227±.022	0.238±.011	0.245±.016	0.092±.054	0.135±.073	0.164±.082
	DEEPHIT	1	0.994±.008	0.965±.028	0.959±.034	0.822±.016	0.796±.019	0.776±.013	0.100±.009	0.130±.004	0.152±.011
		2	0.978±.036	0.983±.027	0.987±.020	0.728±.024	0.705±.019	0.690±.016	0.037±.006	0.056±.003	0.065±.004

Notes: Dataset = (FHS: Framingham Heart Study, SUP: SUPPORT2, PBC: Primary Biliary Cirrhosis); Model= (CRISPNAM-FG: CRISPNAM -Fine Gray, NFG: Neural Fine Gray, DEEPHIT: DeepHit); Risk = (1: Primary, 2: Competing)

Table 4: Comparative 5-fold performance metrics between CRISPNAM-FG, DEEPHIT and Neural Fine Gray models for the GEMINI dataset.

Dataset	Model	Risk	TD-AUC			TD-CI			Brier Score		
			$q_{.25}$	$q_{.50}$	$q_{.75}$	$q_{.25}$	$q_{.50}$	$q_{.75}$	$q_{.25}$	$q_{.50}$	$q_{.75}$
GEMINI	CRISPNAM-FG	1	0.766±.042	0.763±.047	0.763±.052	0.711±.018	0.702±.024	0.702±.022	0.013±.002	0.024±.004	0.033±.007
		2	0.739±.057	0.726±.058	0.707±.060	0.720±.004	0.707±.005	0.690±.011	0.280±.186	0.372±.218	0.402±.212
	NFG	1	0.518±.033	0.524±.035	0.508±.049	0.511±.032	0.515±.033	0.500±.044	0.067±.004	0.111±.004	0.154±.007
		2	0.440±.005	0.396±.012	0.367±.025	0.550±.014	0.532±.015	0.529±.020	0.016±.002	0.034±.004	0.054±.005
	DEEPHIT	1	0.697±.010	0.689±.014	0.684±.019	0.709±.011	0.704±.006	0.698±.009	0.012±.001	0.019±.002	0.089±.130
		2	0.704±.006	0.671±.008	0.655±.015	0.748±.008	0.720±.010	0.690±.022	0.052±.002	0.077±.003	0.142±.098

contribution to the risk of future foot complications, consistent with established medical knowledge. Similarly, elevated HbA1c levels, which is a marker of poor glycemic control, demonstrate increased contribution to foot complication risk, aligning with the known pathophysiology of diabetic foot disease. Age exhibits distinct patterns for the two outcomes: the risk contribution for foot complications increases until approximately 60 years before declining, whereas mortality risk increases monotonically with age. This divergence after age 60 likely reflects competing risks, where older patients succumb to systemic complications before developing foot-specific morbidity. The model correctly identifies that chronic obstructive pulmonary disease (COPD) does not contribute substantially to foot complication risk but significantly increases mortality risk. Similarly, malignancy shows stronger association with mortality than with foot complications. Finally, missing values for creatinine and albumin may indicate selective test ordering, where clinicians did not order these tests in the absence of clinical suspicion for renal dysfunction or hypoalbuminemia-related complications.

6 Conclusions

We have proposed an intrinsically interpretable model called CRISPNAM-FG which uses the Fine-Gray formulation to predict the Cumulative Incidence Function for competing risks scenarios in survival analysis. Our benchmarking results indicate that our model produces very competitive discrimination scores (TD-AUC and TD-CI respectively) while the calibration performance of our model lags behind that of models such as DEEPHIT. Additionally, a novel aspect of our work is the application of CRISPNAM-FG to a real world dataset to predict future foot complications post-discharge for a large cohort of diabetic patients in Ontario. On this dataset, our model's performance

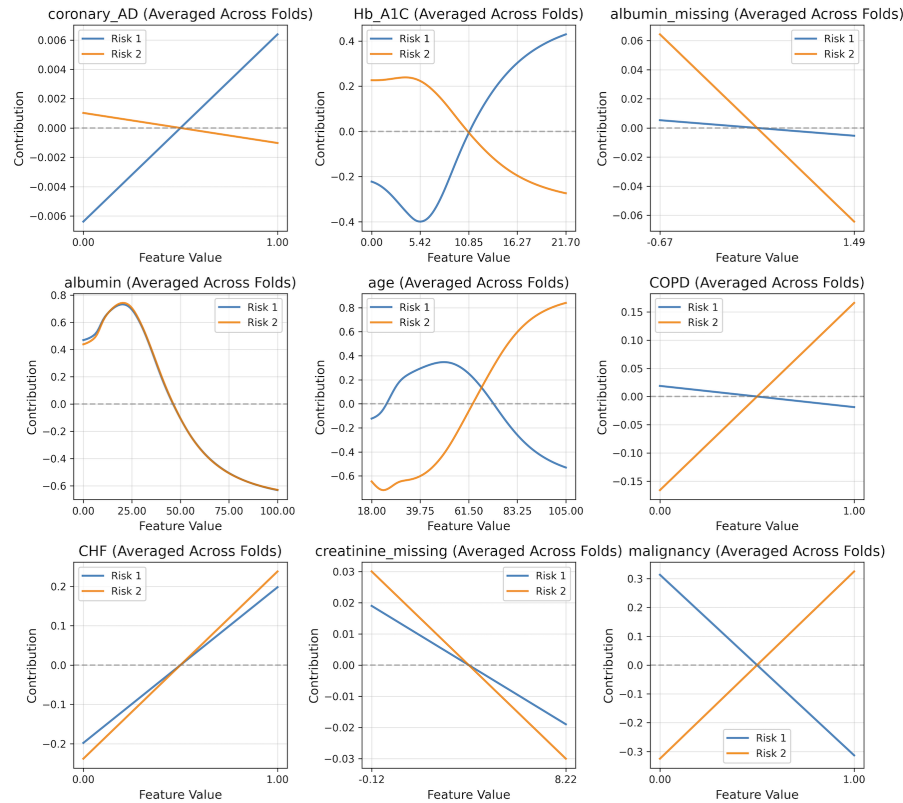


Figure 1: Shape functions generated by CRISPNAM-FG for the GEMINI Foot Complication dataset, averaged across folds. Nine features per risk are shown, randomly selected from the 20 most important.

was better than competing approaches. The core value add for mission critical applications in healthcare is the inherent interpretability provided by our model. The analysis of the shape function plots show in detail how each feature contributes to each risk, which can be valuable for trust and model validation. Since the goal is to prioritize interpretability, our model does not capture feature level interactions that may add complexity to the model and limits visualization to up to pairwise interactions only. Future improvements for the model would include investigating temporal feature nets which can capture temporal variation in feature contributions for each risk and developing an improved loss function that could potentially improve model’s calibration performance.

References

- [1] Rishabh Agarwal, Levi Melnick, Nicholas Frosst, Xuezhou Zhang, Ben Lengerich, Rich Caruana, and Geoffrey E Hinton. Neural additive models: Interpretable machine learning with neural nets. *Advances in neural information processing systems*, 34:4699–4711, 2021.
- [2] Latricia L Allen, Anjali Khakharia, Lawrence S Phillips, Theodore M Johnson, Constance R Uphold, Molly M Perkins, and Elizabeth Vaughan. Annual foot exams are associated with reduced incident amputation among older veterans with diabetes. *Journal of Applied Gerontology*, 42(2):205–212, 2023.
- [3] Norman E Breslow. Analysis of survival data under the proportional hazards model. *International Statistical Review/Revue Internationale de Statistique*, pages 45–57, 1975.
- [4] Ffion Dewi and Robert J Hinchliffe. Foot complications in patients with diabetes. *Surgery (Oxford)*, 38(2):108–113, 2020.

- [5] Jason P Fine and Robert J Gray. A proportional hazards model for the subdistribution of a competing risk. *Journal of the American statistical association*, 94(446):496–509, 1999.
- [6] GEMINI. Gemini: Harnessing hospital data to improve healthcare. <https://geminimedicine.ca/>, 2024. Canada’s largest hospital data sharing network for research and analytics.
- [7] SUPPORT Principal Investigators et al. A controlled trial to improve care for seriously hospitalized patients: The study to understand prognoses and preferences for outcomes and risks of treatments. *Jama*, 274:1591–1598, 1995.
- [8] Vincent Jeanselme, Chang Ho Yoon, Brian Tom, and Jessica Barrett. Neural fine-gray: Monotonic neural networks for competing risks. In *Conference on Health, Inference, and Learning*, pages 379–392. PMLR, 2023.
- [9] William B Kannel and Daniel L McGee. Diabetes and cardiovascular disease: the framingham study. *Jama*, 241(19):2035–2038, 1979.
- [10] Changhee Lee, William Zame, Jinsung Yoon, and Mihaela Van Der Schaar. Deephit: A deep learning approach to survival analysis with competing risks. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32, 2018.
- [11] Scott M Lundberg and Su-In Lee. A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30, 2017.
- [12] Ross L Prentice, John D Kalbfleisch, Arthur V Peterson Jr, Nancy Flournoy, Vernon T Farewell, and Norman E Breslow. The analysis of failure times in the presence of competing risks. *Biometrics*, pages 541–554, 1978.
- [13] Dhanesh Ramachandram and Ananya Raval. Crisp-nam: Competing risks interpretable survival prediction with neural additive models. In *Proceedings of the Second Workshop on Explainable Artificial Intelligence for the Medical Domain (EXPLIMED 2025) co-located with 28th European Conference on Artificial Intelligence (ECAI 2025)*, volume 4059 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2025. URL <https://ceur-ws.org/Vol-4059/>.
- [14] Terry M. Therneau and Patricia M. Grambsch. *The Cox Model*, pages 39–77. Springer New York, New York, NY, 2000. ISBN 978-1-4757-3294-8. doi: 10.1007/978-1-4757-3294-8_3. URL https://doi.org/10.1007/978-1-4757-3294-8_3.
- [15] Lev V Utkin, Egor D Satyukov, and Andrei V Konstantinov. Survnam: The machine learning survival model explanation. *Neural Networks*, 147:81–102, 2022.
- [16] Felix WA Waibel, Ilker Uçkay, Laura Soldevila-Boixader, Christina Sydler, and Karim Gariani. Current knowledge of morbidities and direct costs related to diabetic foot disorders: a literature review. *Frontiers in endocrinology*, 14:1323315, 2024.
- [17] Liangchen Xu and Chonghui Guo. Coxnam: An interpretable deep survival analysis model. *Expert Systems with Applications*, 227:120218, 2023.

Research Ethics

Research ethics approval for model development has been obtained from all participating GEMINI hospitals, as per the study protocols approved by Unity Health Toronto’s REB (SMH REB: 20-216, CTO ID: 3344) and in full compliance with PHIPA and TCPS2.