Predicting Individual Survival Distributions Using ECG: A Deep Learning Approach Utilizing Features Extracted by a Learned Diagnostic Model

Weijie Sun[1,2], Sunil Vasu Kalmady[1,2,3], Shi-ang Qi[2], Nariman Sepehrvand[1,3], Abram Hindle[2], Russell Greiner[2,4], Padma Kaul[1,3]

[1] Canadian VIGOUR Centre, University of Alberta, Edmonton, Canada
[2] Department of Computing Science, University of Alberta, Edmonton, Canada
[3] Department of Medicine, University of Alberta, Edmonton, Canada
[4] Alberta Machine Intelligence Institute, Edmonton, Canada
weijie2@ualberta.ca

Abstract

In the field of healthcare, individual survival prediction is important for personalized treatment planning. This study presents machine learning algorithms for predicting Individual Survival Distributions (ISD) using electrocardiography (ECG) data in two different formats. The models, which predict time until death, are developed and evaluated on a large, population-based cohort from Alberta, Canada. Our results demonstrate that models trained on raw ECG waveforms significantly outperform those trained on traditional ECG measurements in several metrics, including concordance index, hinge L1 loss, margin L1 loss, and margin truncated L1 loss. Additionally, the integration of predicted probabilities from wide-range diagnostic tasks not only enhances our ISD models' performance but also makes them significantly superior to other models across all evaluation metrics in individual survival prediction tasks. This innovative approach highlights the potential to leverage insights from diagnostic models for prognostic tasks, such as individual survival prediction. These findings could have far-reaching implications for the development of personalized treatment plans and open new avenues for future research in survival prediction using ECGs.

Introduction

Electrocardiography (ECG) is a fundamental diagnostic tool in cardiology, providing a non-invasive method to record the electrical activity of the heart over time. However, the interpretation of ECG data is complex and requires specialized knowledge and expertise. Traditional methods often involve manual analysis, which is both time-consuming and requires specialized education and practice. Moreover, traditional methods typically require cardiologists who have undergone extensive training be able to extract ECG measurements, such as Q duration and RR interval, from raw ECG waveforms for identifying patient's heart condition. These traditional methods may not fully capture the wealth of information contained within the ECG signals, potentially limiting clinical utility and their predictive accuracy in diagnosing patient's heart condition. Recently, deep learning techniques have emerged as a promising solution to these challenges, offering the potential to automate and enhance the interpretation of ECG data. These techniques can learn complex patterns from raw ECG data and have been successfully applied in both diagnostic (Ribeiro et al. 2020; Sun et al. 2022) and prognostic (Raghunath et al. 2020; Sun et al. 2023) tasks.

Previous mortality prediction studies often centered on binary outcomes, at a single mortality risk point. This approach, while informative, presents challenges: it may not sufficiently aid clinical decision-making, and many such systems simply ignore censored data. Recognizing these limitations, we extend the application of deep learning to ECG data by developing ISD algorithms that can provide survival probabilities for novel instances across all future time points, while incorporating information obtained from the censored events. We implement ISD models to accommodate both traditional ECG measurements and raw ECG waveforms.

Further, building on our previous work (Sun et al. 2022), which demonstrated the potential of using deep learning techniques to produce models that could accurately predict a wide range of diseases from ECG data, we adopt a transfer learning-inspired approach. Specifically, we enhance our models by leveraging the predicted probabilities generated by a diagnostic model designed to estimate patients' health conditions. This innovative approach aims to refine ISD predictions.

This paper makes several significant contributions:

1. ECG-based ISD Algorithm: We propose a novel ISD algorithm that computes a patient's time-to-death ISD, utilizing ECG voltage time-series, age, and sex.
2. Incorporation of Predicted Probabilities from Diagnosis Model: We explore an ISD model, whose input includes the diagnostic probabilities predicted from ECG, enhancing prediction precision.
3. Novel Evaluation Metric (L1-margin-truncated): We introduce the L1-margin-truncated metric, addressing challenges in survival datasets where long-term tracking may be impractical.

Our findings contribute valuable insights into the potential of deep learning in enhancing survival prediction, offering a foundation for future research and potential clinical applications in cardiology.

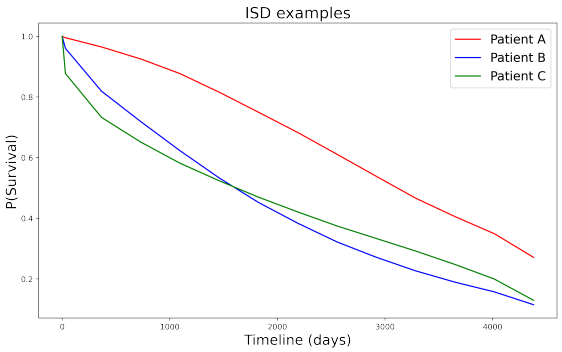

Figure 1: This figure illustrates the individual survival curves for three patients over a 13-year time interval. Patient A exhibits a greater likelihood of survival compared to Patients B and C. Within the first 5 years, Patient C faces a higher mortality risk than Patient B. However, after 5 years, Patient B demonstrates a higher mortality risk relative to Patient C.

Background Work

Clinical studies have demonstrated a significant correlation between ECG abnormalities and the risk of mortality. For instance, Nishime et al. (2000) concluded that ECG abnormalities could be used as a significant predictor of coronary heart disease and all-cause mortality in middle-aged male patients. Goldman et al. (2019) study found that the all-cause mortality rate was significantly higher in individuals with an incidental abnormal ECG finding. They suggested that incidental ECG abnormalities could be used as a marker for increased risk of mortality.

Recently, Raghunath et al. (2020) developed a CNN-based deep learning model that could predict a patient's 1-year mortality from ECG. Sun et al. (2023) extended their work with a ResNet-based model that used ECG traces to predict short- and long-term mortality prediction. However, these binary mortality classification algorithms have some limitations: (1) they may not fully utilize censored data; (2) the single time-point survival probability may not provide sufficient information for clinicians; and (3) the multiple probabilistic binary mortality models may produce counter-intuitive results where the short-term survival probability is lower than the long-term survival probability.

To address these limitations, researchers have proposed ISD algorithms. These algorithms can generate monotonically decreasing survival curves for all future time points for each patient. Figure 1 illustrates the individual survival curves of three patients. In the other domains, researchers have already used multi-modal data in ISD algorithms. Kim, Kazmierski, and Haibe-Kains (2021) proposed a method called Deep-CR MTLR, which uses deep neural networks to learn joint prognostic representations between CT images and clinical information. This approach extends MTLR to medical image data input and considers competing risks for cancer and other causes of survival prediction. Moreover, Li et al. (2019) proposed the deep learning "Deep Correlational Survival Model" (DeepCorrSurv) approach for predicting patient survival based on multi-modal imaging data. In a public cancer survival dataset, DeepCorrSurv outperformed other models in predicting patient survival using both CT images and clinical data.

These studies highlight the potential of ISD algorithms in improving the survival predictive performance and clinical utility of multi-modal data, such as ECG along with the tabular features.

Methods

This section introduces comprehensive approaches to predict ISD using ECG data. We leverage a large dataset from Alberta, Canada, and develop three distinct models, comparing two different ECG representations (raw ECG waveform and ECG measurements). The models are trained using a convolutional neural network (CNN) based on the ResNet architecture (He et al. 2016), with specified hyperparameters and training settings. To address the bias related to overpresentation of severely sick patients, the performance of the models is evaluated using one ECG from each patient, selected at random from the holdout set.

Data

Data Sources and Structure: The data were collected from the province of Alberta, Canada, which operates a single-payer and single-provider healthcare system. This ensures universal access to hospital, ambulatory, laboratory, and physician services for its 4.4 million residents. The dataset includes: Hospitalization Data, Outpatient Clinic Visit Data, Demographic Information, and Vital Status Death Registry.

In addition to these data sources, diagnoses were coded using the World Health Organization's International Classification of Diseases (ICD) (WHO 2016), which is a 3 to 7-character identifier that specifies a specific disease. For example, 'I214' refers to 'Non-ST elevation (NSTEMI) myocardial infarction'. We utilized these ICD codes and their corresponding categories as labels for our prediction modeling. In our dataset, we identified 1,414 distinct ICD codes, each of which was linked to at least 1,000 ECGs. Note that a single patient might have more than one ICD code.

The study cohort included patients presenting to 84 emergency departments or hospitals between February 2007 and April 2020 in the province of Alberta. Clinical characteristics of patient cohorts used for learning and evaluating the models have been described in our earlier studies (Sun et al. 2022, 2023).

ECG data: The ECG data consisted of standard 12-lead ECG traces as well as the associated ECG measurements from the Philips IntelliSpace ECG system (Systems 2009). Each of the 12 leads had a sequence of ECG voltages sampled at 500 Hz for 10 seconds. The ECG measurements were automatically generated by the ECG machine manufacturer's built-in algorithm and included 22 features such as atrial rate, P duration, RR interval, Q wave onset, Fridericia rate-corrected QT interval, heart rate, and others.

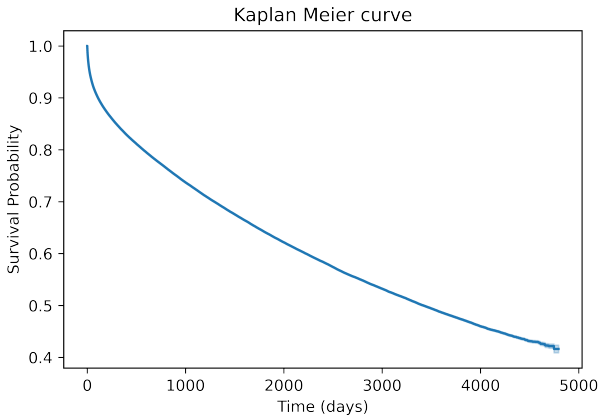

Figure 2: Kaplan Meier (KM) Curve illustrating the survival distribution of the study cohort. Note the median survival time is 3,420 days, corresponding to a 50% survival probability. The curve spans 4,794 days, covering 13 years from 2007 to 2020.

Data Processing and Cohort Selection: After excluding the ECGs that could not be linked to any episode, the ECGs of patients under 18 years of age, and the ECGs with poor signal quality, the final analysis cohort contained 1,605,268 ECGs from 748,773 episodes of 244,077 patients. We transformed our data into a survival dataset denoted as $D = \{[\vec{X}_i, t_i, \delta_i] \mid i = 1, 2, \ldots, n\}$, where: i represents the i^{th} patient in the dataset, $\vec{X}_i$ denotes the feature vector for patient i, t_i specifies the time to either censoring or death after ECG test date for patient i, δ_i serves as an indicator variable for the death event of patient i, where $\delta = 1$ represents death and $= 0$ represents censored. Notably, the censor rate of the dataset is 63.64%, indicating that patients with these ECGs did not have the exact death date. See the Kaplan Meier (KM) curve (Kaplan and Meier 1958) in Figure 2 for the survival distribution of this study cohort.

Dataset Split: We split our ECG dataset into the development set (random 60%: 143,939 patients with 436,508 ECGs, used for training and internal validation) and holdout set (remaining 40%: 95,913 patients with 287,566 ECGs). We ensured that ECGs from the same patient were not shared between the development set and holdout set.

Models

In this study, we explored three distinct architectures for individual survival analysis, a process that leverages personal data to estimate the time until an event of interest will occur – see Figure 3. Each of these models employs the N-MTLR algorithm (Fotso 2018) for learning a model for individual survival prediction, but each first transforms the input data into different intermediate representation.

The implementation of the N-MTLR algorithm in our study is based on Kazmierski (2020)'s torchmtlr codebase, which provides a robust and efficient framework. It consists of three fully connected layers, each containing 128 hidden neurons. This architecture allows the models to capture nonlinear relationships in the data, providing a robust framework for survival analysis. By utilizing multiple layers and a substantial number of neurons, the N-MTLR algorithm enhances the models' ability to generalize from the training data, leading to more accurate and reliable predictions.

Model A: This model is an end-to-end ISD model, designed to provide comprehensive survival predictions for each patient. The learned model takes as input a set of 12-lead ECGs, which are represented as a 12×4096 numeric matrix. This matrix is then processed through a ResNet architecture, which transforms the raw ECG data into a high-level feature representation. In addition to the ECG-derived features, demographic features such as age and sex are also incorporated into the model. These combined features form the comprehensive input for the N-MTLR algorithm. The output of the model is an individual survival distribution.

Model B: This model is a two-step survival prediction approach. In the first step, we utilize a learned model from (Sun et al. 2022) ECG diagnosis study. This model, which is capable of predicting 1,414 different ICD diagnosis labels, produces features that are input as an ECG feature extractor. This model is learned only using same development set as the current study. It processes the ECG data in the development set, transforming each instance into a vector of 1414 predicted diagnosis probabilities (one for each diagnosis). In the second step, these predicted diagnosis probabilities are combined with demographic features as input features. These features are then used to train an N-MTLR model on the development set. This model generates the final individual survival distributions. During both learning as well as evaluation phases, note that we do not use any diagnosis labels from the holdout set, ensuring that our model's performance is evaluated purely on its ability to predict survival based on ECG traces and demographic features.

Model C: This model is designed to accommodate handcrafted ECG features. It uses 22 ECG measurements, along with age and sex, as input features for training with the N-MTLR model.

For Models A and B, we implement a convolutional neural network (CNN) based on the ResNet architecture. This consists of a convolutional layer, four residual blocks with two convolutional layers per block, followed by a dense layer. We use batch normalization (Ioffe and Szegedy 2015), ReLU, and dropout (Hinton et al. 2012) after each convolutional layer. The architecture is based on a model trained on a large ECG dataset from Ribeiro et al. (2020) study to identify abnormalities in 12-lead ECGs.

All three models are trained using negative log likelihood as the loss function, with an initial learning rate of 1×10^{-4}, Adam optimizer (Kingma and Ba 2014), kernel size of 16, batch size of 64, and a dropout rate of 0.2. Other hyperparameters are set to default. The models are trained for a maximum of 70 epochs. The learning process is stopped if the loss in the tuning set does not reduce for 9 epochs. Moreover, the models are implemented using PyTorch 1.11 in Python 3.8. We train all our models with 8 Tesla V100-SXM2 GPUs and 32 GB of RAM per GPU.

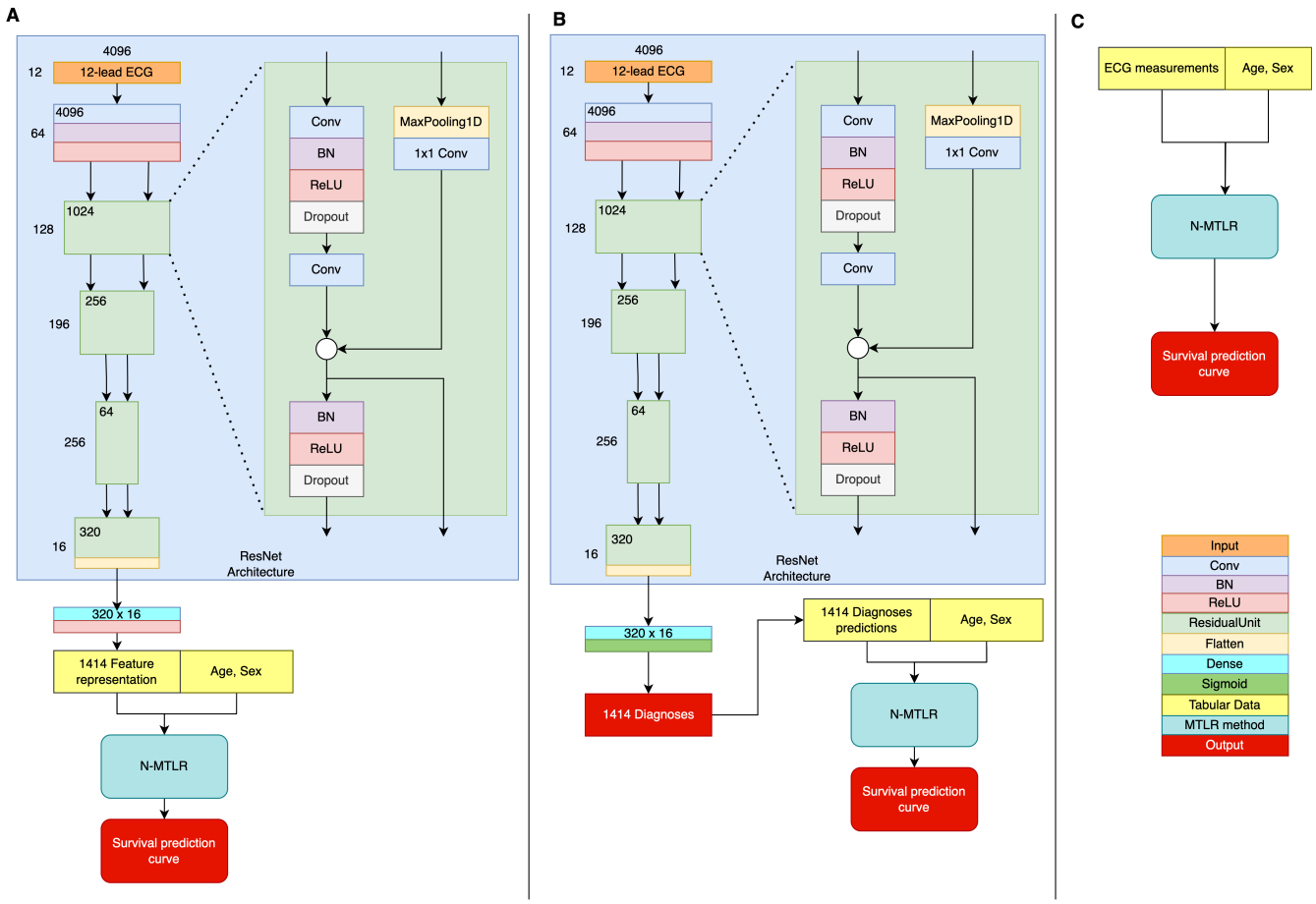

Figure 3: Schematic of ISD models. Three ECG feature representations: Model A (End-to-End): Takes 12-lead ECG waveforms, age, and sex as inputs and directly outputs survival probabilities for future time points; Model B (Two-Step): The first step processes 12-lead ECG waveforms to generate 1,414 diagnostic prediction values. The second step takes these values along with age and sex to output survival probabilities; Model C: Utilizes ECG measurements as input to predict survival probabilities at all future time points. Each model aims to estimate survival probabilities over time based on different feature representations and methodologies.

	Input	Features for survival algorithm	Hinge L1 loss	Marginal Truncated L1 loss	Marginal L1 loss	C index	IBS
Model A	12 lead ECG + Age, Sex	same as input	547.50 (545.51 - 549.20)	1219.54 (1218.21 - 1222.38)	2260.60 (2256.24 - 2263.86)	0.7643 (0.7627 - 0.7660)	0.1503 (0.1490 - 0.1518)
Model B	12 lead ECG	1414 ICD predictions + Age, Sex	**514.78 (513.09 - 516.80)**	**1078.21 (1074.78 - 1082.38)**	**2116.31 (2112.12 - 2120.32)**	**0.8004 (0.7995 - 0.8011)**	**0.1368 (0.1355 - 0.1382)**
Model C	ECG measurements + Age, Sex	same as input	564.24 (563.04 - 566.03)	1228.12 (1224.48 - 1231.67)	2304.69 (2302.16 - 2307.95)	0.7589 (0.7576 - 0.7597)	0.1508 (0.1495 - 0.1517)

Table 1: Evaluation of ECG ISD models' performance in hinge L1 loss, margin truncated L1 loss, margin L1 loss, C-index, and IBS expressed in mean (95% confidence interval)

Results

In this section, we assess the performance of the ISD models using five key metrics: the Concordance Index (C-index), hinge L1 loss, margin truncated L1 loss, margin L1 loss, and the Integral Brier Score (IBS).

Concordance Index (C-index)

The C-index (Antolini, Boracchi, and Biganzoli 2005) is a widely used metric in survival analysis to evaluate the discriminative power of a risk model. It quantifies the agreement between the predicted survival risk and the observed survival times, with values ranging from 0 to 1. A C-index of 0.5 represents the baseline performance of a model that assigns probabilities randomly, while a higher C-index indicates that the predicted survival risks are better ordered when compared across all patients.

To compute the C-index, we must first define the set of all comparable pairs (CP) of patients. A pair of patients (i, j) is considered comparable if patient j is alive when patient i is dead. The comparable pairs can be defined as:

$$CP_{i,j} = I\{t_i < t_j \wedge \delta_i = 1\} + I\{t_i = t_j \wedge \delta_i = 1 \wedge \delta_j = 0\}$$

Next, we identify the correctly ranked comparable pairs, following the approach of Antolini, Boracchi, and Biganzoli (2005) study. Here, $r(\vec{x_i})$ represents the risk score of patient i.

$$\text{CPcorrect}_{i,j} = I\{r(\vec{x_i}) < r(\vec{x_j})\} \cdot CP_{i,j}$$

Finally, the C-index is estimated by calculating the ratio of correctly ranked comparable pairs to all comparable pairs. In other words, given two randomly selected patients i and j in an ISD model with an 80% C-index, if $r(\vec{x_i}) > r(\vec{x_j})$, then there is an 80% probability that patient i's event will occur before patient j's event:

$$\text{C-index} = \frac{\sum_{i=1}^{n}\sum_{j=i}^{n} \text{CPcorrect}_{i,j;i\neq j}}{\sum_{i=1}^{n}\sum_{j=i}^{n} CP_{i,j;i\neq j}}$$

This formulation of the C-index provides a robust measure of the model's ability to accurately rank patients based on their survival risk, reflecting the model's effectiveness in survival prediction.

L1 loss

The ISD model produces an individual survival curve, predicting the median survival time $\hat{t}^{0.5}$ representing a 50% chance for a patient to survive until the event occurs[1].

For an uncensored patient, the L1 loss is defined as the average absolute value of the difference between the predicted median survival time $\hat{t}^{0.5}$ and true event time $t_{event} = d$. Since the true event time is unknown for censored patients, we consider two approaches from Haider et al. (2020) study: hinge L1 loss and margin L1 loss, and a new approach: margin truncated L1 loss.

[1]According to Haider et al. (2020), we use extrapolation to extend the last point if the user's survival curve does not reach the median until the study ends.

Hinge L1 loss: The hinge L1 loss gives 0 loss if expected survival time larger and equal than the censor time. The expression for L1-hinge loss is given by:

$$L1_{\text{hinge}}(D, \hat{t}^{0.5}) = \frac{1}{|D|}[\sum_{i \in D_{\text{uncensor}}} |d_i - \hat{t}_i^{0.5}| + \sum_{k \in D_{\text{censor}}} \max(0, c_k - \hat{t}_k^{0.5})]$$

Margin L1 loss: In the margin L1 loss, we also assign a "Best-Guess" (BG) value to each censored patient, representing the expected survival time given that the patient already survived until c

$$BG(c) = c + \frac{\int_c^{\infty} S_{KM}(t)dt}{S_{KM}(c)}$$

In our evaluation method, we use the Kaplan-Meier estimator $\hat{S}_{KM}(\cdot)$ (Kaplan and Meier 1958) from the training dataset to estimate the survival function. The margin L1 loss is defined as:

$$L1_{\text{margin}}(D, \hat{t_i}^{0.5}) = \frac{1}{|D_{\text{uncensor}}| + \sum_{k \in D_{\text{censor}}} a_k} \times \left[\sum_{i \in D_{\text{uncensor}}} |d_i - \hat{t}_i^{0.5}| + \sum_{k \in D_{\text{censor}}} a_k |BG(c_k) - \hat{t}_k^{0.5}|\right]$$

where α_k is the weight in each Best-Guess estimation contributing to the L1-margin loss. Since instances with early censor times provide less information than cases with late censor times, we assign more weight to the late censor time instances in the margin L1 loss by setting $a_k = 1 - \hat{S}_{KM}(c_k)$.

Margin truncated L1 loss: In the margin truncated L1 loss(L1-margin-T loss), we address a common challenge in survival datasets where tracking a patient's life for an extended period may not be feasible. When using the Best Guess method to estimate a censored patient's event time, we observed that both the predicted time and the best guess could become significantly large if the censor's predicted time is long enough. This large difference may not align with clinical interests, as clinicians often focus on assessing a patient's short-term risk rather than long-term living conditions. To address this issue, we introduce a truncated time τ to limit the impact of extremely long predictions and best guesses. The L1-margin-T loss is mathematically defined as:

$$L1_{\text{margin-T}}(D, \hat{t}^{0.5}) = \frac{1}{|D_{\text{uncensor}}| + \sum_{k \in D_{\text{censor}}} a_k} \times \left[\sum_{i \in D_{\text{uncensor}}} |d_i - \hat{t}_i^{0.5}| + \sum_{k \in D_{\text{censor}}} a_k \times \text{c\_loss}\right],$$

$$\text{where} \quad \text{c\_loss} = \left|\min(\tau, BG(c_k)) - \min(\tau, \hat{t}_k^{0.5})\right|$$

Here τ represents the truncated time, which serves as an upper limit for both the predicted time and the best guess. By using $\min(\tau, \cdot)$, we ensure that neither value exceeds τ, thus mitigating the effect of overly long predictions. In this study, we use the longest time of 4794 days (greater than 13 years), which we can track in our study, as the truncated time τ. This choice of τ reflects the practical constraints of our dataset and aligns the loss function more closely with clinical priorities.

Integral Brier Score(IBS)

The IBS (Graf et al. 1999) provides an average measure of the Brier score (Brier and Allen 1951) across a time interval from 0 to infinity. The Brier score itself is a mean squared error metric between the true event status (for uncensored: either 1 for alive or 0 for dead) and the predicted survival probability at a single time point. A lower IBS indicates a better model, with the baseline score being 0.25 for a model that predicts a survival probability of 0.5 for all patients.

Statistical Comparison

We used a bootstrap statistical test to compare the performances between our models in the evaluation stage. First, we selected a single, randomly chosen ECG record from each patient's collection of ECGs in the holdout set. This step was repeated to generate 10 distinct random sets. Subsequently, for each random set, we performed the traditional bootstrap process, repeating it 10 times, yielding total of 100 replacement sets for evaluating each model. Then, performance metrics were computed for each of 100 sets, and 95% confidence interval (CI) was calculated for each model. This allows us to compare each model's performance using the upper and lower bounds of CI. When 95% CIs were not overlapping, difference between two model performances was considered to be statistically significant.

As illustrated in Table 1, the following key observations can be made:

Deep Learning Raw ECG traces vs. ECG Measurements (Model A vs. Model C): Model A, an end-to-end ISD model that utilizes a ResNet architecture to process raw ECG data with a significantly high C-index of 0.7643, and significantly low hinge L1 loss of 547.50, margin truncated L1 loss of 1219.54, and margin L1 loss of 2260.60 outperforms Model C, which relies on hand-crafted ECG features. This result underscores the effectiveness of deep learning techniques in handling raw ECG data, as opposed to relying solely on ECG measurements. The superiority of Model A over Model C demonstrates the potential of leveraging complex neural network architectures to extract meaningful features from raw data.

Incorporating Predicted Diagnosis Probabilities (Model B vs. Model A): Model B, which employs a two-step learning process with an ICD-10 based feature extractor, achieves a significantly higher C-index of 0.8004 and significantly lower hinge L1 loss of 514.78, margin truncated L1 loss of 1078.21, margin L1 loss of 2116.31, and IBS of 0.1368, outperforming Model A. This finding emphasizes the value of incorporating predicted diagnosis probabilities into the prognosis model. We transformed each instance into a vector of 1414 predicted diagnosis probabilities and combined them with demographic features.

Discussion and Conclusion

In this study, we have made notable strides in leveraging ECG data for survival prediction. Our contributions include the development of an end-to-end individual survival algorithm, the innovative use of predicted probabilities from an ICD-wise diagnosis model, and the introduction of the margin truncated L1 loss to assess short-term performance. These advancements collectively enhance the accuracy and relevance of survival predictions, particularly for high-risk patients.

Our findings reveal that the machine learning models we developed can effectively predict survival probabilities for all future time points. The end-to-end model outperforms traditional models that rely on ECG measurements. Moreover, Model B, which incorporates predicted probabilities, demonstrates superior performance in multiple evaluation metrics, including concordance index, hinge L1 loss, margin L1 loss, margin truncated L1 loss, and IBS. This success underscores the potential of deep learning techniques in leveraging the ECG feature representations from diagnosis tasks on estimating patients' health risks.

Despite these promising results, our study has limitations. First, our models are trained and tested on a single healthcare system's data, which may limit their generalizability to other healthcare systems or populations. Secondly, while our study demonstrates that ECG traces significantly outperform ISD models trained with traditional ECG measurements, it is worth noting that the magnitude of improvement may not be substantial. Finally, while our models perform well in predicting survival probabilities, they do not provide insights into the specific factors contributing to these predictions. This limits their interpretability, which is a crucial aspect of clinical decision-making. Future research will focus on predicting readmission times using the ISD algorithm, enhancing interpretability through explainable AI techniques, and validating the models on diverse healthcare systems to ensure their generalizability and robustness.

The models and algorithms developed in this study hold substantial promise for clinical practice. The introduction of the margin truncated L1 loss aligns the models with clinical priorities, focusing on short-term risks that are often more relevant to patient care. The end-to-end ECG ISD model outperformed ECG measurements ISD model, which can guide personalized treatment strategies and risk assessments. Finally, this pioneering approach, which uses the ECG feature representation from diagnosis models, opens new horizons for further research and development in this vital area.

Acknowledgements

Data was extracted from the Alberta Health Services Enterprise Data Warehouse with support provided by AbSPORU Data and Research Services platform which is funded by CIHR, Alberta Innovates, University Hospital Foundation,

University of Alberta, University of Calgary and Alberta Health Services. The interpretation and conclusions contained herein are those of the researchers and do not necessarily represent the views of Alberta Health Services or any of the funders. We are grateful for the funding from NSERC (Canada's Natural Science and Engineering Research Council), CIFAR (Canadian Institute for Advanced Research) and AMII (Alberta Machine Intelligence Institute).

References

Antolini, L.; Boracchi, P.; and Biganzoli, E. 2005. A time-dependent discrimination index for survival data. *Statistics in medicine*, 24(24): 3927–3944.

Brier, G. W.; and Allen, R. A. 1951. Verification of weather forecasts. In *Compendium of meteorology*, 841–848. Springer.

Fotso, S. 2018. Deep neural networks for survival analysis based on a multi-task framework. *arXiv preprint arXiv:1801.05512*.

Goldman, A.; Hod, H.; Chetrit, A.; and Dankner, R. 2019. Incidental abnormal ECG findings and long-term cardiovascular morbidity and all-cause mortality: A population based prospective study. *International Journal of Cardiology*, 295: 36–41.

Graf, E.; Schmoor, C.; Sauerbrei, W.; and Schumacher, M. 1999. Assessment and comparison of prognostic classification schemes for survival data. *Statistics in medicine*, 18(17-18): 2529–2545.

Haider, H.; Hoehn, B.; Davis, S.; and Greiner, R. 2020. Effective Ways to Build and Evaluate Individual Survival Distributions. *J. Mach. Learn. Res.*, 21(85): 1–63.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778.

Hinton, G. E.; Srivastava, N.; Krizhevsky, A.; Sutskever, I.; and Salakhutdinov, R. R. 2012. Improving neural networks by preventing co-adaptation of feature detectors. *arXiv preprint arXiv:1207.0580*.

Ioffe, S.; and Szegedy, C. 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International conference on machine learning*, 448–456. PMLR.

Kaplan, E. L.; and Meier, P. 1958. Nonparametric estimation from incomplete observations. *Journal of the American statistical association*, 53(282): 457–481.

Kazmierski, M. 2020. torchmtlr: Flexible and Modular Implementation of Multi-Task Logistic Regression in PyTorch. https://github.com/mkazmier/torchmtlr. Accessed: 2022-10-01.

Kim, S.; Kazmierski, M.; and Haibe-Kains, B. 2021. DeepCR MTLR: a Multi-Modal Approach for Cancer Survival Prediction with Competing Risks. In *Survival Prediction-Algorithms, Challenges and Applications*, 223–231. PMLR.

Kingma, D. P.; and Ba, J. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Li, H.; Boimel, P.; Janopaul-Naylor, J.; Zhong, H.; Xiao, Y.; Ben-Josef, E.; and Fan, Y. 2019. Deep convolutional neural networks for imaging data based survival analysis of rectal cancer. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, 846–849. IEEE.

Nishime, E. O.; Cole, C. R.; Blackstone, E. H.; Pashkow, F. J.; and Lauer, M. S. 2000. Heart rate recovery and treadmill exercise score as predictors of mortality in patients referred for exercise ECG. *Jama*, 284(11): 1392–1398.

Raghunath, S.; Ulloa Cerna, A. E.; Jing, L.; VanMaanen, D. P.; Stough, J.; Hartzel, D. N.; Leader, J. B.; Kirchner, H. L.; Stumpe, M. C.; Hafez, A.; et al. 2020. Prediction of mortality from 12-lead electrocardiogram voltage data using a deep neural network. *Nature medicine*, 26(6): 886–891.

Ribeiro, A. H.; Ribeiro, M. H.; Paixão, G. M.; Oliveira, D. M.; Gomes, P. R.; Canazart, J. A.; Ferreira, M. P.; Andersson, C. R.; Macfarlane, P. W.; Meira Jr, W.; et al. 2020. Automatic diagnosis of the 12-lead ECG using a deep neural network. *Nature communications*, 11(1): 1–9.

Sun, W.; Kalmady, S. V.; Salimi, A.; Sepehrvand, N.; Ly, E.; Hindle, A.; Greiner, R.; and Kaul, P. 2022. ECG for high-throughput screening of multiple diseases: Proof-of-concept using multi-diagnosis deep learning from population-based datasets. *arXiv preprint arXiv:2210.06291*.

Sun, W.; Kalmady, S. V.; Sepehrvand, N.; Salimi, A.; Nademi, Y.; Bainey, K.; Ezekowitz, J. A.; Greiner, R.; Hindle, A.; McAlister, F. A.; et al. 2023. Towards artificial intelligence-based learning health system for population-level mortality prediction using electrocardiograms. *NPJ Digital Medicine*, 6(1): 21.

Systems, P. M. 2009. Philips DXL ECG Algorithm Physician's Guide for PH100B, Edition 2. April, 2009.

WHO. 2016. *International statistical classification of diseases and related health problems*, volume 2. World Health Organization.