MultSurv: A Multimodal Deep Learning Model for Hospitalized Patients Survival Analysis in the Context of a Pandemic

Felipe André Zeiser[1], **Cristiano André da Costa**[1], **Gabriel de Oliveira Ramos**[1]

[1]Software Innovation Laboratory - SOFTWARELAB
Applied Computing Graduate Program
Universidade do Vale do Rio dos Sinos (UNISINOS) - São Leopoldo, Brazil

felipezeiser@unisinos.br, cac@unisinos.br, gdoramos@unisinos.br

Abstract. *Respiratory infectious diseases continue to be a significant public health challenge. For example, the COVID-19 pandemic has exposed critical weaknesses in healthcare systems worldwide. The heterogeneity of clinical manifestations in hospitalized patients highlights the need for risk stratification methods to aid in better patient management and resource allocation. In this study, we propose MultSurv MultSurv, a multimodal survival analysis model that integrates clinical, laboratory, and chest X-ray data to capture the temporal dynamics of disease progression. The methodology comprises five main components: pre-processing, feature encoders, temporal attention, CheXReport for image analysis, and multitask neural networks for risk assessment. The model was evaluated using a private dataset, demonstrating superior performance over state-of-the-art methods, with a C-index of* $0.723(t = 1)$ *and* $0.695(t = 7)$*. The proposed approach can improve patient prioritization in pandemic scenarios and be adapted for broader clinical applications, potentially contributing to better decision-making in healthcare settings.*

1. Introduction

The COVID-19 pandemic was the most significant healthcare challenge of the century, exposing system weaknesses, deepening inequalities in developing countries, and revealing issues in governance, misinformation, and resource allocation [Leach et al. 2021, Zeiser et al. 2022]. The rising frequency of zoonotic diseases, driven by globalization and climate change, highlights the need for advanced analytical tools to address emerging health threats [Rahman et al. 2020].

Electronic Health Records (EHRs) have provided valuable data for understanding COVID-19. However, their heterogeneity, sparsity, and longitudinal nature present challenges for analysis. The irregular timing of data collection in hospitalized patients leads to sparse and longitudinal data, further complicated by competing risks where multiple health events can occur simultaneously or in rapid succession. Extracting meaningful insights from such data requires AI models capable of handling non-linear relationships, yet current approaches often rely on the last available measurement for prediction, neglecting the temporal evolution of the disease.

To overcome these challenges, it is crucial to develop techniques that integrate temporal dynamics and leverage multimodal data—including medical images, vital signs, and clinical text—to provide a comprehensive view of a patient's health status. Recent advances, like Dynamic-DeepHit, leverage deep learning to model temporal dependencies and competing risks [Lee et al. 2019], but they still face challenges in effectively processing multimodal data.

In this way, our main goal is to develop and evaluate a survival analysis model that integrates longitudinal and multimodal data from EHRs to enhance the diagnosis and prognosis of COVID-19 patients. With the model, we aim to provide healthcare professionals with an explainable, accurate tool for prioritizing patient care and optimizing resource allocation in hospital settings. By incorporating clinical, laboratory, and imaging data, the model seeks to improve the understanding of the influences of different variables on patient survival and support the implementation of personalized treatment strategies.

Our main scientific contribution is the proposal of a survival analysis model that can process multimodal and longitudinal data from patients hospitalized in the context of a pandemic. In addition, we present the complementary scientific contributions:

- **State-of-the-art survey** on survival analysis, data fusion, and competing risks.
- **CheXReport architecture**, a transformer-based model for radiological report generation.
- **Embedding techniques** for categorical and continuous variables, improving predictive accuracy.
- **Multitask learning framework** for modeling multiple health risks simultaneously.

2. Methodology

We present the MultSurv model (Figure 1), designed to process multimodal clinical, laboratory, and imaging data while preserving its temporal nature. Feature encoders transform tabular data into dense embeddings, while a temporal attention mechanism prioritizes relevant historical information. Chest X-ray features are extracted via the transformer-based CheXReport [Zeiser et al. 2024] architecture, which integrates visual and textual data. Multitask networks model complex relationships between risk factors, improving predictive accuracy.

The model is trained and validated on the MyDigitalHealth (MDH) Dataset, which consists of 1,815 hospitalized COVID-19 patients confirmed via RT-qPCR at HCPA between March 2020 and June 2022 with approval of the Research Ethics Committee (CAAE under number 33540520.6.3004.5327). This dataset includes sociodemographic, clinical, laboratory, imaging, and unstructured electronic medical records. Given the heterogeneity of medical data, preprocessing techniques are applied to normalize continuous variables, impute missing values, and enhance imaging features. Chest X-rays undergo contrast normalization via CLAHE to mitigate intensity variations, improving robustness in feature extraction.

We employ a Gated Recurrent Unit combined with a temporal attention mechanism for sequential data modeling, dynamically assigning importance weights to past observations. This enables the model to focus on the most informative time points, comprehensively representing a patient's health progression. CheXReport enhances imaging analysis using Swin Transformer blocks, with the ability to capture local and global spatial relationships, improving radiological assessment and structured report generation [Zeiser et al. 2024]. Finally, to predict multiple health risks simultaneously, we adopt a multitask learning framework, where shared representations improve generalization while specialized subnetworks refine individual risk predictions.

The MultSurv model is optimized using a combination of loss functions that balance survival event modeling, temporal consistency, and feature extraction from chest X-ray images. The total loss is defined as: $L_{(total)} = L_1 + L_2 + L_3$. The first term, L_1, is a binary cross-entropy loss that estimates the probability of observed survival events $L_1 = \frac{1}{N}\sum_{i=1}^{N}\left[I_i \log_2\left(\sum_{k=1}^{K} m_{i,k}o_{i,k}\right) + (1-I_i)\log_2\left(\sum_{k=1}^{K} m_{i,k}o_{i,k}\right)\right]$, where N is the number of patients, K the event types, I_i an event indicator, $m_{i,k}$ a mask for event k in patient i, and $o_{i,k}$ the predicted event probability. The second term, L_2, is a mean squared error (MSE) loss that regularizes temporal predictions while handling missing values: $L_2 = \frac{1}{N}\sum_{i=1}^{N}\sum_{t=2}^{T} m_{i,t}(1-m_{i,t})(y_{i,t}-x_{i,t})^2$, where T is the total time points, $m_{i,t}$ a missing data mask, $y_{i,t}$ the predicted value, and $x_{i,t}$ the observed value. The last term, L_3, is a cross-entropy loss for optimizing the CheXReport network, incorporating attention regularization: $L_3 = -\sum_{t=1}^{T}\log_2\left(p_\Theta(y_t^*|y_{t-1}^*)\right) + \sum_{l=1}^{L}\frac{1}{L}\sum_{d=1}^{D}\sum_{i=1}^{M^2}\left(1-\sum_{c=1}^{T}\alpha_{ctdl}\right)$, where Θ represents model parameters, y_t^* the ground-truth token, $p_\Theta(y_t^*|y_{t-1}^*)$ the predicted probability, D the attention heads, L the transformer layers, M^2 the image patches, and α_{ctdl} the attention weight for patch i at time t.

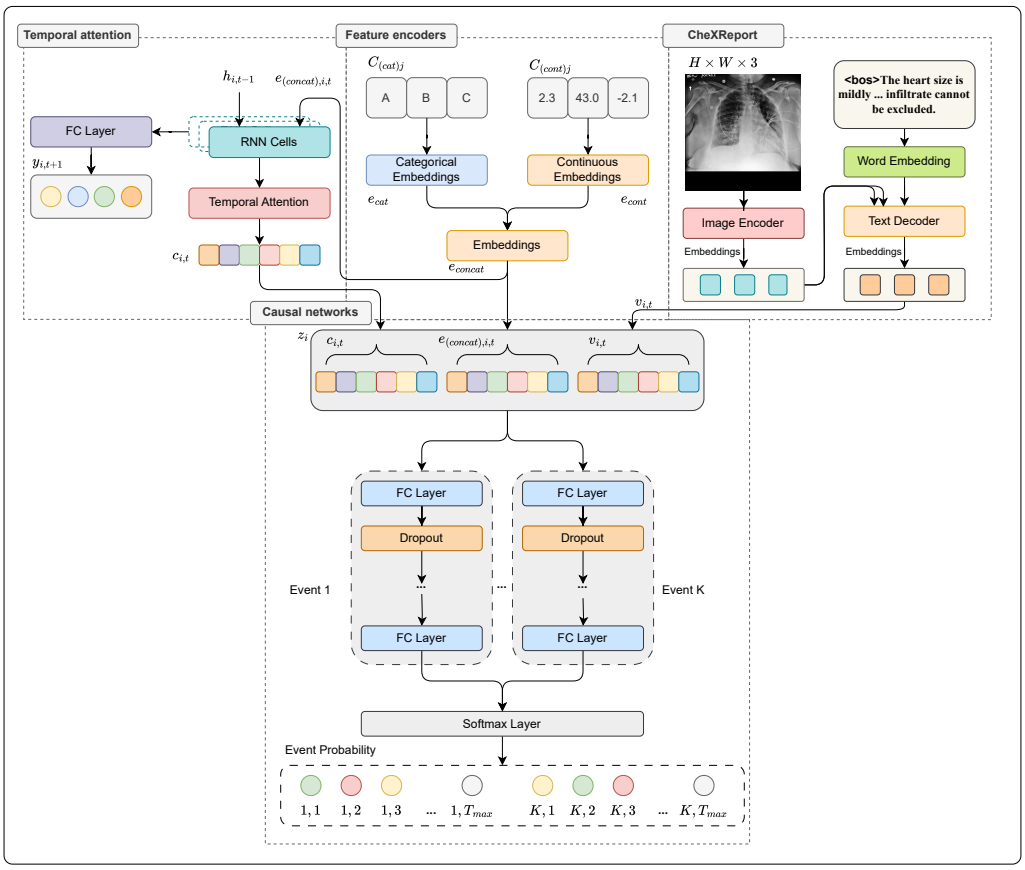


Figure 1. The MultSurv model processes tabular data using feature encoders, generating an embedding vector. For each patient, historical samples are processed by the temporal attention mechanism, producing a contextual vector. Chest X-ray images are analyzed by CheXReport, which extracts a latent space representation of key features. These vectors are concatenated and fed into multitask networks to estimate event probabilities for each time and event.

3. Results and Discussion

To validate the MultSurv model, we conducted an ablation study following an incremental approach. We first assessed the impact of using embeddings for categorical and continuous variables by comparing Model B to the baseline Model A, which lacks embeddings. Next, we explored feature extraction strategies for X-ray images, evaluating a convolutional classification model (Model C) against a transformer-based report generation model (MultSurv model). For methods based purely on data, we only use sociodemographic, clinical, and laboratory data for training. All methods were trained with the same dataset (MDH Dataset). In Table 1, we compare the results obtained by Models A, B, C, and MultSurv in terms of C-index.

Table 1. Comparison of MultSurv model in relation to various methods for the C-index (average and $\pm$ standard deviation). The bigger, the better.

Algorithms	$\Delta t = 1$	$\Delta t = 3$	$\Delta t = 5$	$\Delta t = 7$
Prediction Time $t = 1$				
CoxTime†	0.377 ± 0.08	0.312 ± 0.01	0.380 ± 0.01	0.406 ± 0.04
CoxCC†	0.463 ± 0.07	0.340 ± 0.01	0.400 ± 0.01	0.436 ± 0.02
DeepSurv†	0.361 ± 0.01	0.326 ± 0.07	0.378 ± 0.09	0.410 ± 0.02
PCHazard†	0.555 ± 0.08	0.552 ± 0.01	0.506 ± 0.03	0.524 ± 0.03
DeepHit†	0.533 ± 0.01	0.462 ± 0.01	0.480 ± 0.06	0.456 ± 0.07
N-MTLR†	0.453 ± 0.01	0.452 ± 0.06	0.465 ± 0.03	0.453 ± 0.05
Model A†	0.666 ± 0.02	0.622 ± 0.02	0.629 ± 0.01	0.648 ± 0.01
Model B†	0.693 ± 0.02	0.708 ± 0.02	0.702 ± 0.02	0.701 ± 0.01
Model C	0.695 ± 0.01	0.711 ± 0.02	0.701 ± 0.01	0.703 ± 0.01
MultSurv model	0.723 ± 0.08	0.735 ± 0.01	0.711 ± 0.02	0.706 ± 0.01
Prediction Time $t = 7$				
CoxTime	0.371 ± 0.08	0.404 ± 0.05	0.382 ± 0.03	0.358 ± 0.02
CoxCC†	0.420 ± 0.06	0.443 ± 0.07	0.437 ± 0.04	0.430 ± 0.03
DeepSurv†	0.396 ± 0.06	0.432 ± 0.06	0.440 ± 0.04	0.422 ± 0.03
PCHazard†	0.494 ± 0.09	0.478 ± 0.09	0.496 ± 0.08	0.479 ± 0.08
DeepHit†	0.563 ± 0.01	0.510 ± 0.07	0.520 ± 0.04	0.533 ± 0.07
N-MTLR†	0.420 ± 0.01	0.467 ± 0.05	0.476 ± 0.04	0.452 ± 0.03
Model A†	0.605 ± 0.02	0.617 ± 0.03	0.614 ± 0.02	0.611 ± 0.02
Model B†	0.694 ± 0.03	0.696 ± 0.02	0.695 ± 0.01	0.688 ± 0.01
Model C	0.692 ± 0.02	0.698 ± 0.01	0.696 ± 0.01	0.690 ± 0.01
MultSurv model	0.725 ± 0.01	0.715 ± 0.02	0.702 ± 0.03	0.695 ± 0.03

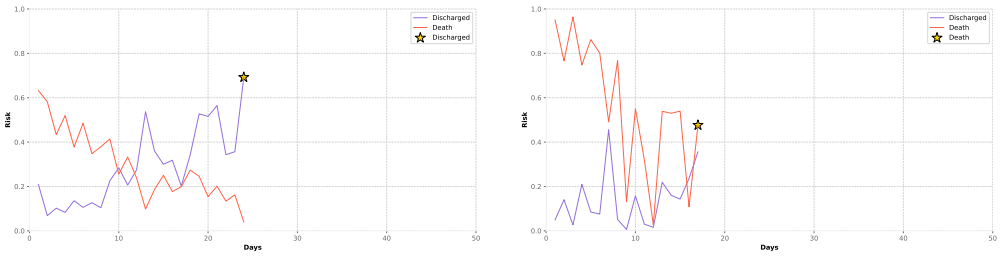
† Trained only with tabular data.

Traditional Cox-based methods, such as CoxTime, CoxCC, and DeepSurv, are widely used for survival analysis due to their simplicity and interpretability. However, they underperform compared to the MultSurv model, particularly in long-term predictions, as they assume proportional hazards, which may not hold in complex clinical scenarios. Although CoxCC achieves a C-index of 0.463 ± 0.07 at $t = 1$, DeepSurv and CoxTime perform worse, and all exhibit limitations in handling temporal variability. While neural network-based approaches like N-MTLR and PCHazard reduce reliance on proportional hazards and improve short-term predictions, they deteriorate faster over time. DeepHit, which models time-dependent covariates and competing risks, further enhances performance but lacks integration of patient health history.

The MultSurv model improves survival analysis by incorporating contextual embeddings for tabular data, capturing non-linear relationships that enhance discrimination and calibration of predictions. While including visual features initially provided minimal performance gains, adopting the CheXReport architecture enables better correlation between radiological reports and visual features, offering richer insights. The Mult-

Surv model achieves a C-index of 0.695 ± 0.03 for $t = 7$ and $\Delta t = 7$, outperforming DeepHit (0.533 ± 0.07), and even Model A (0.611 ± 0.02). These results highlight the advantage of multimodal learning, demonstrating that integrating embeddings and imaging data leads to more accurate survival predictions. By providing a comprehensive health assessment, the MultSurv model can enable healthcare professionals to make more precise and effective treatment decisions, optimizing both patient outcomes and healthcare resource allocation.

We also present a qualitative analysis of the MultSurv model, evaluating risk predictions for death and discharge (Figure 2), and key X-ray regions influencing predictions (Figure 3). For Patient A, the MultSurv model correctly predicted discharge, progressively increasing the associated risk as hospitalization progressed, suggesting its ability to detect health improvements. Additionally, the model consistently assigned a higher initial risk of death, likely reflecting the severity of hospitalized cases. On the other hand, Patient B exhibited fluctuations in the predicted risk of death, indicating variations in health status over time. These fluctuations highlight the importance of continuous monitoring and dynamic risk assessment, reinforcing the need for adaptive treatment strategies.



Patient A Patient B

Figure 2. Independent risks for discharge and death for patients in the test set over the 50 days. In purple is the risk for discharge, and in red is the risk for death. The star indicates when and what event happened.

When we analyze the attention of the CheXReport architecture on the specific patient's X-ray image (Figure 3), we can see that the model gives greater attention to the upper areas of the lungs, as indicated by the darker regions in the attention scale. This distribution suggests that the model identifies these areas as critical for assessing the patient's health status. Focusing attention on these regions may be associated with the detection of common anomalies in cases of COVID-19, such as lung opacities or infiltrations, which are often observed in the upper part of the lungs.

Figure 3. Attention over the image for the last layer of the CheXReport.

3.1. Main Publications

- Zeiser, et al. CheXReport: A transformer-based architecture to generate chest X-ray reports suggestions. **Expert Systems with Applications**, 2024.
- Zeiser, et al. Integration of Epidemiologic, Socioeconomic, and Sociodemographic Indicators to Predict Early COVID-19 In-Hospital Outcomes. In: **20th Encontro Nacional de Inteligência Artificial e Computacional**, 2023.
- Zeiser, et al. First and second COVID-19 waves in Brazil: A cross-sectional study of patients' characteristics related to hospitalization and in-hospital mortality. **The Lancet Regional Health - Americas**, 2022.
- Zeiser et al. Evaluation of Convolutional Neural Networks for COVID-19 Classification on Chest X-Rays. In: **X Brazilian Conf. on Intelligent System**, 2021.
- Zeiser et al. DeepBatch: A hybrid deep learning model for interpretable diagnosis of breast cancer in whole-slide images. **Expert Systems with Applications**, 2021.
- Zeiser et al. Breast cancer intelligent analysis of histopathological data: A systematic review. **Applied Soft Computing**, 2021.

4. Conclusion

We proposed MultSurv model, a deep learning architecture designed for survival analysis using dynamic multimodal data. The research addressed gaps in existing methods for integrating tabular, temporal, and imaging data with concurrent risk modeling. The proposed approach has the ability to capture nonlinear relationships in patient data through embeddings. Furthermore, we propose CheXReport for radiological finding suggestions. Finally, MultSurv model can provide explainable risk predictions. Experimental results demonstrated the superiority of MultSurv model over traditional and unimodal models, achieving better C-index scores.

Future work should prioritize expanding datasets across diverse hospitals, integrating more imaging modalities, and improving model adaptability via continual learning. Prospective validation in clinical settings is vital to evaluate its effect on decision-making and patient outcomes. Additionally, addressing ethical issues, enhancing transparency, and ensuring secure deployment are key for broader adoption. This approach establishes a foundation for applying survival analysis models to other critical care contexts.

Acknowledgment

The authors would like to thank CAPES (C.F. 001) and CNPq (No. 309537/2020-7).

References

Leach et al. (2021). Post-pandemic transformations: How and why covid-19 requires us to rethink development. *World development*.

Lee et al. (2019). Dynamic-deephit: A deep learning approach for dynamic survival analysis with competing risks based on longitudinal data. *IEEE Transactions on Biomedical Engineering*.

Rahman et al. (2020). Zoonotic diseases: etiology, impact, and control. *Microorganisms*.

Zeiser, F. A. et al. (2022). First and second covid-19 waves in brazil: A cross-sectional study of patients' characteristics related to hospitalization and in-hospital mortality. *The Lancet Regional Health-Americas*.

Zeiser, F. A. et al. (2024). Chexreport: A transformer-based architecture to generate chest x-ray reports suggestions. *Expert Systems with Applications*.