2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Deep Neural Network-Based Survival Analysis for Skin Cancer Prediction in Heart Transplant Recipients*

Kuo-Chun Chiu[1], Dongping Du[1], Nandini Nair[2], Yuncheng Du[3]

Abstract— **Heart-transplant recipients are at high risk of developing skin cancer, while Squamous Cell Carcinoma (SCC) and Basal Cell Carcinoma (BCC) are commonly detected. This paper utilized the database from the United Network for Organ Sharing (UNOS) to study the incidence rate of SCC and BCC among heart transplant recipients. Cox proportional hazards model and two deep neural network-based models were studied, and their performance were compared. In addition, Lasso regression, Chi-square test, and Wilcoxon signed-rank test were applied to identify key risk factors. The neural network-based survival models showed better accuracy compared to the standard Cox regression model, which indicates the advantage of deep learning approaches in survival analysis and risk prediction for post-transplant skin cancer.**

Clinical relevance— **This study investigates the performance of deep learning (DL) models in clinical applications for predicting the risk of skin cancer in heart transplant recipients. The DL models outperform the standard models in assessing the incidence rate of skin cancer across different time spans.**

I. INTRODUCTION

Heart transplantation substitutes a healthier donor heart for a weakened heart to extend the survival of end-stage heart failure patients. The transplanted heart will be regarded as an outsider by the immune system of recipients and triggers rejection. Hence, organ transplant recipients need to take immunosuppressive drugs to reduce the capability of the immune system rejecting the new heart. Unfortunately, immunosuppressive drugs is a double-edged sword. It could weaken immune system and increase the chance of developing cancers. As reported previously, heart transplant recipients are at a higher risk of developing cancer, i.e., 2.9% of cancer rate at 1 year and 31.9% at 10 years [1]. More, skin cancer is the most common cancer type among organ transplant recipients. The study by Oechslin et al [2] shows 54% of the post-transplant malignancy belongs to skin cancer. Therefore, it is crucial to predict and screen skin cancer for improving long term survival.

The two most commonly detected skin cancers among heart transplant recipients are Squamous Cell Carcinoma (SCC) and Basal Cell Carcinoma (BCC). There are generally three types of skin cancer, including SCC, BCC, and Melanoma. 40% of the post-transplant skin cancer goes to SCC, and 23% goes to BCC according to United Network for Organ Sharing (UNOS) database between 1985 and 2015. Thus, predicting the incidence rate of post-transplant SCC and BCC plays a pivotal role in post-transplant health management. With risk prediction models, researchers can evaluate the risk of developing skin cancer based on observations from individual patients, e.g., age, gender, etc.

Prevalent analytical models for risk predictions include Kaplan-Meier estimator, Logistic regression, and Cox proportional hazards model. Kaplan-Meier model is commonly used in medical studies [3]. This method calculates the cumulative survival rate as a function of time. The survival probability at any time interval can be estimated by: S_t= (number of survived patients in the beginning - number of deceased patients) / (number of survived patients in the beginning). The visualization of this model is well-known as Kaplan-Meier curve which indicates the survival probability within a time interval. However, Kaplan-Meier estimator has several limitations such as it does not consider risk factors and cannot handle time-varying risk factors. Logistic regression model is another popular model used in medical studies to analyze data with binary labels such as "Cancer" or "No-cancer", but it cannot evaluate how risk varies over time. Cox proportional hazards model is one of the extensively used survival models. This method estimates the contribution of covariates from individual subjects to the probability of an event, e.g., patient gets cancer.

Nair et al. applied Cox model to identify key risk factors with regard to the development of cutaneous SCC (cSCC) in a post heart transplant study [4]. The model offers Area Under the ROC Curves (AUCs) of 0.77-0.79 on prediction performance. Furthermore, Cox proportional hazards model was also utilized to study the effect of risk factors on the formation of post-heart transplant BCC by Nair et al [5]. The AUCs of the prediction at 5, 8 and 10 years after heart transplant are between 0.73 and 0.74. Cox model assumes that the hazard function of individual subject can be inferred from a linear combination of some risk factors (covariates). However, in most real-life cases, such a linear combination is not efficient in describing the complex nonlinear relationships among the covariates. Therefore, survival models with nonlinear hazard functions were developed to address the limitation. A feed-forward neural network-based nonlinear proportional hazards model was proposed by Faraggi et al [6]. This model applied a neural network to evaluate the log-

*This work was supported by National Science Foundation (CMMI-1728338, CMMI-1727487)

[1]Kuo-Chun Chiu and Dongping Du are with the Department of Industrial, Manufacturing, and Systems Engineering, Texas Tech University, Lubbock, TX 79409, USA . (E-mails: kuchiu@ttu.edu,dongping.du@ ttu.edu)

[2]Nandini Nair is with Center for Cardiovascular Health, Texas Tech University Health Science Center, Lubbock, TX 79430 USA. Nandini.Nair@ttuhsc.edu

[3]Yuncheng Du is with the the Department of Chemical and Biomolecular Engineering, Clarkson University, Potsdam, NY 13699 USA. ydu@clarkson.edu

risk function and modeled the relationship among nonlinear covariates to the potential risk. Further, Katzman et al. [7] developed a deep neural network-based Cox proportional hazards model called "DeepSurv" to compute complex interactions between covariates of interest. Moreover, a tree method-based non-linear survival model, named random survival forest (RSF), was developed to evaluate the cumulative hazard function in a medical study [8].

Capitalizing on the advances of survival models, we proposed to apply the neural network-based survival models to predict the risk of skin cancer among heart transplant recipients and compare their performance to the standard Cox proportional hazards model. The rest of the paper is organized as follows. Section II provides the mathematical details of the three models. Section III describes the risk factors and the UNOS dataset. Section IV gives the descriptions on feature selection and model performance evaluation, and Section V presents the experimental results. Finally, the conclusions of this study are summarized in section VI.

II. Methods

In survival analysis, time to event (e.g., development of skin cancer) and/or censoring data (e.g., follow up time) are collected to infer the probability of the event as a function of time. Cox proportional hazards model computes the probability of event from a set of covariates in a linear relationship. While, the neural network-based approaches capture a more complex relationship among the covariates. Two advanced neural network-based models will be discussed in this section, which are Cox proportional hazards deep neural network [7] and a neural network-based non-proportional Cox model [9].

A. Cox Proportional Hazards Model

Standard Cox proportional hazards model defines the survival function as [10]

$$S(t) = P(T \leq t) = 1 - F(t) \tag{1}$$

where $F(t)$ is the cumulative distribution function of an time to event T (e.g., development of skin cancer). The survival function can be derived from a hazard function as

$$S(t) = exp[-\int_0^t h(s)ds] \tag{2}$$

The hazard function is further determined by the covariates $\mathbf{x}$ and a baseline hazard function $h_0(t)$ as

$$h(t|\mathbf{x}) = h_0(t)exp[g(\mathbf{x})] \tag{3}$$

where $g(\mathbf{x}) = \beta^T\mathbf{x}$, and $exp[g(\mathbf{x})]$ is referred to as "hazard ratio", "risk score", or "relative risk function"; β is a coefficient vector which measures the influence of each covariate on the development of a event. The coefficient vector can be inferred by maximizing the Cox partial likelihood

$$\begin{aligned} L(\beta) &= \prod_{i=1}^{n} \left(\frac{h_i}{\sum_{j \in R_i} h_i} \right)^{D_i} \\ &= \prod_{i=1}^{n} \left(\frac{h_0 exp(\beta^T x_i)}{\sum_{j \in R_i} h_0 exp(\beta^T x_j)} \right)^{D_i} \end{aligned} \tag{4}$$

where n is number of subjects, and D_i stands for event/censoring indicator where $D_i = 1$ indicates a cancer event and $D_i = 0$ indicates loss to follow-up. For an individual subject i, let T_i be the possible censoring time, R_i contains all subjects who were at risk, alive, and uncensored at T_i. To simplify the calculation, the *log*-partial likelihood is derived

$$\ell(\beta) = \sum_{i=1}^{n} D_i \left(\beta^T x_i - ln \left(\sum_{j \in R_i} exp(\beta^T x_j) \right) \right) \tag{5}$$

By setting $\ell(\beta) = 0$, β can be estimated by maximizing log-partial likelihood.

B. Neural Network-Based Cox Proportional Hazards Model

Katzman et al. developed a deep neural network Cox model in 2018 [7]. While this method has a similar structure as the Cox proportional hazards model, deep neural network was used to evaluates the risk, i.e., $\beta^T x$. This advanced model is capable of assessing a more complex relationship among covariates of the interested dataset.

To train the feed-forward neural network, the objective function is set as the averaged log negative partial-likelihood with l_2 regularization

$$\ell(w) = -\frac{1}{n} \sum_{i:D_i=1} \left(\hat{h}(x_i|w) - ln \sum_{j \in R_i} exp[\hat{h}(x_j|w)] \right) + \lambda ||w||_2^2 \tag{6}$$

where n denotes the total number of subjects who suffered the event and λ is regularization parameter. R_i set indicates the subjects who were still at risk, alive, and uncensored at a point of time at T_i. By minimizing equation (6), the weights of neural network, w, among covariates can be estimated.

C. Neural Network Based Non-Proportional Cox Model

Neural Network Based Non-Proportional Cox Time Model is another extension of the neural network-based Cox proportional hazards model developed by Kvamme et al. [9]. This technique takes covariates and time into consideration in the hazard ratio term to capture the interactive relationship between covariates and time

$$h(t|\mathbf{x}) = h_0(t)exp[g(t, \mathbf{x}|w)] \tag{7}$$

$g(t, \mathbf{x})$ is no longer a linear time-independent function. Instead, it is time-varying and estimated by a deep neural

network. Thus, CoxTime can model interactions between covariates and time. The objective function is

$$\ell(\theta) = -\frac{1}{n}\sum_{i:D_i=1}\left(g(T_i, x_i|\theta) - ln\sum_{j\in R_i} exp[h(x_j|w)]\right) + \lambda\sum_{i:D_i=1}\sum_{j\in R_i}|g(x_j)| \quad (8)$$

where $\lambda\sum_{i:D_i=1}\sum_{j\in R_i}|g(x_j)|$ is the penalty on $g(\mathbf{x})$, and λ is the tuning parameter. The objective function is averaged by n which is the number of events in the dataset.

III. Dataset

UNOS dataset which collects organ transplant recipients' data of waiting list, matching, post-transplant follow-ups was used to develop the risk models for post-heart-transplant skin cancer. The original dataset contains 152,095 subjects with 490 covariates collected between 1985 and 2015. We kept the covariates and the recipients that meet the following criteria. 1. The covariates with low ratio of missing data. 2. Patients with age greater than 18. 3. Transplant organ is heart. 4. Successful transplant surgery. 5. Post-heart transplant patients had or were free from BCC or SCC. 6. Heart transplant surgery was performed between January 01, 2000 and December 31, 2005. The final cohort includes 20,205 eligible heart transplant recipients. 39 covariates were extracted, which are listed in Table I. Time to event or censoring were defined as the time span by day from the completion of the heart transplant to the date of being diagnosed with skin cancer (i.e., SCC and BCC) or loss to follow-up. The number of censoring and events are 18,715 and 1,490 respectively in a duration of 5,130 days, and the ratio is 93:7 (censoring: event).

TABLE I
Description of attributes

Recipient Age*	Recipient-Donor Ethnicity Match*
Donor Age	Recipient-Donor Gender Match*
Recipient Blood Type*	Induction with OKT3*
Donor Blood Type*	Induction with Basiliximab *
Recipient Weight*	Induction with Daclizumab *
Donor Weight	Recipients Tattoos*
Recipient Height*	Donor Cancer History
Donor Height	Number of Previous Transplants
Recipient BMI	Recipient Creatinine*
BMI	Days on Waiting List*
Recipient Diabetes	Latitude of Recipient's Location *
Donor Diabetes	Distance between Donor's Residence to Hospital
Recipient Bilirubin*	Recipient's status at Tx*
Donor Bilirubin	Recipient A1 antigen*
Recipient Gender*	Recipient A2 antigen*
Donor Gender*	Recipient B1 antigen
Recipient Ethnicity*	Recipient B1 antigen
Donor Ethnicity*	Recipient DR1 antigen
Donor Hypertension History	Recipient DR2 antigen*
Induction with Thymoglobulin	

* denotes significant covariate verified by Lasso regression, Chi-square test, and Wilcoxon signed-rank test with p-value < 0.05

IV. Design of Experiments

Feature selection was performed on the dataset with 39 covariates by Lasso regression, Chi-square test, and Wilcoxon signed-rank test. Lasso regression screened out the significant covariates with non-zero coefficients. Next, categorical features underwent Chi-square test, and Wilcoxon signed-rank test was carried out for numerical features. The significant covariates selected by the three methods were considered to be highly correlated with SCC or BCC and were included in the survival models.

Subsequently, we performed 5-fold cross-validation. The dataset was split into a testing set (20%) and a training set (80%). 20% of the training set was then selected as the validation set, and the remaining data was used as the final training set. The training set and the validation set were used to train the model and tune the hyperparameters, and the testing set was utilized for performance evaluation.

Receiver Operating Curves (ROCs) were computed to derive Area Under the Curves (AUCs) to assess the prediction performance 1 year, 3 years, 5 years, and 10 years after heart transplant surgeries. In addition, Brier score and concordance (c-index) were used to evaluate the models' predictive capability. AUCs and concordance are the higher the better, yet Brier score is the lower the better. All the experiments were implemented through Python.

V. RESULTS

A. Feature Selection

39 covariates in the UNOS dataset were initially selected. Then, Lasso regression was carried out and 12 covariates had non-zero coefficients after L1 penalization. These covariates were significant features adversely or positively contributing to the development of SCC or BCC skin cancer.

We also performed Chi-square test, and Wilcoxon signed-rank test to ascertain significant covariates. The dataset excluding "time to event" and "event labels" consists 19 categorical and 20 numerical attributes. To identify the significant covariates, Chi-square test was done to assess the effects of categorical features on the cancer event, and 11 of them were selected as significant with p-value < 0.05.

A similar procedure was implemented for numerical covariates. Each numerical attribute was split by event into two groups, i.e., skin cancer and no-skin cancer, and was analyzed by Wilcoxon signed-rank test. Finally, 9 significant numerical features were identified. The significant covariates (see Table I) were used in the survival models to predict the risk of developing skin cancer after heart transplantation.

B. Risk Prediction

The significant covariates were used to develop three different survival models, i.e., Cox proportional hazards model, neural network-based Cox model (DeepSurv), and neural network-based non-proportional Cox model(CoxTime), and their prediction performance were evaluated by Brier score and concordance. More, AUCs was applied to assess the models' prediction capability 1 year, 3 years, 5 years, and 10 years after heart transplantation.

Fig. 1 (left) shows the Brier score of the three models. The average Brier scores of the two neural network-based models were 0.091±0.0089 and 0.089±0.0084, respectively. Cox proportional hazards model had a higher average Brier score of 0.111±0.0126, which indicates this standard model falls behind the neural network-based models on prediction. Between the CoxTime model and the DeepSurv model, the CoxTime model slightly outperforms the DeepSurv model. Likewise, we got a similar conclusion from the concordance

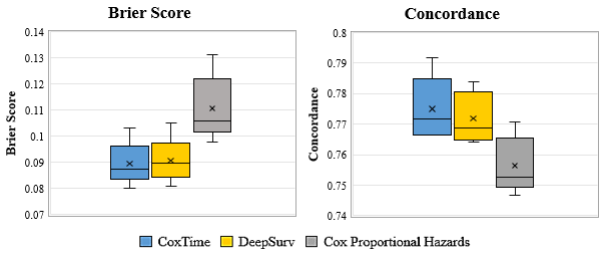

Fig. 1. Brier score and Concordance of CoxTime, DeepSurv, and Cox Proportional Hazards Models respectively

as seen in Fig. 1 (right). Concordance measures the percentage of pairs, where true incidence rates of skin cancer are greater than the score of negative cancer event. The higher the concordance, the better the performance of the model is. It is obvious that neural network-based models outperformed Cox proportional hazards model. The concordance of DeepSurv, CoxTime, and Cox proportional hazards model are 0.772±0.0084, 0.775±0.0105, and 0.756±0.0092, respectively.

Further, we computed AUCs to compare the model performance at 1 year, 3 years, 5 years, and 10 years after heart transplantation. Still, DeepSurv and CoxTime significantly surpassed Cox proportional hazards model in all four checkpoints as shown in Fig. 2 and Table II.

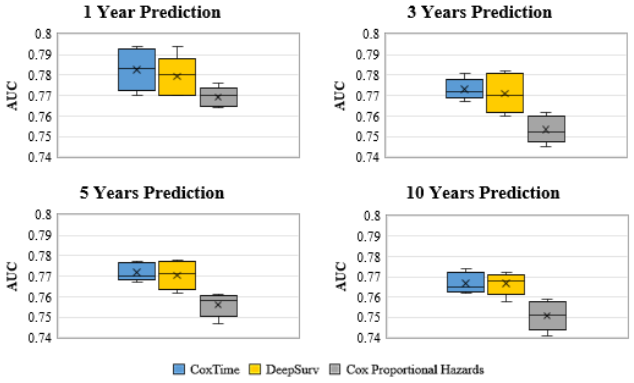

Fig. 2. AUCs of 1-year, 3-year, 5-year, and 10-year prediction of skin cancer by CoxTime, DeepSurv, and Cox Proportional Hazards models respectively

VI. CONCLUSIONS

This study identified 23 signficant risk factors highly associated with post-transplant skin cancer event, which include recipient's age, blood type, height, weight, creatinine, bilirubin, gender, ethnicity, tattoos, latitude, days on waiting list, and status at transplant, donor's blood type and ethnicity, inductions with OKT3, Basiliximab, and Daclizumab, A1, A2, and DR2 antigens, donor and recipient's ethnicity and gender matches. These risk factors are used in different survival models to predict the incidence rate of post-heart-transplant skin cancer. The Brier score and Concordance indicated that the neural network-based models had superior performance over the traditional Cox proportional hazards model, and the AUCs at different time points after heart transplant proved the consistent finding. This study demonstrated that deep neural network-based models can capture the complex relationship among covariates, thus providing more accurate predictions than the traditional Cox model.

TABLE II

AUCs OF 1, 3, 5, AND 10 YEARS PREDICTION OF POST-TRANSPLANT SKIN CANCER

	CT	DS	CPH	CT	DS	CPH
	1-year prediction			3-year prediction		
Mean	0.783	0.779	0.769	0.773	0.771	0.753
Std.	0.0091	0.0089	0.0044	0.0047	0.0088	0.0060
	5-year prediction			10-year prediction		
Mean	0.772	0.770	0.756	0.767	0.767	0.751
Std.	0.0040	0.0062	0.0051	0.0045	0.0049	0.0064

*DS:DeepSurv, CT:CoxTime, CPH: Cox Proportional Hazards

ACKNOWLEDGMENT

The authors would like to acknowledge the funding support from National Science Foundation USA (CMMI-1728338, CMMI-1727487) and The Medical Practice Income Plan (MPIP) grant from Texas Tech Health Sciences Center.

REFERENCES

[1] N. Lateef, K. A. Basit, N. Abbasi, S. M. H. Kazmi, A. B. Ansari, and M. Shah, "Malignancies after heart transplant," *Exp Clin Transpl*, vol. 14, pp. 12–6, 2016.

[2] E. Oechslin, W. Kiowski, J. Schneider, F. Follath, M. Turina, and A. Gallino, "Pretransplant malignancy in candidates and posttransplant malignancy in recipients of cardiac transplantation," *Annals of oncology*, vol. 7, no. 10, pp. 1059–1063, 1996.

[3] M. K. Goel, P. Khanna, and K. Jugal, "Goel mk, khanna p, kishore j. understanding survival analysis: Kaplan-meier estimate," *Int J Ayurveda Res.*, vol. 1, pp. 274–278, 2010.

[4] N. Nair, Z. Hu, D. Du, and E. Gongora, "Risk prediction model for cutaneous squamous cell carcinoma in adult cardiac allograft recipients," *World Journal of Transplantation*, vol. 11, no. 3, p. 54, 2021.

[5] N. Nair, D. Du, Z. Hu, and E. Gongora, "Risk prediction model for basal cell carcinoma in cardiac allograft recipients," *The Journal of Heart and Lung Transplantation*, vol. 39, no. 4, p. S128, 2020.

[6] D. Faraggi and R. Simon, "A neural network model for survival data," *Statistics in medicine*, vol. 14, no. 1, pp. 73–82, 1995.

[7] J. L. Katzman, U. Shaham, A. Cloninger, J. Bates, T. Jiang, and Y. Kluger, "Deepsurv: personalized treatment recommender system using a cox proportional hazards deep neural network," *BMC medical research methodology*, vol. 18, no. 1, pp. 1–12, 2018.

[8] H. Ishwaran, U. B. Kogalur, E. H. Blackstone, M. S. Lauer *et al.*, "Random survival forests," *Annals of Applied Statistics*, vol. 2, no. 3, pp. 841–860, 2008.

[9] H. Kvamme, Ø. Borgan, and I. Scheel, "Time-to-event prediction with neural networks and cox regression," *arXiv preprint arXiv:1907.00825*, 2019.

[10] P. M. Therneau, Terry M.and Grambsch, *Modeling survival data: Extending the cox model*. New York: Springer, 2000.