Proceedings of CONF-APMM 2025 Symposium: Controlling Robotic Manipulator Using PWM Signals with Microcontrollers
DOI: 10.54254/2753-8818/2025.AD26245

Construction and Validation of a Survival Prediction Model for Eye Cancer Patients Based on Cox Regression

Ming Gao

Department of Mathematics and Applied Mathematics, Nanchang University, Nanchang, China
2DA6047F6AF500BD@outlook.com

Abstract. Cancer poses significant health risks to humans, and its prognosis plays a crucial role in treatment decisions. This study developed and validated a survival prediction model for eye cancer patients using Cox regression. Eye Cancer Patients dataset from Kaggle was utilized, containing 5,000 patient records. Using random sampling, the cohort was divided into a training set (n=3,421, 70%) for model development and an independent validation set (n=1,467, 30%) for performance evaluation. Prognostic factors were initially screened via LASSO regression, followed by multivariate Cox analysis to derive a risk score. This score was then used as a univariate Cox variable to build a nomogram. Model performance was validated through calibration curves and decision curve analysis. Treatment type, chemotherapy, family history, and country were identified as combined prognostic factors for overall survival. The constructed logistic regression model demonstrated excellent calibration curve fit, while the decision curve indicated high clinical utility. The Cox regression-based logistic regression model effectively predicts overall survival in eye cancer patients.

Keywords: Eye cancer, Total survival, Cox regression

1. Introduction

Cancer mortality rates have been declining steadily until 2022, yet the incidence of common cancers continues to rise and pose significant health risks [1]. Ocular carcinoma, a prevalent and highly aggressive type of cancer, encompasses malignant tumors affecting various eye tissues including eyelids, conjunctiva, eyeball structures, and ocular appendages. These include sarcomas, lymphoma, malignant eyelid carcinoma, and ocular surface neoplasms (OSSNs) [2]. The incidence patterns vary significantly by age group: retinoblastoma predominantly affects children under five, while choroidal melanoma is more common among adults. These malignancies not only present challenges for early detection but also cause substantial damage to ocular structures and visual functions. When cancer cells metastasize to the brain and proliferate, they may become life-threatening. Effective prevention, treatment, and prognosis remain critical components of care, making survival prediction particularly vital for ocular carcinoma patients [3].

In terms of cancer prognosis, many people have used different approaches to study it. Liu and colleagues developed a survival prediction model using univariate and multivariate Cox regression analysis, revealing that previously recognized prognostic factors in middle-aged and elderly patients

© 2025 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2753-8818/2025.AD26245

with liver metastases from rectal adenocarcinoma, such as differentiation degree, PI, and T/N staging, no longer demonstrated predictive value [4]. Xie and colleagues compared line graph models with traditional staging systems, demonstrating the line graph model's superior performance in identifying high-risk patients [5]. Pan and colleagues identified characteristic genes through differential expression analysis (DEGs) of three molecular clusters (MRGs) using LASSO and COX regression analyses [6].

Zhang and colleagues utilized the Fine-Gray sub-distribution risk model and Random Survival Forest method to predict postoperative CSS in elderly stage I-III CRC patients. Their results highlighted the critical need to account for competing risks, as omission led to statistically significant overestimation of CSS ($p<0.05$) [7]. To improve survival prediction for PCBM patients, Zhang and colleagues developed and validated an XGBoost model with fivefold cross-validation, which outperformed other machine learning models [8]. Zhao G. and Li D. proposed the TreeEM model, a novel computational framework for multi-omics data analysis. Experimental validation on the BRCA (breast cancer) and ARCENE datasets confirmed that TreeEM achieves significantly higher prediction accuracy ($p<0.01$) than existing models, highlighting its potential for precision oncology applications [9].

Previous studies indicate that survival prediction models are typically constructed through two approaches: traditional survival prediction and machine learning-based survival prediction. However, machine learning survival prediction demonstrates greater accuracy in individualized predictions but lacks explanatory power regarding variable impacts [3]. Given this context, our study focuses on identifying survival factors for eye cancer patients rather than individualized prediction, thus adopting traditional survival prediction methods. Wang Xin, Peng Fanyu, Zong Dan, and colleagues proposed the LASSO-Cox regression method, which validated clinical efficacy and risk stratification capabilities through model evaluation metrics including fit, discrimination, calibration, and significance differences [10]. Building on this foundation, the peopple propose a LASSO-Cox regression model to identify survival factors in eye cancer patients. Statistical tests on fit and significance differences further confirm the model's validity.

2. Method

2.1. Data acquisition and preprocessing

This study extracted the Eye Cancer Patients dataset from the Kaggle database. The dataset contains 14 features and 5,000 samples. Through screening, we identified 112 missing values: 112 cases with zero radiotherapy sessions in treatment plans and 112 cases with zero chemotherapy doses in treatment plans. We adopted a direct deletion method to handle these missing values. Consequently, the experimental dataset was reduced to 14 columns and 4,888 rows.

2.2. Indicator selection and explanation

Clinically, the 5-year survival status is generally used to evaluate the treatment effect of cancer. Therefore, the predicted survival time in this paper is 60 months. Meanwhile, Variables were coded according to predefined criteria (Table 1).

DOI: 10.54254/2753-8818/2025.AD26245

Table 1. Variable setting and assignment table

Variable	Type	Levels
Age	Date	Quantitative Date
Gender	Category	1=M,2=F,3=Other
Cancer Type	Category	1=Retinoblastoma,2=Melanoma,3=Lymphoma
Laterality	Category	1=Left,2=Right,3=Bilateral
Stage at Diagnosis	Category	1=Stage I,2=Stage II,3=Stage III,4=Stage IV
Treatment Type	Category	1=Radiation,2=Chemotherapy,3=Surgery
Surgery Status	Category	1=FALSE,2=TRUE
Radiation Therapy	Date	Quantitative Date
Chemotherapy	Date	Quantitative Date
Outcome Status	Category	1=Deceased,2=Active,3=In Remission
Survival Time Months	Date	Quantitative Date
Genetic Markers	Category	1=Non,2=BRAF Mutation
Family History	Category	1=FALSE,2=TRUE
Country	Category	1=Australia,2=Brazil,3=Canada,4=France,5=Germany, 6=India,7=Japan,8=South Africa,9=UK,10=USA

2.3. Method introduction

This study employed the Cox regression method for data analysis (Figure 1). Using random sampling, the cohort was divided into a training set (n=3,421, 70%) for model development and an independent validation set (n=1,467, 30%) for performance evaluation. Variables were compared between sets using t-tests/χ^2 tests. Prognostic factors were screened by LASSO regression (λ: 1-SE) and finalized via multivariate Cox analysis to derive the risk score. This score was then used as a univariate Cox variable to build a nomogram. Model performance was validated through calibration curves (assessing the concordance of model-predicted versus actual outcomes) and decision curve analysis (DCA, measuring clinical net benefit across varying risk thresholds). With the research objective of developing a statistically rigorous and widely applicable survival prediction model, strict statistical screening criteria ($P<0.05$) were applied throughout the study.



Figure 1. LASSO-Cox regression analysis flowchart (picture credit: original)

3. Results and discussion

3.1. Model construction

The preprocessed dataset in this study included 4,888 eye cancer patients with a mean age of 50.5 years (range 1-90). The mean follow-up time was 97 months (range 1-120 months). Using random sampling, the cohort was divided into a training set (n=3,421, 70%) for model development and an independent validation set (n=1,467, 30%) for performance evaluation. Demographic and clinical

DOI: 10.54254/2753-8818/2025.AD26245

characteristics were well-balanced between the independent training (n=3,421, 70%) and validation (n=1,467, 30%) cohorts, with all between-group comparisons showing P values > 0.05 (Table 2).

Table 2. The differences of baseline features between training set and validation

Characteristic	Statistic	P
Age ($\bar{x} \pm s$, years old)	-0.199	0.448
Gender	0.904	0.924
Cancer Type	3.185	0.107
Laterality	3.199	0.525
Stage at Diagnosis	1.344	0.501
Treatment Type	0.214	0.298
Surgery Status	0.132	0.716
Radiation Therapy($\bar{x} \pm s$,Gy)	-2.03	0.831
Chemotherapy($\bar{x} \pm s$,cycles)	0.262	0.197
Outcome Status	0.132	0.716
Survival Time Months($\bar{x} \pm s$,months)	-0.071	0.931
Genetic Markers	0.006	0.937
Family History	1.047	0.306
Country	88.576	0.264

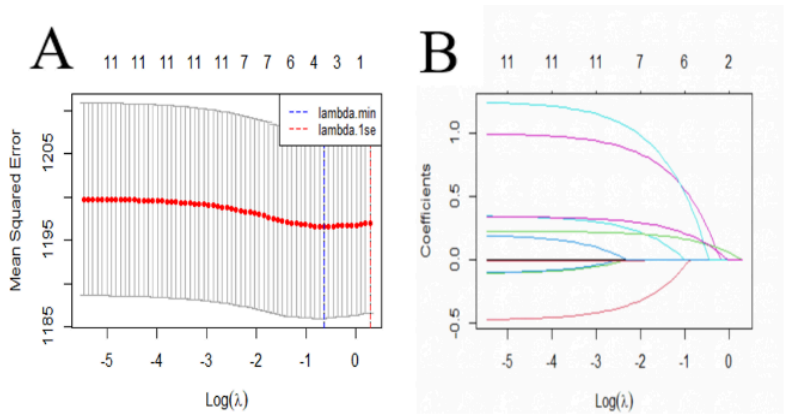


Figure 2. LASSO regression analysis. (A: Cross-validated MSE (mean ± SE) for λ selection. The dotted line indicates the optimal λ (minimum MSE + 1 SE. B: Coefficient shrinkage paths. Variables surviving at optimal λ are highlighted in red.) (Picture credit: Original)

After variable screening of the data set using LASSO regression analysis (Figure 2).

The optimal regularization parameter (0.99) was selected by using 10-fold cross-validation, and the variables with non-zero regression coefficients were Treatment Type, Chemotherapy, Family History and Country.

DOI: 10.54254/2753-8818/2025.AD26245

Table 3. Coefficients of risk scoring calculation formula

	exp(coef)
Treatment Type=2	1.0983
Treatment Type=3	0.9619
Chemotherapy	0.9922
Family History=2	0.9579
Country=2	1.0026
Country=3	1.0219
Country=4	1.0833
Country=5	0.9796
Country=6	1.0449
Country=7	1.0046
Country=8	0.9791
Country=9	1.0378
Country=10	0.8837

The variables selected through LASSO regression analysis were substituted into a multivariate Cox proportional hazard regression to derive the risk score calculation formula (Table 3): Risk Score =1.0983 × Treatment Type 2 + 0.9619 × Treatment Type 3 + 0.9922 × Chemotherapy + 0.9579 × Family History 2 + 1.0026 × Country2 + 1.0219 × Country3 + 1.0833 × Country4 + 0.9796 × Country5 + 1.0449 × Country6 + 1.0046 × Country7 + 0.9791 × Country8 + 1.0378 × Country9 + 0.8837 × Country 10.

Subsequently, the risk score formula was incorporated into a univariable Cox proportional hazards regression analysis to construct a nomogram model, with visual representation provided in Figure 3.

DOI: 10.54254/2753-8818/2025.AD26245

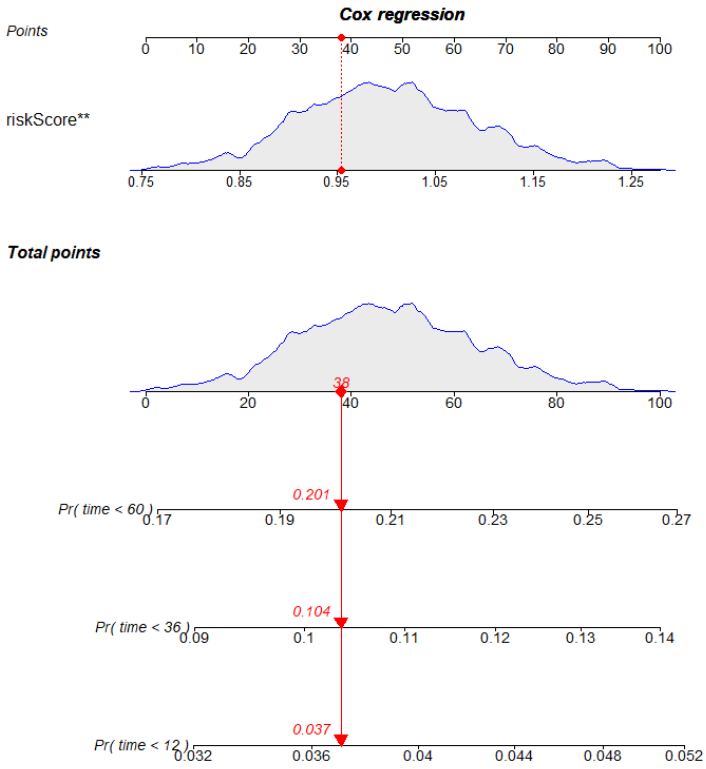


Figure 3. Nomogram for predicting time-related death in eye cancer patients. (picture credit: original)

3.2. Validation model

The constructed line chart model was introduced into the verification set, and the calibration curves of the verification set and training set were well fitted with the diagonal (Figure 4).

DOI: 10.54254/2753-8818/2025.AD26245

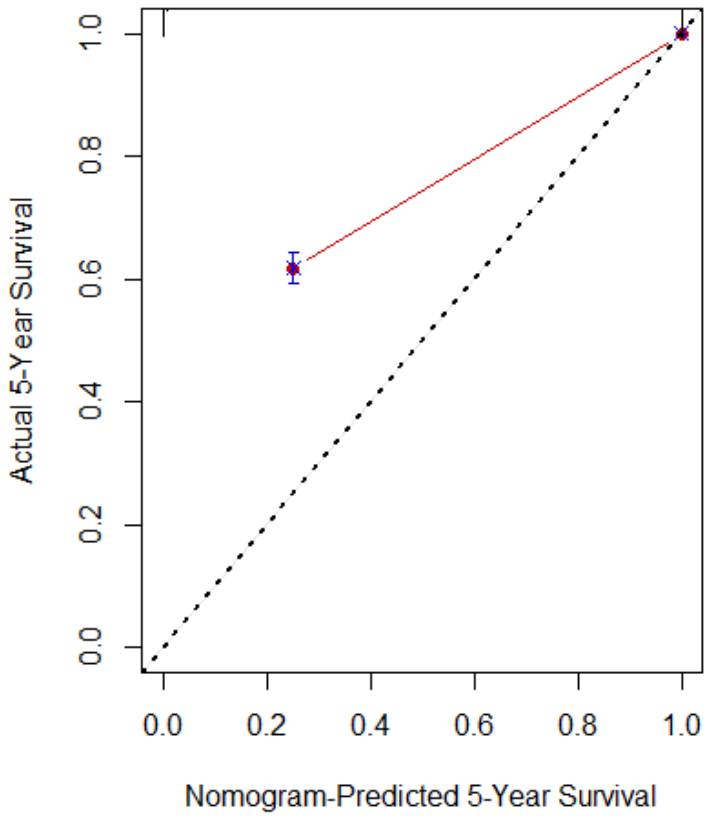


Figure 4. Calibration curves for 5-year overall survival based on the nomogram mode (picture credit: original)

The close alignment of calibration curves with the diagonal line in both training and validation sets indicates strong agreement between predicted and observed 5-year overall survival rates in ocular cancer patients, demonstrating the nomogram model's robust discriminative ability and calibration accurac; DCA curve analysis showed that it had high net benefit and clinical effectiveness (Figure 5).

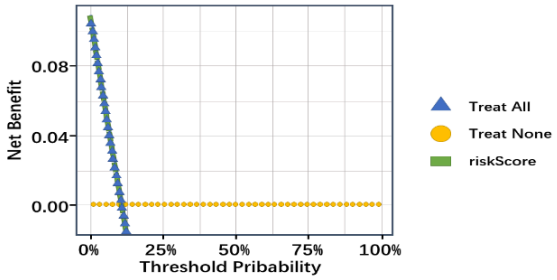


Figure 5. Decision curve analysis of a nomogram model for predicting overall survival in patients with ocular cancer (picture credit: original)

DOI: 10.54254/2753-8818/2025.AD26245

3.3. Model analysis

The DCA curve analysis demonstrates that within the 25%-75% threshold probability range, the risk Score-based decision significantly outperforms both "all treatment" and "no treatment" approaches in net benefit. This indicates the model's effectiveness in optimizing treatment strategies, proving clinically practical for preventing both overtreatment and missed diagnoses. Furthermore, the Cox regression line plot converts clinical variables including Treatment Type, Chemotherapy, Family History, and Country into risk scores to predict survival probabilities at different time points. This provides a prognostic assessment tool for individual patients, assisting clinicians in making informed decisions (e.g., recommending intensified therapy for high-risk patients).

Factors affecting survival prognosis in patients with eye cancer include Family History and Country, and other potential predictors (such as genomic data) can be explored to further improve the accuracy of factors affecting survival prognosis.

4. Conclusions

To investigate survival prognosis factors in ocular carcinoma patients, this study proposes the LASSO-Cox regression method to explore influencing factors and predict survival probabilities. The results indicate that Treatment Type, Chemotherapy, Family History, and Country are common prognostic factors for ocular carcinoma patients. According to evaluation metrics from calibration curves and decision curve analysis (DCA) models, the LASSO-Cox regression model demonstrates good predictive performance and model fit in identifying prognostic factors for ocular carcinoma patients. Future research could further explore other potential variables and more specific family history factors to enhance the accuracy of survival prognosis prediction for ocular carcinoma patients.

References

[1] Nierengarten, B. M. (2025). “Cancer Statistics, 2025” provides data on cancer occurrence and outcomes. Cancer, 131(9): e35826-e35826.
[2] Xu, H., Rokohl, C. A., Ju, X. (2025). Global incidence and trends of ocular cancer: A bibliometric analysis. Advances in Ophthalmology Practice and Research, 5(1): 22-31.
[3] Zhai, Y. J., Li, G. P., C, J. Y. & Zhang, K. X. (2024). Research Progress of Deep Learning in Cancer Survival Prediction. Biomedical Engineering Research, 43(5), 394-401.
[4] Liu, Y. J., Xu, D. C., Yu, H. W., G, Y. H. & X, P. L. (2025). Construction and validation of survival prediction model for elderly patients with rectal adenocarcinoma liver metastasis. Modern Medical and Health, 1-12.
[5] Xie, H. Y., Z, L., K, L. W. & L, W. (2025). Prognostic prediction model for head and neck squamous cell carcinoma patients based on SEER database. Chinese Journal of Evidence-Based Medicine, 25(4), 380-386.
[6] Pan, X., Z, S., Mao, N., Yao, F., Guo, Y., Zhou, W. (2025). Development and validation of a prognostic model for predicting survival and immunotherapy benefits in melanoma based on metabolism-relevant genes..Discover oncology, 16(1), 1321.
[7] Zhang, Q., Xu, R., Zhen, W., Bai, X., Li, Z., Zhang, Y., & Li, X. (2025). Competing risk and random survival forest models for predicting survival in post-resection elderly stage I-III colorectal cancer patients. Scientific reports, 15(1), 24269.
[8] Zhang, H., Dong, B., Han, J. & Huang, L. (2025). Interpretable machine learning models for survival prediction in prostate cancer bone metastases. Scientific Reports, 15(1), 24150-24150.
[9] Zhao, G. & Li, D. (2025). TreeEM: Tree-enhanced ensemble model combining with feature selection for cancer subtype classification and survival prediction.Results in Applied Mathematics, 27, 100605-100605.
[10] Wang, X., Peng, F. Y., Zong, D., Qian, P. D. & He, X. (2025). Construction of a survival prediction model for high-grade glioma patients based on LASSO-Cox regression. Journal of Oncology, 31(5), 419-426.