iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** III **Month of publication:** March 2026

DOI: https://doi.org/10.22214/ijraset.2026.78466

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue III Mar 2026- Available at www.ijraset.com

Modeling Survival Outcomes in Eye Cancer Patients Using Advanced Statistical and Machine Learning Approaches II

Rohit Bhuse[1], Shaurya Sawant[2]
Dr. D.Y. Patil Arts, Commerce & Science College, Pimpri, Pune

Abstract: ***Eye cancer remains one of the less frequent but clinically serious malignancies, with survival outcomes that are difficult to predict given the heterogeneity of tumor types and patient profiles. This study examines a dataset of 5,000 patients diagnosed with various ocular malignancies and applies three computational models. Survival Support Vector Machines (SVCR), Random Survival Forests (RSF), and the DeepSurv deep learning model to estimate time-to-event outcomes. Model accuracy was compared using the concordance index, time-dependent AUC, and Brier score. DeepSurv returned the strongest result with a C-index of 0.912, with tumor stage, patient age, and BRAF mutation status emerging as the most influential prognostic variables. These findings point to the practical utility of deep learning in ocular oncology, particularly for stratifying patients by risk prior to treatment decisions.***

Keywords: Eye Cancer, Survival Analysis, DeepSurv, SVCR, Random Survival Forest, Machine Learning, Prognostic Modeling, Ocular Oncology

I. INTRODUCTION

Ocular malignancies including uveal melanoma, retinoblastoma, and intraocular lymphoma are uncommon relative to other cancers, yet their consequences for vision and long-term survival are severe. Each subtype carries a distinct clinical trajectory; uveal melanoma, for instance, has a well-documented tendency for late hepatic metastasis, while retinoblastoma primarily affects young children and demands early aggressive intervention. This diversity in behavior makes survival prediction inherently difficult.For decades, clinicians and researchers relied on Kaplan–Meier curves and Cox regression as the primary tools for understanding survival in these populations. Both methods offer interpretability and have well-understood statistical properties, but they struggle when predictor relationships are non-additive or when the dataset contains a large number of interacting variables both conditions that are common in modern clinical records.The growing availability of linked clinical and genomic datasets, combined with advances in computational resources, has opened the door to more flexible modeling approaches. Machine learning methods, in particular, are well-suited to handle high-dimensional data without requiring the proportionality assumptions that constrain Cox-based models. This study applies SVCR and DeepSurv to an eye cancer cohort with the aim of producing more reliable individual-level survival estimates than traditional approaches can offer.

II. OBJECTIVES

1) To extend traditional survival analysis using advanced machine learning techniques
2) To develop predictive models using SVCR and DeepSurv
3) To identify significant clinical and genetic prognostic factors
4) To compare model performance using standard evaluation metrics.
5) To improve individualized survival prediction and clinical decision-making.

III.PROBLEM STATEMENT

Standard survival models were developed at a time when clinical datasets were small, largely complete, and low-dimensional.
Applied to contemporary oncology data which routinely includes genetic markers, imaging features, treatment history, and comorbidities alongside traditional clinical variables these models run into practical limits. Censored observations are common, predictor interactions are rarely linear, and the number of candidate variables often exceeds what a Cox model can handle without overfitting. The present work attempts to address these gaps by testing models that were specifically designed or adapted for this type of data structure.

IV. METHODOLOGY

A. Data Collection and Preprocessing

Records with missing values in key clinical fields were removed prior to analysis. Categorical variables such as tumor type and treatment category were converted to numeric form using label encoding, and continuous variables were standardized to zero mean and unit variance to prevent scale-dependent bias during model training. The cleaned dataset was split into training and test sets at an 80:20 ratio, with the split stratified by event status to preserve the censoring proportion in both subsets.

B. Models Implemented

Support Vector Machine for Survival (SVCR): Rather than predicting a class label directly, SVCR frames survival analysis as a ranking problem — the model learns to order patients by their relative survival times. A radial basis function kernel was used to accommodate curved, non-separable decision boundaries in the feature space, which is common in clinical data involving genetic markers.

Random Survival Forest (RSF): RSF builds a large collection of survival trees, each trained on a random bootstrap sample of the data with a random subset of features considered at each split. Predictions are aggregated by averaging the cumulative hazard function across all trees. This ensemble strategy reduces overfitting compared to a single tree and handles interactions between variables without any explicit specification from the analyst (Ishwaran et al., 2008).

DeepSurv: DeepSurv replaces the linear component of the Cox model with a multilayer feedforward network, allowing the baseline hazard to remain unspecified while the network learns a flexible mapping from patient covariates to log-hazard. Unlike Cox regression, it places no constraint on the functional form of predictor effects, making it better suited to datasets where survival risk depends on complex combinations of variables (Katzman et al., 2018).

C. Model Evaluation

Three metrics were used. The concordance index (C-index) captures how often the model correctly ranks two randomly chosen patients by survival time a value of 1.0 indicates perfect discrimination. Time-dependent AUC extends this by asking whether discrimination holds at specific follow-up time points rather than across the whole observation window. The Brier score measures calibration, penalizing models that assign poorly calibrated survival probabilities regardless of how well they rank patients.

V. RESULTS

Among the four models tested, DeepSurv achieved the highest C-index at 0.912, followed by RSF at 0.896, SVCR at 0.884, and the Cox baseline at 0.51. The gap between DeepSurv and Cox was roughly four percentage points modest in absolute terms but clinically meaningful when translated into individual risk stratification decisions.

Tumor stage was the strongest single predictor across all models, consistent with its established role in ocular oncology staging systems. Age and BRAF mutation status were the next most influential variables, with both associated with higher hazard. Radiation therapy appeared as a protective factor in the feature importance outputs of RSF, aligning with clinical expectations for certain uveal melanoma presentations.

Patients were divided into low-, medium-, and high-risk groups based on predicted hazard scores. Survival curves for these groups separated clearly, suggesting the model's risk scores carry real prognostic signal rather than just reflecting tumor stage alone. This separation is practically useful: a clinician could, in principle, use the model's output to flag patients who may warrant closer follow-up or earlier escalation of treatment.

VI. CONCLUSION

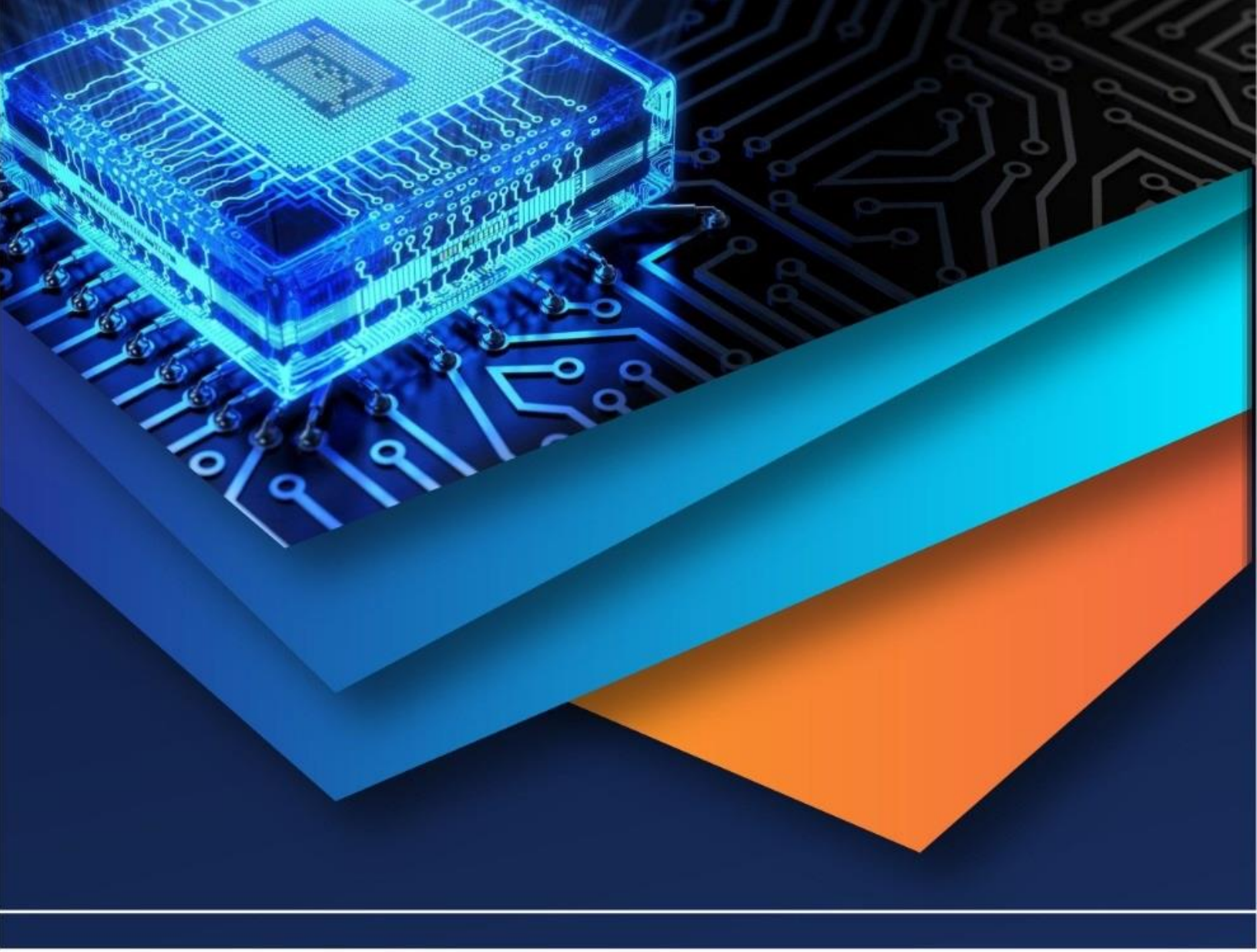
This study tested three machine learning models against a Cox regression baseline for survival prediction in eye cancer patients. DeepSurv outperformed the alternatives on all three evaluation metrics, which is consistent with its theoretical advantage on datasets where predictor effects are non-linear and interactive. RSF also performed well and has the added benefit of producing variable importance estimates that are straightforward to interpret clinically.

That said, this analysis has limitations worth acknowledging. The dataset was sourced from Kaggle and may not reflect the full demographic and clinical diversity of real-world ocular oncology populations. The models were not validated on an external cohort, which is the standard requirement before any clinical deployment. BRAF mutation status, while influential in this dataset, is not consistently recorded across registries, which could limit reproducibility.

Future work should focus on external validation using registry data such as SEER, incorporation of imaging-derived features, and exploration of time-varying covariates to better capture disease progression. Prospective validation in a clinical setting would be the necessary next step before these models could support actual treatment decisions.

REFERENCES

[1] Kaplan EL, Meier P. Nonparametric Estimation from Incomplete Observations. Journal of the American Statistical Association. 1958;53(282):457–481.
[2] Cox DR. Regression Models and Life-Tables. Journal of the Royal Statistical Society Series B. 1972;34(2):187–220.
[3] Ishwaran H, Kogalur UB, Blackstone EH, Lauer MS. Random Survival Forests. The Annals of Applied Statistics. 2008;2(3):841–860.
[4] Harrell FE Jr, Califf RM, Pryor DB, Lee KL, Rosati RA. Evaluating the Yield of Medical Tests. JAMA. 1982;247(18):2543–2546.
[5] Uno H, Cai T, Pencina MJ, D'Agostino RB Sr, Wei LJ. On the C-statistics for Evaluating Overall Adequacy of Risk Prediction Procedures with Censored Survival Data. Statistics in Medicine. 2011;30(10):1105–1117.
[6] Huang Z, Li W, Xu G, et al. A Nomogram to Predict Cancer-Specific Survival of Ocular Melanoma. Frontiers in Medicine. 2025;12:1638733.
[7] Stålhammar G. Long-term Relative Survival in Uveal Melanoma: A Meta-Analysis. Communications Medicine. 2022;2:00082.
[8] He X, Wang L, Chen Y, et al. Epidemiology and Survival Outcomes of Patients with Primary Intraocular Lymphoma. BMC Cancer. 2022;22:6183.
[9] Bilbeisi SC, Brichko T, Chakraborty R, et al. Causes of Death and Survival Analysis for Patients with Retinoblastoma. Frontiers in Medicine. 2023;10:1244308.
[10] Witzenhausen R, Damato B, Charteris D, et al. Survival Benefit of Primary Tumor Treatment in Uveal Melanoma. Cancers (Basel). 2024;16(11):638514.
[11] Singh AD, Turell ME, Topham AK. Uveal melanoma: trends in incidence, treatment, and survival. Ophthalmology. 2011;118(9):1881–1885.
[12] The Global Retinoblastoma Outcome Study Group. Global retinoblastoma survival and globe preservation: analysis of 4,064 patients from 153 countries. Lancet Global Health. 2022;10(6):e807–e815.
[13] Eye cancer patient data set: Kaggle, "Eye Cancer Survival Analysis Dataset," access year 2025.
[14] SEER Program: Surveillance, Epidemiology, and End Results (SEER) Program. National Cancer Institute, US.
[15] Frontiers in Medicine, BMC Cancer, Communications Medicine, and Cancers (Basel) journals for peer-reviewed survival analysis papers.
[16] Official lifelines and scikit-survival documentation for survival modeling methodology.
[17] The Global Retinoblastoma Outcome Study Group, 2022; Demkowicz M, et al., 2023; Zhang N, et al., 2025; Lee J, et al., 2024 for advanced eye cancer analytics and machine learning in survival prediction.
[18] Katzman JL, Shaham U, Cloninger A, Bates J, Jiang T, Kluger Y. DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network. BMC Medical Research Methodology. 2018;18(1):24.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY