

Blood Biomarkers and Machine Learning in Prognosis of Nasopharyngeal Carcinoma

Adel Bajcsi
adel.bajcsi@ubbcluj.ro
Babes-Bolyai University
Cluj-Napoca, Romania

Laura Diana Cernau
laura.cernau@ubbcluj.ro
Babes-Bolyai University
Cluj-Napoca, Romania

Camelia Chira
camelia.chira@ubbcluj.ro
Babes-Bolyai University
Cluj-Napoca, Romania

Laura Diosan
laura.diosan@ubbcluj.ro
Babes-Bolyai University
Cluj-Napoca, Romania

Teodora Ursu
teodora.ursu@stud.ubbcluj.ro
Babes-Bolyai University
Cluj-Napoca, Romania

ABSTRACT

Nasopharyngeal carcinoma (NPC) is an aggressive malignancy characterised by a heterogeneous prognosis despite standardised staging systems. This study examines the prognostic potential of combining systemic inflammatory markers with machine learning models to enhance survival prediction and support personalised clinical decision-making in NPC patients. We conducted a retrospective analysis of patients diagnosed and/or treated for NPC at IOCN(Institutul Oncologic Prof. Dr. Ion Chiricuta, Cluj-Napoca, Romania) and other treatment centres, utilising anonymised data in accordance with ethical guidelines. Predictive models—including Random Forest and Support Vector Machine regression, as well as Decision Tree classification—were developed to estimate overall survival time, predict survival beyond 60 months, and assess the risk of second malignancies based on blood-derived inflammatory markers such as NLR, LMR, PLR, BLR, SIRI, ELR, and SII. Multiple data splitting strategies (random, balanced and stratified) were applied to evaluate model performance across these tasks. Our findings suggest that these systemic inflammatory markers possess significant predictive value for survival outcomes in NPC and that integrating blood-based biomarkers with machine-learning techniques enhances prognostic accuracy. These results suggest a promising role for such approaches in guiding personalised treatment strategies and improving patient outcomes.

KEYWORDS

Nasopharyngeal carcinoma, Blood biomarkers, Machine learning, Survival prediction, Predictive modelling, Clinical decision support

ACM Reference Format:

Adel Bajcsi, Laura Diana Cernau, Camelia Chira, Laura Diosan, and Teodora Ursu. 2025. Blood Biomarkers and Machine Learning in Prognosis of Nasopharyngeal Carcinoma. In . ACM, New York, NY, USA, 2 pages. <https://doi.org/10.1145/nnnnnnn.nnnnnnn>

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
Conference'17, July 2017, Washington, DC, USA

© 2025 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-x-xxxx-xxxx-x/YY/MM
<https://doi.org/10.1145/nnnnnnn.nnnnnnn>

1 INTRODUCTION

Nasopharyngeal carcinoma (NPC) is a highly aggressive malignancy of the head and neck with a complex multifactorial aetiology involving genetic predisposition, environmental factors, and Epstein-Barr virus (EBV) infection. Despite advances in diagnosis and treatment, prognosis remains variable even within the same TNM stage, highlighting the need for more accurate prognostic markers.

Haematological inflammatory markers have emerged as potential predictors of clinical outcomes, but their role in NPC requires further investigation. Previous studies have shown the importance of systemic inflammatory markers in NPC. In particular, in their work, Zeng et al.[2] highlighted that factors like the neutrophil-to-lymphocyte ratio (NLR), platelet-to-lymphocyte ratio (PLR), systemic immune-inflammation index (SII), and systemic inflammation response index (SIRI) were strongly linked to overall survival and disease-free survival in patients with NPC.

Building on these findings, we conducted a retrospective study to evaluate the prognostic potential of a broader set of inflammatory biomarkers by integrating them into machine learning models. Our experiments focused on three tasks: predicting overall survival time, estimating the likelihood of survival beyond 60 months, and identifying patients at risk of developing second malignancies. The classification models, particularly Random Forest, demonstrated promising predictive performance, while regression models faced challenges, likely due to the high variability and range in survival times. Overall, the integration of blood-based biomarkers and machine learning techniques proved to be a valuable approach for enhancing prognostic accuracy and supporting more informed, personalised clinical decision-making.

2 RELATED WORK

In their work, Sun et al. [1] proposed an approach using machine learning algorithms to predict distant metastasis, overall survival, and cancer-specific survival in over 8,000 NPC patients, utilising the SEER database. Some of the best results were obtained during their experiments using the Random Survival Forest model. For their experiments, they primarily used demographic and clinical variables such as gender, tumour grade, and stage as predictive features.

On the other hand, in their research, Zeng et al.[2] adopted a different approach, utilising systemic inflammatory markers as indices. Their findings support the use of combined inflammation markers to enhance prognostic accuracy beyond standard clinical staging. However, their work relied on traditional statistical methods, such as Cox regression, and did not explore machine learning approaches.

Building upon these findings, we incorporated the same inflammation biomarkers into machine learning models to predict survival outcomes and assess the risk of second malignancies. This methodological shift enables the capture of complex, nonlinear relationships within the data and offers improved predictive performance, supporting more informed and personalised clinical decision-making.

3 OBJECTIVES

This study aims to develop predictive models for overall survival and the risk of second malignancies in patients with nasopharyngeal carcinoma, utilising clinical, pathological, and haematological parameters, particularly systemic inflammatory markers.

In our experiments, we aim to evaluate the following tasks:

- Can the classifier accurately predict whether a patient will survive beyond 60 months?
- Can the classifier accurately predict whether a patient will develop a second malignancy?
- Can the regressor reliably estimate the number of months a patient will survive?

4 METHODOLOGY

The datasets included clinical, pathological, and haematological information routinely collected during diagnosis and treatment. However, the preprocessing phase posed significant challenges due to inconsistencies in data format across different patient records. Substantial efforts were made to harmonise the structure of the dataset, enabling a unified analysis pipeline.

In addition to standard clinical and haematological parameters, we computed several systemic inflammatory markers known to have prognostic relevance. These derived biomarkers were calculated as follows:

- Systemic Inflammation Response Index (SIRI): $SIRI = \frac{NEU \times MONO}{LYM}$
- Neutrophil-to-Lymphocyte Ratio (NLR): $NLR = \frac{NEU}{LYM}$
- Systemic Immune-Inflammation Index (SII): $SII = \frac{NEU \times PLT}{LYM}$
- Platelet-to-Lymphocyte Ratio (PLR): $PLR = \frac{PLT}{LYM}$
- Lymphocyte-to-Monocyte Ratio (LMR): $LMR = \frac{LYM}{MONO}$
- Eosinophil-to-Lymphocyte Ratio (ELR): $ELR = \frac{EO}{LYM}$
- Basophil-to-Lymphocyte Ratio (BLR): $BLR = \frac{BASO}{LYM}$

These blood-derived markers were integrated into machine-learning models to evaluate their prognostic value.

To address the study objectives, we developed predictive models using Random Forest and Support Vector Machine (SVM) regression to estimate overall survival time in months. In parallel, Decision Tree classifiers were employed to predict survival beyond 60 months and to assess the risk of developing a second malignancy. Multiple data splitting strategies, including random, balanced, and stratified

sampling, were employed to evaluate model performance across various tasks.

5 RESULTS

Our experiments demonstrate that integrating systemic inflammatory markers into machine-learning models can yield clinically relevant predictions. In the classification tasks, particularly those assessing the risk of second malignancy and survival beyond five years, the Random Forest classifier consistently provided accurate and robust results across various data-splitting strategies. These findings underscore the potential of blood-derived biomarkers in enhancing prognostic models for nasopharyngeal carcinoma.

In contrast, the regression task, aimed at predicting the exact number of months a patient would survive, yielded unsatisfactory outcomes. We believe this may be attributed to the wide variability in survival times within the dataset, which complicates the model's ability to generalise across the full prediction range. To address this limitation, future work will focus on stratifying the dataset, for example, by restricting the analysis to patients with survival durations of less than 60 months in order to reduce prediction variance and improve model performance.

6 CONCLUSIONS

Our findings demonstrate that systemic inflammatory markers—including BLR, SIRI, NLR, LMR, ELR, SII, and PLR—possess significant prognostic value for patients with nasopharyngeal carcinoma (NPC). By integrating these blood-based biomarkers into machine learning models, we were able to enhance the accuracy of survival outcome predictions and stratify risk for second malignancies. This approach offers a promising, noninvasive tool for improving clinical decision-making and guiding personalised treatment strategies.

Further validation through large-scale, multicentre studies is necessary to confirm these results and facilitate the integration of such models into routine clinical workflows for NPC management.

REFERENCES

- [1] Yanbo Sun, Jian Tan, Cheng Li, Di Yu, and Wei Chen. 2024. Creating an interactive database for nasopharyngeal carcinoma management: applying machine learning to evaluate metastasis and survival. *Front. Oncol.* 14 (Oct. 2024), 1456676.
- [2] Xiaojiao Zeng, Guohong Liu, Yunbao Pan, and Yirong Li. 2020. Development and validation of immune inflammation-based index for predicting the clinical outcome in patients with nasopharyngeal carcinoma. *J. Cell. Mol. Med.* 24, 15 (Aug. 2020), 8326–8349.