

Predictive Modeling of Survival in Buccal Mucosa Carcinoma using Decision Tree Techniques

Dipti Yadav, Abhay Khamborkar
<https://orcid.org/0009-0000-1363-7228>
Department of Statistics, Institute of Science,
Rashtrasant Tukadoji Maharaj Nagpur University, Nagpur,
Maharashtra, India.

Abstract - Despite multimodal treatment approaches, buccal mucosa carcinoma, a frequent kind of oral cancer in the Vidarbha area, is linked to low survival rates. We used decision tree categorization, an understandable Machine Learning approach, in this study to forecast 3-year survival results utilizing a clinical dataset of 539 individuals. Treatment approaches used to forecast survival included postoperative radiotherapy (PORT), radiation alone, radiation with chemotherapy, and PORT coupled with chemotherapy.

Clinical stage, histological distinction and treatment mode were found to be independent and statistically significant predictors of survival using multivariate Cox proportional hazard regression analysis.

Additionally, treatment method was found to be the most significant predictor of survival outcomes using the Classification and Regression Tree (CRT) model in SPSS, which produced clear and clinically understandable decision criteria.

All things considered, this study emphasizes of interpretable machine learning models, like decision trees, in bolstering evidence-based clinical decision making and refining treatment approaches for buccal mucosa carcinoma. The results also demonstrate the growing contribution of artificial intelligence and applied mathematics to the advancement of oncology and biomedical research.

Keywords:

Buccal mucosa carcinoma, Decision tree, Survival analysis, SPSS, Cancer treatment

INTRODUCTION

In India, oral cancer is a serious public health issue, particularly in the Vidarbha region. Poor survival rates are associated with buccal mucosa carcinoma, which is frequently diagnosed late and exhibits aggressive behavior. It is still difficult to accurately predict patient survival with the available treatments.

Medical research has seen a sharp increase in the use of machine learning. Decision tree classification, which is well-known for its ease of use and interpretability, can help make predictions based on patient and treatment characteristics by revealing patterns in clinical data that are difficult to find using traditional statistics.

The purpose of this study is to use a decision tree model to forecast, based on various treatment modalities, the 3- year survival outcomes of patients with buccal mucosa carcinoma. By determining which combinations of therapies offer the best chance of long-term survival, the objective is to aid in treatment planning.

MATERIALS AND METHODS

Data Source and Variables

This study was carried out using **real patient records** from **RST Regional Cancer Hospital, Nagpur**. A total of **539 patients** diagnosed with **buccal mucosa carcinoma** were included.

The dataset included the following variables:

- **Demographic information:** Age, gender.
- **Clinical features:** Tumor stage, histopathological differentiation, and lymph node involvement.
- **Treatment modalities:**
 - a. Radiotherapy (RT) alone

- b. Radiotherapy with chemotherapy (RT+CT)
 - c. Post-operative radiotherapy (PORT)
 - d. PORT with chemotherapy (PORT+CT)
- **Outcome variable:** 3-year survival status (alive or dead).

Analytical Method

- A **decision tree classification** approach was applied using the **Classification and Regression Tree (CRT)** algorithm in **SPSS software**.
- The decision tree model was trained to identify the most important predictors of the **3-year survival outcome**.

All data were entered into **SPSS software** for analysis. The study was done with proper **permission** from the hospital authorities, and patient confidentiality was strictly maintained.

Statistical Analysis

The data collected from RST Regional Cancer Hospital were analysed using **SPSS (Version 20)**.

1. Descriptive Statistics

- The characteristics of the patients, including their age, sex, cancer stage, type according to histopathology, and the various treatment groups. Were compiled through the use of frequencies and percentages.
- Mean and standard deviation were calculated for continuous variables (like age).

2. Survival Analysis (Classical Method)

- Kaplan–Meier methodology was applied to estimate Overall Survival (OS).
- Survival curves among RT, RT + CT, PORT, and PORT + CT groups were compared using the log-rank test.
- Results were expressed as **3-year survival probabilities** with p-values (significance at $p < 0.05$).

3. Machine Learning Method (Decision Tree)

- A **Classification and Regression Tree (CRT)** decision tree model was applied to predict **3-year survival status (alive or dead)**.
- Predictor variables included **treatment modality, stage of disease, histopathological differentiation, and lymph node involvement**.
- The **importance of variables** was calculated to see which factors were strongest in predicting survival.

By combining **Kaplan–Meier survival analysis** and **Decision Tree classification**, the study provides both “**traditional statistical evidence** and an **interpretable machine learning model** to support clinical decision-making.

“Ethical approval for the study was obtained from the Institutional Ethics Committee of RST Regional Cancer Hospital, Nagpur.”

Follow -up

All patients were regularly followed up at RST Regional Cancer Hospital, Nagpur, and their visit dates were recorded until December 2024.

Survival outcomes

Overall survival (OS) was calculated from the first treatment date to death or the most recent follow-up among surviving patients.

Statistical analysis

SPSS software (Version 20.0) was used to conduct statistical tests. Chi- square test was used to investigate differences in demographic and clinicopathologic features among individuals with buccal mucosal carcinoma across several treatments, with a

significant level fixed at $P < 0.05$. The Kaplan-Meier approach was used to estimate general survival, and the log-rank test was used to compare survival variations among therapy groups ($P < 0.05$). Additionally, a decision tree classification model was developed using the Classification and Regression Tree (CRT) algorithm in SPSS to identify the most significant predictors of 3-year survival.

RESULTS

A total of 539 patients (426 males and 113 females) were analyzed who have buccal mucosa carcinoma. The surgeries were carried out in 348 (64.6) cases, and in the rest of the cases, patients were deemed unfit to undergo the operation, or patients did not opt to undergo surgery. Additionally, 191 (35.4%) patients were designated for palliative care involving only radiotherapy or a combination of radiotherapy and chemotherapy because of medical reasons that made surgery unadvisable.

Table 1 presents the demographic, clinical, and histopathological characteristics of the patients. Among individuals aged over 50 years, 117 (70.5%) received curative treatment consisting of surgery followed by adjuvant radiotherapy and chemotherapy, while the remaining patients were managed with Palliative care.

Gender: There was no statistically significant difference in the distribution of males and females across the treatment groups ($p = 0.091$).

Age: No statistically significant difference in age categories (≤ 50 vs. > 50 years) was observed between treatment groups ($p = 0.301$).

Histopathology: There was a statistically significant association between histopathological differentiation and treatment groups ($p = 0.002$). Patients with moderately differentiated tumors were more frequently treated with **PORT + CT**, whereas well-differentiated tumors were more common in the **RT** and **RT+CT** groups.

Stage of Tumor: A highly significant difference in tumor stage distribution was observed across treatment groups ($p < 0.001$). All patients receiving **RT** and **RT+CT** were in the **late stage**, whereas a large proportion of patients undergoing **PORT** or **PORT + CT** were in the **early stage**.

Lymph Node Status: No significant association was observed between lymph node involvement and treatment type ($p = 0.369$).

Clinical staging was performed according to the Union for International Cancer Control (UICC) Tumor-Node -Metastasis (TNM) classification system. Based on disease stage, patients were categorized into two groups: Group 1 comprising early-stage disease (i.e., Stage I and II) and Group 2 comprising advanced-stage disease (i.e., Stage III and IV). In the present study, 295 patients (54.73%) were diagnosed at an advanced stage, while regional lymph node involvement was observed in 111 patients (20.59%).

Depending on the disease stage and level of medical fitness, different treatment approaches were used. In our study, 99 (67.3%) in the PORT group and 145 (58.5%) in the PORT+CT group had early stages, while 22 (100%) in the RT group and 122 (100%) in the RT+CT group had advanced stages. Similarly, 4 (18.2%) subjects in the RT group, 32 (26.2%) in the RT+CT group, 29 (19.7%) in the PORT group, and 46 (18.5%) in the PORT+CT group had regional lymph node involvement.

The two factors (type of treatment and clinical stage) had a strong correlation ($P < 0.001$) and no significant correlation ($P = 0.369$).

The histopathological findings became an important prognosis variable to the survival of patients. The survival of patients with poorly differentiated squamous cell carcinoma was worse than the moderately differentiated, or well- differentiated ones, of 117(47.2), and 91(36.7), respectively. Histopathological differentiation and the form of treatment showed a significant correlation ($P = 0.002$).

They applied a multivariate Cox proportional hazards regression analysis to determine independent variables that had effects on overall survival of people with buccal mucosa carcinoma. The model included such variables as clinical stage, histopathological differentiation, lymph node involvement, and the type of treatment that were statistically significant ($\chi^2 = 74.679$, $df=8$, $P < 0.001$).

At the early disease stage, it was associated with a higher probability of survival in its own right with a significantly lower risk of mortality as compared to the late stage disease ($HR = 0.422$, 95% CI: 0.314-0.566 $P < 0.0001$). Also, histopathological differentiation remained a significant prognostic variable (overall $P = 0.010$), whereby well or moderately differentiated tumours had a greater survival when compared to poorly differentiated tumours.

After the regression of other variable, lymph node status could not maintain its independent value ($p = 0.560$). Treatment modality has been indicated as one of the most important predictors of patient survival, showing a strong statistical association (overall $p < 0.0001$). Postoperative radiotherapy alone ($HR = 0.396$, 95% CI: 0.276- 0.569, $P < 0.001$).

DECISION TREE MODEL

The CRT decision tree confirmed **treatment modality** as the strongest predictor of 3-year survival. Patients receiving multimodal therapy (PORT+CT) had higher survival than those receiving single-modality treatments.

Table 1: Distribution of Sociodemographic & Clinical Variables Across Treatment Groups					
Characteristics	RT	RT+CT	PORT	PORT+CT	P value
Gender					
Male	15 (68.2)	102 (83.6)	108 (73.5)	201 (81.0)	0.091
Female	7 (31.8)	20(16.4)	39(26.5)	47(19.0)	
Age					
≤ 50	13 (59.1)	82 (67.2)	97 (66.0)	181 (73.0)	0.301
>50	9(40.9)	40 (32.8)	50 (34.0)	67(27.0)	
Histopathology					
Well differentiated	10 (45.5)	68 (55.7)	64 (43.5)	91(36.7)	0.002
Moderately differentiated	5(22.7)	29 (23.8)	56 (38.1)	117 (47.2)	
Poor Differentiated	0 (00.0)	3(2.5)	1(00.7)	4(1.6)	
Squamous cell carcinoma	7 (31.8)	22 (18.0)	26 (17.7)	36 (14.5)	
Stage of tumor					
Early	00 (0.00)	00 (00.0)	99 (67.3)	145 (58.5)	0.000
Late	22 (100)	122(100.0)	48 (32.7)	103 (41.5)	
Lymph- Node					
Positive	4(18.2)	32 (26.2)	29(19.7)	46(18.5)	0.369
Negative	18 (81.8)	90 (73.8)	118 (80.3)	202 (81.5)	
*Significance exists at the P < 0.05 level by the chi-square method. PORT: Postoperative Radiotherapy					

Table 2: Comparison of Overall Survival Among Treatment Modalities

Treatment groups	Survival time (Mean ± SE)	95% CI	P
Radiotherapy alone	18.91 ± 0.73	17.491 ± 20.33	0.000
Radiotherapy and Chemotherapy	21.55 ± 0.44	20.70 ± 22.42	
PORT	19.04 ± 0.29	18.46 ± 19.62	
PORT+ Chemotherapy	19.92 ± 0.29	1. ± 20.51	

*Statistically Significant at P < 0.05 by Kaplan – Meier (log-rank) Test. PORT: Postoperative Radiotherapy, SE: Standard Error, CI: Confidence Interval

Table 3: Three-Year Survival by Clinical Stage and Treatment Modality

Clinical Stage	Treatment Modality	Total	Survived ≥3 Years, n(%)	Not Survived, n%
Early (n=244)	PORT + CT	145	95 (65.5%)	50 (34.5%)
	PORT	99	55 (55.6%)	44 (44.4%)
Late (n= 295)	RT + CT	122	67 (54.9%)	55 (45.1%)
	RT, PORT, PORT+CT	173	59 (34.1%)	114 (65.9%)
Overall (N= 539)	-----	539	276 (51.2%)	263 (48.8%)

*Statistically significant at P < 0.05 level, PORT: Postoperative Radiotherapy

Multivariate Cox proportional hazards regression analysis was conducted to identify independent predictors of overall survival. The results are summarized in Table 4.

Table 4. Result of Multivariate Cox Proportional Hazards model for Overall Survival

Variable	B	SE	Wald χ^2	HR (Exp(B))	95% CI	p-value
Clinical Stage (Early vs Late)	-0.863	0.15	33.031	0.422	0.314–0.566	< 0.001
Histopathological Differentiation (overall)			11.316			0.01
Well differentiated	-0.528	0.178	8.813	0.59	0.416–0.836	0.003

Moderately differentiated	-0.506	0.181	7.851	0.603	0.423–0.859	0.005
Poorly differentiated	-0.948	0.486	3.803	0.387	0.149–1.005	0.051
Lymph Node Status (Positive vs Negative)	-0.090	0.155	0.34	0.914	0.674–1.238	0.56
Treatment Modality (overall)			39.213			< 0.001
RT + CT vs RT	-0.281	0.293	0.918	0.755	0.425–1.341	0.338
PORT vs RT	-0.926	0.185	25.15	0.396	0.276–0.569	< 0.001
PORT + CT vs RT	0.339	0.15	5.134	1.404	1.047–1.882	0.023

Survival analysis

The average follow-up period of the study population held 17.8 months. By the last follow up, 276 out of 539 were alive. Based on that, the estimated overall survival rate was 3- year, which means that patients were still alive at the final follow up. Therefore the 3 and the survival rate of the whole population was 51.20. The analysis of Kaplan -Meier survival analysis and the log-rank test showed a significant. association between treatment modalities and overall survival. Patients who received postoperative radiotherapy (PORT) combined with adjuvant chemotherapy showed better survival outcomes compared to those treated with other therapeutic approaches (Table 2).

Decision Tree Classification Results:

A classification and regression tree (CRT) model was constructed to predict survival outcomes among patients diagnosed with buccal mucosa carcinoma. The model had a total elevator classification accuracy of 61.4%. The specificity to recognize death was 43.3 where the sensitivity to rightly predict survival was 78.6. The risk assessment carried out by cross-validation yielded a value of 0.386 +/- 0.021 indicating that the risk assessment is satisfactory in the degree of predictive power. Moreover, the survival rate of the combined treatment methods is more effective than that of individual treatment methods. The presence of decision trees in Figure 1 supports this claim.

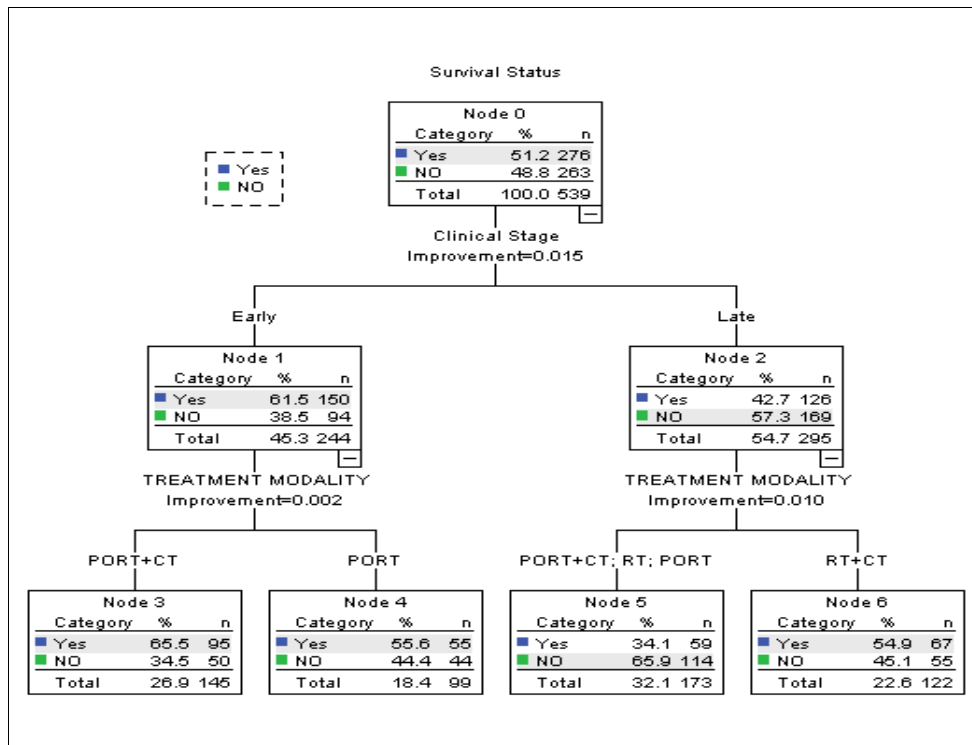


Figure 1: Decision Tree for Predicting Survival Status by Clinical Stage and Treatment Modality

Observed Outcome	Predicted: Survived (Yes)	Predicted: Died (No)	Percent Correct (%)
Survived (Yes)	217	59	78.6
Died (No)	149	114	43.3
Overall Accuracy (%)			61.4

DISCUSSION

We used different treatment options, stages of cancer progression, and tumor types in this study to assess their influence on survival of the patients having buccal mucosa carcinoma. Most of the patients presented in the hospital were in severe stages of the disease which is in line with the previous study carried out in India. This late diagnosis is a reason behind the low survival rates that are commonly witnessed.

We found that the patients receiving multimodality therapy especially those who received surgery along with radiotherapy and chemotherapy survived longer than those receiving radiotherapy only or just by having radiotherapy followed by a combination of chemotherapy and radiotherapy. This has also been observed in other research studies both in India and also international research. Histopathological type of tumor had a great bearing on the survival rates. Tumors with poor differentiation of patients had the least good prognosis compared to those with well differentiated or moderately differentiated tumors who had a greater prognosis of survival.

A new feature in our study presented a decision tree model, which was helpful in emphasizing the fact that the treatment strategy was the most significant variable in predicting the survival rates of 3 years. The use of decision trees to analyze the outcome of cancer survival has shown to be useful with previous studies as well.

Whereas Our finding is positive, some limitations in this study were known. The available treatment options were dependent on the overall health and fitness of the patient instead of being randomly specified which could also have influenced the outcomes. Also, we paid attention to a 3-year survival analysis, but it should be followed by a longer period.

This study highlights that clinical stage, level of histopathological differentiation and the kind of treatment used are vital independent variables that affect survival among patients with buccal carcinoma. It was found that the presence of the disease at an early stage was associated with much improved prognoses, which highlights the importance of early disease diagnosis and management.

The differentiation of tumors remained one of the key prognostic factors, with less differentiated tumours having lower chances of survival, and likely because they are more aggressive genetically. Although the involvement of lymph nodes is associated with the stage of the disease, as well as treatment decisions, and is generally considered as a prognostic variable, the multivariate analysis failed to show its independent effect.

The significant survival benefits of postoperative radiotherapy do support the use of multimodality treatment programs amongst suitably identified patients. The matched results of Cox regression with Kaplan-Meier analyses coupled with the decision tree modelling increases the validity and clinical importance of the study.

To conclude here, our investigation is similar to the findings on carcinoma of the buccal mucosa [1],[2],[3],[4] with a combination of therapies providing better survival chances. Decision tree method opens a new and simple tool that enables a physician to predict outcomes and plan treatment.

CONCLUSION

The current research demonstrates that the treatment plan has significant impact on the prognosis of buccal mucosa cancer. Better survival outcomes were associated with multimodal treatment-specifically, post-operative radiotherapy (PORT) combined with adjuvant chemotherapy-compared to other treatment groups. Using multivariate Cox proportional hazards regression analysis, clinical stage, histopathological differentiation, and treatment technique were discovered to be independent and statistically significant predictors of survival. These findings emphasize how important for carcinoma of the buccal mucosa stage-appropriate, optimized multimodal treatment planning is. Future multicentric and longitudinal research should also include clinical and biological factors in order to increase prognostic accuracy and guide treatment decisions.

ACKNOWLEDGMENT

“I would like to express my gratitude to Rashtra Sant Tukadoji Cancer Hospital & Research Centre, Nagpur, for allowing sample collection. We also appreciate the HBCR Department's assistance with sample collection.”

REFERENCES

- [1] \item [1] P. N. Mishra, D. Kumar, and R. K. Gupta, “Survival analysis of oral cancer patients: A hospital-based study,” *Indian Journal of Cancer*, vol. 55, no. 1, pp. 45–50, 2018, doi: 10.4103/ijc.IJC_132_17.
- [2] \item [2] S. Sharma, A. Madan, and V. S. Mohanti, “Treatment outcomes of carcinoma buccal mucosa: Experience from a regional cancer center in India,” *Journal of Cancer Research and Therapeutics*, vol. 16, no. 2, pp. 291–297, 2020, doi: 10.4103/jcrt.JCRT_497_18.
- [3] \item [3] A. Gupta, P. Kumar, and M. S. Rathi, “Prognostic significance of tumor stage and treatment modality in oral squamous cell carcinoma,” *Asian Pacific Journal of Cancer Prevention*, vol. 21, no. 9, pp. 2603–2609, 2020, doi: 10.31557/APJCP.2020.21.9.2603.
- [4] \item [4] A. Rezaianzadeh, J. Peacock, D. Reidpath, A. Talei, S. V. Hosseini, and D. Mehrabani, “Survival analysis of 1148 women diagnosed with breast cancer in Southern Iran,” *BMC Cancer*, vol. 9, no. 168, pp. 1–11, 2009, doi: 10.1186/1471-2407-9-168.
- [5] \item [5] M. M. Justesen, H. Stampe, K. K. Jakobsen, A. O. Andersen, J. M. Jensen, K. J. Nielsen, A. B. Gothelf, I. Wessel, A. Christensen, C. Grønhoj, and C. von Buchwald, “Impact of tumor subsite on survival outcomes in oral squamous cell carcinoma: A retrospective cohort study from 2000 to 2019,” *Oral Oncology*, vol. 149, no. 106684, pp. 1–10, 2024, doi: 10.1016/j.oraloncology.2024.106684.
- [6] \item [6] H. A. Kumar, R. S. Patil, and N. B. Ingle, “Predictive modeling using decision trees in survival analysis of head and neck cancers,” *Journal of Applied Statistics and Health Sciences*, vol. 3, no. 2, pp. 75–84, 2021, doi: 10.1007/s41045-021-00231-x.
- [7] \item [7] M. P. Fay, “Estimating age conditional probability of developing disease from surveillance data,” *Population Health Metrics*, vol. 2, no. 6, pp. 1–9, Jul. 2004, doi: 10.1186/1478-7954-2-6.
- [8] \item [8] O. Bissinger, A. von den Hoff, E. Maier, K. T. Obermeier, H. Stimmer, A. Kolk, K.-D. Wolff, and C. Götz, “The Value of Surveillance Imaging of Oral Squamous Cell Carcinoma,” *Cancers*, vol. 16, no. 1, p. 207, Jan. 2024, doi: 10.3390/cancers16010207.
- [9] \item [9] A. S. Kapali, N. A. George, E. M. Iype, S. Thomas, and others, “Retrospective Outcome Analysis of Buccal Mucosal and Lower Alveolar Squamous Cell Carcinoma from a High-Volume Tertiary Cancer Centre,” *Indian Journal of Surgical Oncology*, Feb. 2019, doi: 10.1007/s13193-019-00896-8.
- [10] \item [10] J. T. Rich, J. G. Neely, R. C. Paniello, C. C. J. Voelker, B. Nussenbaum, and E. W. Wang, “A Practical Guide to Understanding Kaplan-Meier Curves,” *Otolaryngology–Head and Neck Surgery*, vol. 143, no. 3, pp. 331–336, Sep. 2010, doi: 10.1016/j.otohns.2010.05.007.
- [11] \item [11] A. Krishna, R. K. Singh, S. Singh, P. Verma, U. S. Pal, and S. Tiwari, “Demographic Risk Factors, Affected Anatomical Sites and Clinicopathological Profile for Oral Squamous Cell Carcinoma in a North Indian Population,” *Asian Pacific Journal of Cancer Prevention*, vol. 15, no. 16, pp. 6755–6760, 2014, doi: 10.7314/APJCP.2014.15.16.6755.
- [12] \item [12] P. I. Adamu, P. E. Oguntunde, H. I. Okagbue, and O. O. Agboola, “Statistical Data Analysis of Cancer Incidences in Insurgency Affected States in Nigeria,” *Data in Brief*, vol. 18, pp. 2029–2046, 2018, doi: 10.1016/j.dib.2018.04.135.