



Comparison of Bayesian, Frequentist and Machine learning models for predicting the two-year mortality of patients diagnosed with squamous cell carcinoma of the oral cavity

Sachit Ganapathy^a, K.T. Harichandrakumar^a, Prasanth Penumadu^b, Kadiravan Tamilarasu^c, N. Sreekumaran Nair^{a,*}

^a Department of Biostatistics, Jawaharlal Institute of Post Graduate Medical Education and Research (JIPMER), Puducherry, India

^b Department of Surgical Oncology, Jawaharlal Institute of Post Graduate Medical Education and Research (JIPMER), Puducherry, India

^c Department of Medicine, Jawaharlal Institute of Post Graduate Medical Education and Research (JIPMER), Puducherry, India

ARTICLE INFO

Keywords:

Bayesian
Frequentist
Machine learning
Prediction
Squamous cell carcinoma

ABSTRACT

Background: Statistical models developed in frequentist and Bayesian context along with machine learning algorithms can encompass the multifactorial effect of the prognostic factors in predicting the outcome. This paper is aimed to compare the effect estimates and predictive performance of Bayesian, frequentist and machine learning algorithm in predicting the two-year mortality of patients diagnosed with squamous cell carcinoma (SCC) of oral cavity.

Materials and methods: Logistic regression (LR), Binary Discriminant analysis (BDA), Naïve Bayes (NB), Bayesian regression (BLR), K nearest neighbor (KNN), Artificial neural network (ANN) and Random Forest (RF) models were built. The effect estimate of each prognostic factor was estimated and compared by LR and BLR model. 10-fold cross-validation was performed for internal validation of the models. The predictive performance of the models was assessed and compared.

Results: BLR model had lower and narrower effect estimates in comparison to LR model. Age and smoking are the biggest prognostic risk factors for SCC whereas surgery had the best response amongst the mode of treatment. Random forest had an AUROC of 0.86 (0.82, 0.90) whereas it was estimated to be 0.77 (0.71, 0.82) for both BLR and LR models.

Conclusions: BLR model had better precision in estimating the effect size of prognostic factors and can be an alternative for predicting mortality in patients with SCC of oral cavity. Machine learning classifiers had the best predictive ability as compared to statistical models.

1. Introduction

Head and neck cancer (HNC) accounts for nearly 4.8% of the cancers and a similar percentage of mortality due to cancer around the world.¹ Among the various subsites of HNC, ninety percentage of it is associated with the squamous epithelial cells along the lining of the oral cavity. Oral cavity squamous cell carcinoma (OCSCC) is a huge burden for health care setting and is the sixth most common cancer in the world and more than 50% of the patients have advanced stage of cancer at the time of diagnosis.² Pathological tumor (pT) stage and Pathological node (pN) stage of the cancer along with the age, perineural invasion and margin status of the patients are considered to be important factors affecting the

prognosis of OCSCC.³

Statistical models in health care have been extensively developed to help in medical decision-making.⁴ These models assist in the process of making important decisions to achieve specific clinical outcomes and also in managing resources to be allocated. Prognostic modelling has had immense application in the field of medicine.⁵ They estimate the probability of an outcome of a condition and also explore the relationship of factors affecting this outcome. Unlike other models which incorporate a single explanatory variable and consider other variables as confounders, prognostic models can combine the effect of all variables in the model to predict the outcome.⁵

Logistic Regression (LR) model is the preferred method for modelling

* Corresponding author.

E-mail address: nsknairmanipal@gmail.com (N.S. Nair).

<https://doi.org/10.1016/j.cegh.2022.101145>

Received 23 February 2022; Received in revised form 2 September 2022; Accepted 6 September 2022

Available online 15 September 2022

2213-3984/© 2022 The Authors. Published by Elsevier B.V. on behalf of INDIACLEN. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

the prognosis of disease outcome when the outcome variable is binary. LR model encompasses the effect of predictor variables on the binary dependent variable by linearizing the relationship using a log-link function.⁶ Although the performance of LR as a prognostic model has been good, practically, various assumptions of the models are generally violated.⁷ One of the most important assumptions of LR model is that the predictor variables are independent of one another. This assumption is almost never true in medical research, especially in the context of prognostic model. Logistic regression models have other assumptions of independence of error terms, homoscedasticity among the predictor variables and linearity for continuous variables in logit/probit which are regularly violated. In spite of these assumptions being violated, LR model is widely used. Alternative prognostic models are proposed in literature which uses a wide range of different algorithms to predict the disease outcome. Each of these models under consideration have specific assumptions associated with them. In the application of these models for real world data, these assumptions are hardly satisfied. Under such conditions, we are bound to consider the models which are not sensitive to the assumptions being violated while performing the classification. This study would further provide evidence on the prowess of each prognostic model being considered in the ability to predict the mortality. It will also help us in shaping the prognostic tools to be used by clinicians. Many studies have suggested that the application machine learning algorithms have a better outcome in predictive ability as compared to traditional logistic regression or Cox regression models. Various types of cancers have been studied with different sets of variables being considered for prediction.

The objective of this study was to compare the predictive accuracy of various prognostic models developed in frequentist, Bayesian and machine learning approaches for predicting mortality in patients diagnosed with head and neck cancer. We hypothesize that machine learning algorithms will outperform traditional statistical models in predictive ability with limited number of prognostic variables and average sample size.

2. Methodology

2.1. Dataset and variables

The dataset for this analysis was obtained as a secondary data from a cohort study conducted at a tertiary care hospital in South India after ethical clearance from the Institute Ethics Committee (IEC). The data collection for the primary study was collected retrospectively between the period from 2014 to 2017 and had a minimum of two years of follow up for all patients. There was a total of 437 patients in the study with no missing data for any of the variables considered for the study. Smoking and alcohol consumption have been documented risk factors for the development of oral cavity cancer.⁸ Smokeless/Chewing tobacco has also been documented as one of the most important carcinogenic factors which has a complicated pathway and affects the prognosis of the disease. Patients in the dataset had been subjected to three different modes of treatment namely Surgery, Radical Radiotherapy (RT) and Radical Concurrent Radiotherapy (CCRT). It was also considered whether neo-adjuvant chemotherapy (NACT) was given to the patients. Age and gender have also considered as potential prognostic factors. The age of the patients was categorized as more than or less than 50 years to have played a part in the prognosis of the condition. The subsite of the oral cavity cancer was grouped as Gingivo buccal complex which consisted of buccal mucosa, alveolus and retromolar trigone, tongue and floor of the mouth, hard palate and lip. Each subsite has different degree of aggressive prognosis and vary in their survival performance. The tumor related variables such as the cT staging, cN staging, histological grading and tumour thickness have also been recognised as significant prognostic factors for SCC of oral cavity.⁹ The clinical and pathological tumor and node stage of the patients in the study were categorized based on the AJCC (American Joint Committee on Cancer) cancer staging manual –

8th edition.¹⁰ Pathological tumor (pT) stage and Pathological node (pN) stages were considered for only patients who underwent surgery. Stage of the cancer was characterized for each patient based on the TNM staging. The histological differentiation of the tumor was also considered as studies have shown that poorly differentiated tumor has lower survival rate. The outcome variable considered in the study was death due to any reasons by end of two years of follow-up from the date of diagnosis.

2.2. Models

Predictive models assign subjects into groups based on the predicted probabilities estimated from the effects of a set of predictor variables. Although the principal idea of these models is to predict the outcome variable, they can also estimate the effect of the covariates on the outcome variable. Each of the models has underlying principles based on which they are built and the performance of these models are assessed using metrics on classification accuracy.

Logistic regression is a method used to model the relationship between a set of explanatory variables and categorical outcome variable.⁷ It models the effect of independent variables, linearises the relationship to the outcome variable and estimates the probability of occurrence of each of the possible event in the outcome variable. In binary LR model, the probability of occurrence of the event, $Y = 1$ for given set of covariates is given by,

$$P(Y = 1|X) = \frac{e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k}}{1 + e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k}} \quad (1)$$

where k is the number of independent variables (X). Based on the pre-specified cut-off in this predicted probability, the classification is carried out in the regression model.

Binary discriminant analysis is a procedure that classifies subjects into groups based on the posterior probability of the outcome variable. It denotes group membership of the outcome variable by assuming a joint normal distribution of the explanatory variables, which results in a linear discriminant function.¹¹ The linear discriminant function is presented as,

$$L(x) = \left\{ x - \frac{1}{2}(\bar{x}^{(1)} + \bar{x}^{(2)})^T \right\} S^{-1} \left(\bar{x}^{(1)} - \bar{x}^{(2)} \right) \quad (2)$$

where $\bar{x}^{(i)}$ are the sample mean vectors for $i = 1, 2, \dots, n$ for n different categories and S is the sample covariance matrix.

Naïve Bayes classifier models are probabilistic classifiers that is based on Bayes theorem which uses the properties of conditional independence to compactly represent high-dimensional probability distribution.¹² Bayes theorem in classification is framed as estimating the conditional probability of experiencing the outcome given the data.

$$P(\text{Outcome}|\text{Data}) = \frac{P(\text{Data}|\text{Outcome}) * P(\text{Outcome})}{P(\text{Data})} \quad (3)$$

The joint probability distribution of NB model accounting for the assumption is given by,

$$P(Y, X_1, \dots, X_n) = P(Y) \prod_{i=1}^n P(X_i|Y) \quad (4)$$

where n is the number of predictors in the model, Y is the outcome variable and X_1, \dots, X_n are the set of independent variables.

Bayesian logistic regression model is a statistical method for modelling relationship between explanatory variables and categorical outcome variable with Bayesian inference providing for effective method of regularization.¹³ Using prespecified prior distribution and link function corresponding to the prior, the group assignment is given by,

$$\operatorname{argmax}_w p(w|D) \propto p(w)p(D|w) \quad (5)$$

In the study, we assumed a Beta (1,1) distribution for the outcome variable as a non-informative prior and binomial link function to estimate the posterior distribution of the variables.

K-Nearest neighbor model assigns group to the unknown sample by considering the labels of the 'k' most similar examples in the dataset using Euclidean distance.¹⁴

$$\text{Euclidean distance} = \sqrt{(X_2 - X_1)^2 + (Y_2 - Y_1)^2} \quad (6)$$

The choice of 'k' is arbitrarily chosen for which the accuracy of cross-validation and classification is maximum. In the study, we considered, five nearest neighbors to be optimum for classification.

Artificial Neural network algorithm can be modelled as a set of interconnected nodes which adapt, process and store information between input and output layer.¹⁵ The nodes are connected by edges, which are given weights that are adjusted through iterative training. To obtain the predicted probability, a relative error function¹⁶ of back-propagation defined by,

$$E_i = \sum_k \left[\frac{1}{2} (1 + T_i^k) \log \frac{1 + T_i^k}{1 + O_i^k} + \frac{1}{2} (1 - T_i^k) \log \frac{1 - T_i^k}{1 - O_i^k} \right] \quad (7)$$

where T_i^k is the largest tangent function value for the output unit k and O_i^k is the output value. Outputs of +1 and -1 corresponds to two labels of the outcome variable. In the study, we used a multilayer perceptron with backpropagation given above to obtain the output value.

Random forests are ensemble learning algorithms that combines bagging and random selection of variables which develops into a collection of decision trees by selecting the best node to split on.¹⁷ The classification of the outcome is carried out by measuring the impurity in the data by Gini index given by

$$\text{Gini}(t) = 1 - \sum_{i=1}^N P(C_i|t)^2 \quad (8)$$

where t is a condition, N is the number of categories in the data and C_i is the i th class label of the dataset.

2.3. Validation and accuracy of prognostic models

Validation of any statistical model is to establish whether the model is able to perform well generally for the target population.¹⁸ The entire sample was split into a ratio of 70:30 as training and testing dataset. All the models were built using the training dataset and validated independently in testing dataset. Internal validation was carried out in the training dataset using 10-fold cross-validation technique.¹⁹ The kappa statistic and accuracy of the model was estimated to assess the reliability of the cross-validation. Sensitivity, specificity, F1 score, precision and AUROC curve was estimated to assess the predictive ability of each model. The accuracy measures that were estimated are described in Table 1 and Fig. 1.

3. Results

All the statistical analysis was performed in R Studio Version 1.2.1335²⁰ and SPSS Version 19.²¹ Chi-square test/Fisher's exact test results suggested that subsite, smoking status, stage of cancer, treatment modality, differentiation status and age had significant association with mortality at the end of two years follow up (Table 2). The effect estimates of each factor on the mortality of the patients were estimated using multiple LR and BLR models. It was found that smoking, alcohol, stage of cancer, treatment modality and age were significant prognostic factors from both the models (Table 3). It was observed that although the results were similar in both the models, the interval estimates were narrower and hence the precision was greater in BLR model. It was also

Table 1

Definition of accuracy measures for the model used for predicting mortality in squamous cell carcinoma patients.

Measure	Definition
Accuracy	Proportion of patients who were correctly classified as dead or alive from each model.
Precision	Ability of the model to detect mortality of the patients correctly. This is also termed as positive predictive value (PPV).
F1 score	A metric which combines the PPV and sensitivity of the model by their harmonic mean.
Sensitivity	Proportion of patients whose death status was correctly predicted by the model.
Specificity	Proportion of patients whose alive status was correctly predicted by the model.
Area under the curve	The overall discriminative ability of the model.

observed that there was underestimation of effect estimates in BLR model in comparison to LR model.

The 10-fold cross-validation gave satisfactory results for each of the model considered. The kappa statistic was found to be 0.36 for LR, 0.36 for LDA, 0.35 for BLR, 0.39 for KNN, 0.56 for ANN, 0.38 for NB and 0.52 for RF. The machine learning models of ANN and RF had the best averaged accuracy of cross-validated samples with 0.78 and 0.76 respectively whereas it was comparable accuracy for all the models in testing dataset.

The predictive accuracy of all the models were assessed by various metrics in both training and testing dataset (Table 4). In training dataset, it was observed that RF model had the best AUROC of 0.86 whereas it was 0.85 for ANN model whereas it was estimated to be 0.77 for both LR and BLR models. RF had the best precision of 0.81 whereas it was 0.71 for BLR and LR models. F1 scores were the highest for ANN model with 0.82, whereas it was estimated to be 0.70 for BLR and LR models. The sensitivity was estimated to be 89.5 for ANN and 72.2 for RF whereas BLR and LR models had an estimate of 68.5 each. The specificity was estimated to be the best for RF model at 80.6 whereas BLR and LR model had 67.6 each. In the testing dataset, it was found that BLR, KNN and LR models had the estimated AUROC of 0.76. The precision of RF and BDA was 0.80 which was the highest whereas BLR and LR models had a good precision 0.78 each. F1 scores were estimated to be the best for BLR, BDA and ANN with 0.75 each whereas it was estimated to be 0.74 for LR model. The sensitivity of ANN was estimated to be 0.79 whereas it was estimated to be 71.6 and 70.4 for BLR and LR model respectively. The specificity of BDA and RF was estimated to be 74.6 whereas it was estimated to be 70.9 for both BLR and LR models.

The overall predictive accuracy was the best for machine learning algorithms like RF and ANN models even with a relatively smaller sample size. While comparing the predictive accuracy of the models, we could observe that there was no overfitting due to cross validation, and the metrics were showing lower values for testing dataset as compared to training dataset.

4. Discussion

In the current study, our primary objective was to compare the predictive performance of chosen set of statistical and machine learning models. In that context, we had chosen logistic regression model, Bayesian logistic regression model, binary discriminant function, Naïve Bayes model, Brieman's random forest, K-nearest neighbour algorithm and artificial neural network model for assessing the predictive accuracy for two-year mortality of patients diagnosed with squamous cell carcinoma of oral cavity. We found that machine learning models were significantly better in their predictive ability as compared to statistical models developed both in frequentist and Bayesian approach. Considering F1 score and AUC to be the benchmark metric for assessing the predictive accuracy of models, both RF and ANN models outperformed other statistical models. This indicates that although individual models

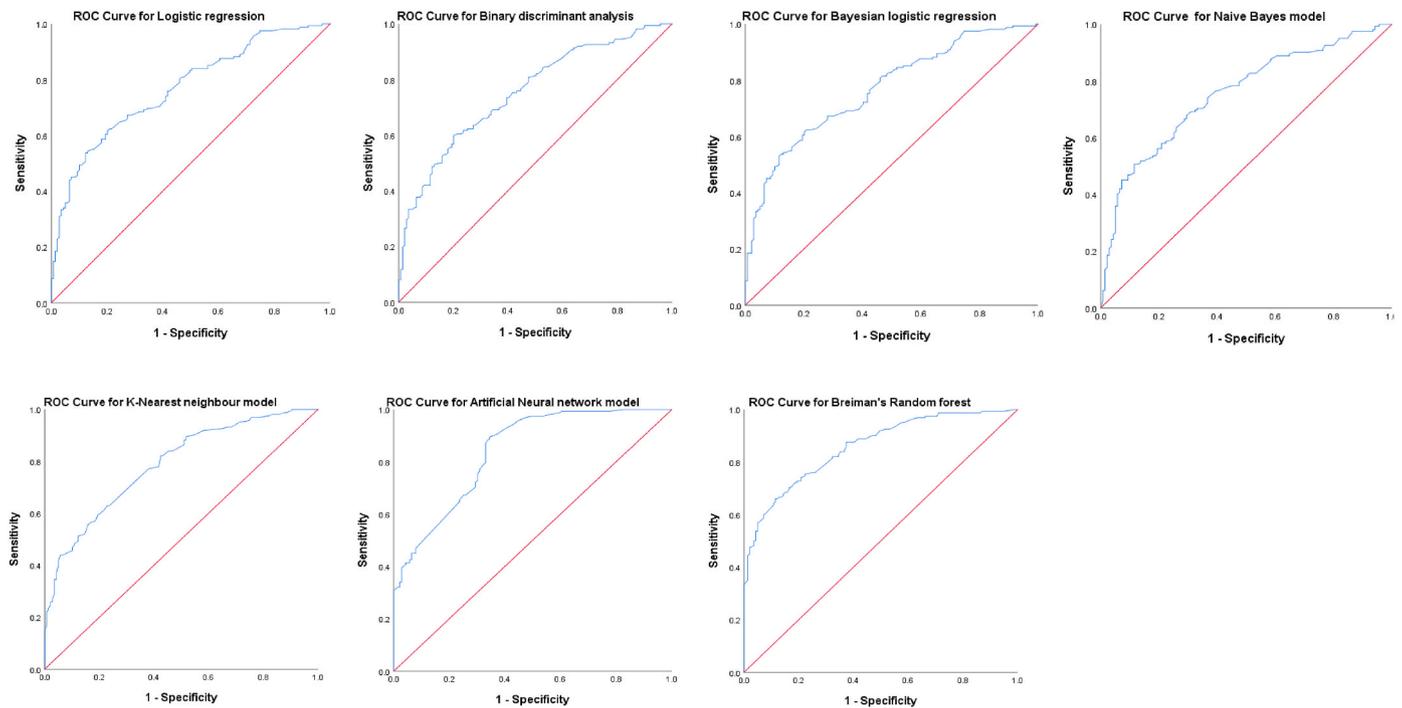


Fig. 1. Receiver operating characteristic (ROC) curves of the prognostic models.

Table 2

Distribution of the prognostic factors according to mortality status at two years follow up among the patients diagnosed with squamous cell carcinoma of the oral cavity.

Variables		Outcome		P-value
		Death (N = 243)	No death (N = 194)	
Gender	Male	156 (64.2)	122 (62.9)	0.777
Subsite ^a	Gingivo Buccal complex	126 (51.9)	119 (61.3)	0.048
	Tongue and floor of mouth	108 (44.4)	67 (34.5)	
	Hard palate	5 (2.1)	1 (0.5)	
	Lip	4 (1.6)	7 (3.6)	
Smoking		89 (36.6)	53 (27.3)	0.039
Alcohol		47 (19.3)	38 (19.6)	0.949
Tobacco		159 (65.4)	114 (58.8)	0.153
Stage	Advanced	200 (82.3)	124 (63.9)	<0.001
	Surgery	110 (45.3)	165 (85.1)	
	Radical CCRT	63 (25.9)	13 (6.7)	
Treatment	Radical RT	70 (28.8)	16 (8.2)	<0.001
	NACT	32 (13.2)	31 (16)	
	Differentiation ^a	172 (70.8)	160 (82.5)	
Differentiation ^a	Well differentiated	68 (28)	30 (15.5)	0.406
	Moderately differentiated	3 (1.2)	4 (2.1)	
	Poorly differentiated			
Age	>50 years	169 (69.5)	116 (59.8)	0.033

NACT – Neoadjuvant chemotherapy, RT – Radiotherapy, CCRT – Radical Chemo-radiotherapy.

^a Fishers exact P-value.

differ in application, the approach to predictive modelling seems to be significantly better in machine learning algorithms.

The overall predictive accuracy of all the models were average considering that the sample size was significantly large. The patients considered in our study were non-metastatic patients diagnosed with squamous cell carcinoma of oral cavity with treatment given with curative intent. As about 90% of oral cavity cancer patients have similar characteristics, the generalizability of the results from the study is

Table 3

Effect estimates of prognostic factors from multiple logistic and Bayesian logistic regression models in predicting the two-year mortality of patients diagnosed with squamous cell carcinoma of oral cavity.

Variables		Logistic regression OR (95% confidence interval)	Bayesian logistic regression OR (95% credible interval)
Gender	Male	1.15 (0.69, 1.95)	1.11 (0.69, 1.78)
Subsite	Tongue & floor of mouth	1.46 (0.89, 2.40)	1.35 (0.86, 2.12)
	Hard palate	2.81 (0.26, 30.62)	2.46 (0.35, 23.86)
	Lip	0.67 (0.17, 2.50)	0.67 (0.19, 2.32)
	Smoking	2.05 (1.01, 4.16)	1.82 (0.96, 3.56)
Alcohol	0.44 (0.22, 0.88)	0.50 (0.27, 0.94)	
Tobacco	0.93 (0.56, 1.56)	0.90 (0.58, 1.52)	
Stage	Advanced	2.84 (1.66, 4.87)	2.46 (1.48, 3.97)
	Radical CCRT	5.61 (2.79, 11.28)	4.48 (2.42, 8.26)
	Radical RT	6.82 (3.58, 13.01)	4.95 (2.93, 9.17)
Treatment	NACT	0.71 (0.38, 1.34)	0.74 (0.42, 1.35)
	Differentiation	1.36 (0.78, 2.39)	1.22 (0.78, 2.10)
Differentiation	Moderately differentiated	0.27 (0.05, 1.44)	0.33 (0.07, 1.50)
	Poorly differentiated		
Age	>50 years	1.91 (1.19, 3.05)	1.65 (1.12, 2.62)

†Reference category: Gender – Female; Subsite – Gingivo Buccal complex; Smoking – No; Alcohol – No; Tobacco – No; Stage – Early; Treatment – Surgery; NACT – No; Differentiation – Well differentiated; Age - < 50 years.

robust. Although significant progress has been made in the prognosis of oral cavity cancer, due to the lack of awareness and aggressive nature of the cancer, the mortality rates are high in tertiary care hospitals in India.²³ Therefore, depending upon the diagnostic stage of the patients, statistical models are supposed to help clinicals take decisions on treatment modality. In this context, predictive models need to be concise with regularly assessed prognostic variables and validated across different datasets. The aim of this study was to compare the approach of predictive model building and to understand the accuracy of models developed with these approaches in oral cavity cancer.

Studies have shown that machine learning models performed better in predictive accuracy as compared to other statistical models.^{24,25} As

Table 4

Performance statistic of the prognostic models in predicting the two-year mortality of patients diagnosed with squamous cell carcinoma of the oral cavity.

Model		Accuracy	F1 score	Precision	Sensitivity	Specificity	AUC ^a (95% confidence interval)
Logistic regression	Training	0.68 (0.63, 0.73)	0.70	0.71	68.5	67.6	0.77 (0.71, 0.82)
	Testing	0.71 (0.62, 0.78)	0.74	0.78	70.4	70.9	0.76 (0.68, 0.85)
Naïve Bayes model	Training	0.69 (0.64, 0.74)	0.70	0.73	67.9	70.5	0.75 (0.69, 0.81)
	Testing	0.70 (0.61, 0.77)	0.73	0.78	69.1	70.9	0.74 (0.68, 0.79)
Binary Discriminant analysis	Training	0.67 (0.61, 0.72)	0.69	0.70	69.1	64.8	0.75 (0.70, 0.81)
	Testing	0.72 (0.64, 0.79)	0.75	0.80	70.4	74.6	0.75 (0.66, 0.83)
Bayesian Logistic regression model	Training	0.68 (0.62, 0.73)	0.70	0.71	68.5	66.9	0.77 (0.71, 0.82)
	Testing	0.71 (0.63, 0.79)	0.75	0.78	71.6	70.9	0.76 (0.68, 0.85)
K-Nearest neighbor model	Training	0.69 (0.64, 0.75)	0.71	0.73	69.1	69.8	0.79 (0.73, 0.84)
	Testing	0.71 (0.62, 0.78)	0.74	0.78	70.4	70.9	0.76 (0.68, 0.85)
Artificial Neural Network	Training	0.78 (0.73, 0.83)	0.82	0.75	89.5	65.5	0.85 (0.80, 0.89)
	Testing	0.69 (0.61, 0.77)	0.75	0.72	79.0	54.6	0.74 (0.65, 0.82)
Brieman's Random Forest	Training	0.76 (0.71, 0.81)	0.76	0.81	72.2	80.6	0.86 (0.82, 0.90)
	Testing	0.71 (0.62, 0.78)	0.73	0.80	67.9	74.6	0.74 (0.66, 0.83)

^a Area under the ROC curve.

the sample size increases, machine learning algorithms performs better than statistical models.²⁶ Therefore, with an increase in the sample size in the present context of the study would only boost their performance. Machine learning algorithms like KNN, RF and ANN gave very good predictive results. Studies based on the prognosis of oral cavity cancer previously comparing the machine learning models have also showed machine learning models outperform conventional methods.²⁷ There are several studies that have pointed out that ANN have performed well in predicting the oral cavity cancer outcome in comparison to logistic regression model.^{24,28} This prompts us to note that predictive models are to be developed specific to disease condition. It also suggests that the inter-relationship between the prognostic factors is important to be considered in the model and there is a need to explore other crucial variable which play a role in the prediction of mortality in these patients.

In our study, we also found that smoking, age, stage of cancer and treatment modality had significant association with the mortality at the end of follow-up at two years. It was noted that the significance of the effect estimates for each of the potential prognostic factors associated with the outcome were similar in both LR and BLR models. But it was consistently observed that the confidence interval from former model was wider as compared to the credible interval from the latter. Although, there are studies which have compared the Bayesian approach to LR model approach in predictive accuracy,²² we observed that BLR model performed on par with LR model in terms of predictive metric.

Implementation of any prognostic model requires rigorous validation, within the available data as well as external validation before bringing it into clinical practice. Validation of models with a larger dataset would provide crucial information about the robustness of the prognostic accuracy. In the current study, although cross-validation was done for internal validation, the developed models were not tested on an independent dataset.²⁹ Calibration plots and discriminative ability of these models for external validation needs to be performed. Our sample size may not have been sufficiently large for the number of variables considered in the model building. The application of the models in a large dataset and external validation would be crucial in understanding the enhanced performance of the above considered models for the prognosis of squamous cell carcinoma of oral cavity patients. The current study aimed at assessing the importance of considering machine learning alternate approaches in developing prognostic model in oral cavity cancer. It is also important to understand that the models need to be considered specifically for particular disease condition and should be modelled with variables that are routinely considered in clinical diagnosis. Online application of these models as web-based tools can have a bedside impact, although that can only be looked at after sufficient external validation of the models.

5. Conclusion

Our study shows that, machine learning algorithms, including ANN, RF, and KNN algorithms outperformed conventional models in the predictive accuracy for the two-year survival of squamous cell carcinoma of oral cavity. It also showed that BLR model estimated the risk factors with more precise interval estimates and had on par predictive accuracy as compared to LR model.

Funding source

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of competing interest

None.

Acknowledgements

1. Dr. S. Pradeep, Department of Surgical Oncology, JIPMER, Puducherry for helping in understanding the practicality of oral cavity cancer and sharing the data.
2. Dr. P. Venkatesan, Scientist-F(Retd.), ICMR; Professor, SRM University for his valuable advices in the conceptual framework in this study.

References

- 1 Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*. 2018;68(6):394–424.
- 2 Jardim JF, Francisco ALN, Gondak R, Damascena A, Kowalski LP. Prognostic impact of perineural invasion and lymphovascular invasion in advanced stage oral squamous cell carcinoma. *Int J Oral Maxillofac Surg*. 2015 Jan;44(1):23–28.
- 3 Zanoni DK, Montero PH, Migliacci JC, Shah JP, Wong RJ, Ganly I, Patel SG. Survival outcomes after treatment of cancer of the oral cavity (1985–2015). *Oral oncology*. 2019 Mar 1;90:115–121.
- 4 Malehi AS, Pourmoghari F, Angali KA. Statistical models for the analysis of skewed healthcare cost data: a simulation study. *Health Economics Review*. 2015 May 27;5(1): 11.
- 5 Vogenberg FR. Predictive and Prognostic Models: Implications for Healthcare Decision-Making in a Modern Recession. 2(6):5.
- 6 Halabi S, Owzar K. The importance of identifying and validating prognostic factors in oncology. *In Seminars in oncology*. 2010 Apr 1;37(2):9–18.
- 7 Hosmer David W, Lemeshow S, Rodney X. *Sturdivant. Applied Logistic Regression*. New York: Wiley; 2000.
- 8 Lam Lisetta, Logan Richard M, Luke Colin, Guy L, Rees. Retrospective study of survival and treatment pattern in a cohort of patients with oral and oropharyngeal tongue cancers from 1987 to 2004. *Oral Oncology*. 2007 Feb;43(2):150–158.
- 9 Arduino PG, Carrozzo M, Chiecchio A, Broccoletti R, Tirone F, Borra E, Bertolusso G, Gandolfo S. Clinical and histopathologic independent prognostic factors in oral squamous cell carcinoma: a retrospective study of 334 cases. *Journal of Oral and maxillofacial surgery*. 2008 Aug 1;66(8):1570–1579.

- 10 Doescher J, Veit JA, Hoffmann TK. The 8th edition of the AJCC cancer staging manual: updates in otorhinolaryngology, head and neck surgery. *Hno*. 2017 Dec 1;65(12):956–961.
- 11 Hand DJ. A comparison of two methods of discriminant analysis applied to binary data. *Biometrics*. 1983 Sep 1:683–694.
- 12 Rish I. An empirical study of the naive Bayes classifier. *InIJCAI 2001 workshop on empirical methods in artificial intelligence*. 2001 Aug 4;3(22):41–46.
- 13 Wakefield J. *Bayesian and Frequentist Regression Methods*. Springer Science & Business Media; 2013 Jan 4.
- 14 Parry RM, Jones W, Stokes TH, Phan JH, Moffitt RA, Fang H, et al. k-Nearest neighbor models for microarray gene expression analysis and clinical outcome prediction. *Pharmacogenomics J*. 2010 Aug;10(4):292–309.
- 15 Uddin S, Khan A, Hossain ME, Moni MA. Comparing different supervised machine learning algorithms for disease prediction. *BMC medical informatics and decision making*. 2019 Dec;19(1):1–6.
- 16 Street WN. A neural network model for prognostic prediction. *INICML*. 1998 Jul 24: 540–546.
- 17 Fawagreh K, Gaber MM, Elyan E. Random forests: from early developments to recent advancements. *Systems Science & Control Engineering: An Open Access Journal*. 2014 Dec 1;2(1):602–609.
- 18 Steyerberg EW, Harrell Jr FE, Borsboom GJ, Eijkemans MJ, Vergouwe Y, Habbema JD. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. *Journal of clinical epidemiology*. 2001 Aug 1;54(8): 774–781.
- 19 Refaeilzadeh P, Tang L, Liu H. Cross-validation. *Encyclopedia of database systems*. 2009 Jan;5:532–538.
- 20 RStudio Team. *RStudio*. Boston, MA: Integrated Development for R. RStudio, Inc.; 2018 (URL <http://www.rstudio.com/>).
- 21 IBM Corp. Released. *IBM SPSS Statistics for Windows, Version 19.0*. Armonk, NY: IBM Corp; 2010.
- 22 Avali VR, Cooper GF, Gopalakrishnan V. Application of Bayesian logistic regression to mining biomedical data. *InAMIA Annual Symposium Proceedings*. 2014;2014:266 (American Medical Informatics Association).
- 23 Borse V, Konwar AN, Buragohain P. Oral cancer diagnosis and perspectives in India. *Sensors International*. 2020 Jan 1;1, 100046.
- 24 Alkhadar H, Macluskey M, White S, Ellis I, Gardner A. Comparison of machine learning algorithms for the prediction of five-year survival in oral squamous cell carcinoma. *Journal of Oral Pathology & Medicine*. 2021 Apr;50(4):378–384.
- 25 Burke HB, Rosen DB, Goodman PH. Comparing the prediction accuracy of artificial neural networks and other statistical models for breast cancer survival. *Advances in neural information processing systems*. 1994;7:1063–1067.
- 26 Enokida T, Fujii S, Takahashi M, Higuchi Y, Nomura S, Wakasugi T, et al. Gene expression profiling to predict recurrence of advanced squamous cell carcinoma of the tongue: discovery and external validation. *Oncotarget*. 2017 Jun 27;8(37): 61786–61799.
- 27 Alabi RO, Elmusrati M, Sawazaki-Calone I, Kowalski LP, Haglund C, Coletta RD, Mäkitie AA, et al. Comparison of supervised machine learning classification techniques in prediction of locoregional recurrences in early oral tongue cancer. *International journal of medical informatics*. 2020 Apr 1;136, 104068.
- 28 Alabi RO, Elmusrati M, Sawazaki-Calone I, Kowalski LP, Haglund C, Coletta RD, Mäkitie AA, Salo T, Leivo I, Almagush A. Machine learning application for prediction of locoregional recurrences in early oral tongue cancer: a Web-based prognostic tool. *Virchows Archiv*. 2019 Oct;475(4):489–497.
- 29 Datema FR, Ferrier MB, Vergouwe Y, Moya A, Molenaar J, Piccirillo JF, et al. Update and external validation of a head and neck cancer prognostic model. *Head & Neck*. 2013;35(9):1232–1237.