Supplementary materials

a	Univariate Cox Regr	ession Analysis		Multivari	iate Cox Regression Analy	sis
Variables	Unadj.HR(95%	CI)	P value	Adj.HR(95%CI)		P value
CDKN2A	····· · · · · · · · · · · · · · · · ·					
High vs. Low	0.703(0.532-0.9	(3) I 🖷 I	0.014	0.698(0.517-0.943)	I • ≡ •I	0.019
Gender Male vs. Female	0.738(0.547-0.9	96) I ·= ·	0.047	0.875(0.623-1.227)	I • 🖬 • I	0.437
>60 vs. <60	1.332(0.997-1.2	8)	0.052	1.154(0.84-1.587)	1 .	0.376
HPV_status		···				
Positive vs. Negative	0.357(0.107-1.1	92) [0.094	0.536(0.149-1.925)	[[····]	0.339
Unknown vs. Negative	1.139(0.734-1.7	67) I····I	0.563	1.167(0.728-1.869)	··■··	0.521
Systematic_therapy	0.994(0.735-1.3	12)	0.967	1 4(0 961-2 04)		0.08
Radiotherapy	0.554(0.755 1.5	-2)	0.907	1.4(0.501 2.04)	• • • •	0.00
YES vs. NO	0.473(0.356-0.6	29) 🛛 🗖 🗧 🕇	< 0.001	0.365(0.262-0.508)	• ■ •	< 0.001
Pathologic_stage			_			
III/IV vs. I/II	1.899(1.248-2.8	89)	0.003	2.252(1.442-3.518)	1 · • · ·	< 0.001
Unknown vs. 1/11 Buninger	1.735(1.015-2.9	54)	0.044	2.046(1.124-3.725)	1	0.019
Unknown vs. NO	1 862(1 275-2 7	19)	0.001	1 529(1 011-2 312)	1	0.044
YES vs. NO	2.169(1.516-3.1	05)	< 0.001	2.039(1.398-2.973)	i i e e i e	< 0.001
Margin status		• -	• •••••		• - •	
Positive vs. Negative	1.901(1.312-2.7	55) I • ■ •	< 0.001	1.664(1.124-2.465)	[·■··]	0.011
Unknown vs. Negative	0.969(0.567-1.6	58)	0.91	1.077(0.575-2.015)	l 🖓 👘 🖓 🚽	0.817
Close vs. Negative	1.307(0.785-2.1	76) I ·· = ·· I	0.304	1.072(0.625-1.839)	1	0.801
Primary_tumor_site	1 115/0 201 1 5		0.525	1.02/0./07 1.405	. 1 .	0.010
Oral Cavity vs. Larynx	1.115(0.791-1.5 0.802(0.48-1.2)		0.535	1.02(0.696 - 1.495) 1.07(0.502 - 1.021)		0.918
Primary diagnosis	x vs. Larynx 0.802(0.48-1.54	I) I	0.4	1.07(0.595-1.951)	1	0.825
Keratinizing vs. NOS	0.827(0.477-1.4	33)	0.498	0.776(0.443-1.36)	1 • • • • • •	0.376
Others vs. NOS	0.368(0.117-1.1	55) • • • • • • • • • • • • • • • • • •	0.087	0.619(0.183-2.089)	[····i.∎····i.]	0.439
Histologic_grade						
G3/G4/GX vs. G1/G2	0.917(0.668-1.2	58) 1 • • •	0.59	0.945(0.668-1.337)	<u> </u>	0.749
) Variables	U	Dl.u	SHD/050/ CI		DC	
Gender	Unadj.SHK(95%CI)	r value	SHK(95%C) for interaction	F I	or interaction
Female	0.889(0.543,1.456)	0.64				
Male	0.679(0.484,0.953)	• • • • 0.025	0.754(0.412,1	.381)	1 • • • • • •	0.36
Age	0.648(0.41.1.022)	0.063				
>60	0.777(0.547,1.105)	0.082	1.214(0.68.2.1	65)	1 • • • • • •	0.51
Systematic_therapy	5111 (010 11,11100)				· - ·	
NO	0.825(0.59,1.154)	• ••• • 0.26				
YES	0.503(0.306,0.826)	↓・・ ■・・・↓ 0.007	0.621(0.346,1	.115)	1 • • • • • •	0.11
Radiotherapy	0.731/0.51.1.047)	0.087				
YES	0.648(0.42.0.999)	0.05	0.879(0.5.1.54	15)	1 • • •	0.65
Pathologic_stage		•				
1/11	0.97(0.459,2.051)	0.94				
III/IV	0.679(0.49,0.942)	0.021	0.719(0.31,1.6	57)	1	0.44
Onknown Perineural invasion	0.623(0.259,1.497)	0.29	0.685(0.217,2	.156)	1	0.52
NO	1.067(0.599,1.899)	0.83				
Unknown	0.684(0.423,1.106)	0.12	0.641(0.307,1	.341)	· · · = · · <mark> </mark> · 	0.24
YES	0.574(0.375,0.88)	[··■··] 0.011	0.519(0.251,1	.074)	I · · · ■ · · · I	0.077
Margin_status	0.76(0.54.1.060)	0.12				
Positive	0.605(0.307.1.193)	0.12	0.846(0.4]4.1	.73)	1	0.65
Unknown	0.457(0.169,1.236)	0.12	0.458(0.147,1	.428)	·····•	0.18
Close	0.716(0.267,1.921)	••••••••••••••••••••••••••••••••••••••	1.145(0.424,3	.092)	[····]	0.79
Primary_tumor_site	0.007/0.5551.700					
Larynx Oral Cavity	0.997(0.556,1.789)	0.99	0 713(0 37 1 3	(71)	1	0.31
Oropharynx/Hypopharynx	0.335(0.135,0.83)	0.002	0.329(0.12.0 9)) I.		0.03
Histologic_grade	•••••			·		
G1/G2	0.884(0.638,1.226)	I · ■··I 0.46				
G1/G2 G3/G4/GX	0.884(0.638,1.226) 0.403(0.233,0.698)	Ⅰ · ■ · Ⅰ 0.46 0.001	0.451(0.235,0	.868)	[····	0.017

Fig-1 Comparative Analysis of CDKN2A and Prognostic Factors

In univariate analysis, high expression of CDKN2A was a protective factor for overall survival (OS) with statistical significance (HR=0.703, 95% CI 0.532-0.93, P=0.014). In multivariate analysis, after adjustment for multiple factors, high expression of CDKN2A (HR=0.698, 95% CI 0.517-0.943, P=0.019) remained a significant protective factor for OS. In subgroup analysis, for the female subgroup, elevated CDKN2A was a protective factor for OS (HR=0.889, 95% CI 0.543-1.456, P=0.64), although this was not statistically significant. Conversely, in the male subgroup, an increase in CDKN2A was also a protective factor for OS (HR=0.679, 95% CI 0.484-0.953, P=0.025), which was statistically significant. The interaction test yielded a P-value of 0.36, indicating no significant interaction between CDKN2A and the different gender subgroups, suggesting that the effect of CDKN2A on OS is consistent across gender subgroups.

Description about mRMR_RFE feature



Fig-2 The mRMR method selected the top 20 features, and then RFE was used to select four features.

The mRMR algorithm (Maximum Relevance, Minimum Redundancy)^[1] is used to select features, considering not only the correlation between features and the target variable to be predicted, but also the correlation between features themselves. The metric used is Mutual Information. For the mRMR method, the relevance of a feature subset to the class is calculated by the mean of the information gain of each feature with respect to the class, while the redundancy between features is measured by the sum of the mutual information between each feature and all other features in the subset, divided by the square of the number of features in the subset.RFE (Recursive Feature Elimination)^[2] is a feature selection method that ranks the predictor variables before modeling and sequentially eliminates the least important ones. Its goal is to find the subset of predictors that can generate an accurate model. The model is repeatedly trained, with the n least important features being eliminated after each training. Then, the new feature subset is re-trained, and the feature importance is reassessed. The n least important features are again eliminated, until an optimal feature subset is obtained.

Description about built GBM Model



Fig-3 The importance of the selected features in the GBM algorithm.

Gradient Boosting Machine (GBM) algorithm^[3] uses a set of weak classifiers (usually decision trees) to train new weak classifiers based on the negative gradient information of the current model's loss function. Then, the trained weak classifiers are combined with the existing model in an additive manner to build a predictive model. By using GBM to model the selected pathological features, it can predict gene expression.

Table-1 Comparison of parameters between training set and validation set

	ACC	SEN	SPE	Brier score
Training set	0.775	0.762	0.789	0.196
Validation set	0.650	0.690	0.605	0.223

PS: Accuracy (ACC), Specificity (SPE), Sensitivity (SEN)

References

- Zhang Y, Ding C, Li T. Gene selection algorithm by combining reliefF and mRMR. BMC GENOMICS. 2008.
 9: 1-10.
- [2] Granitto PM, Furlanello C, Biasioli F, Gasperi F. Recursive feature elimination with random forest for PTR-MS analysis of agroindustrial products. CHEMOMETRICS AND INTELLIGENT LABORATORY SYSTEMS. 2006. 83(2): 83-90.
- [3] Ayyadevara VK, Ayyadevara VK. Gradient boosting machine. Pro machine learning algorithms: A hands-on approach to implementing algorithms in python and R. 2018 : 117-134.