SUPPLEMENTARY DATA

Machine learning for predicting the survival in osteosarcoma patients: Analysis based on American and Hebei Province cohort

Table S1. The 3-year and 5-year CSS of osteosarcoma in different datasets

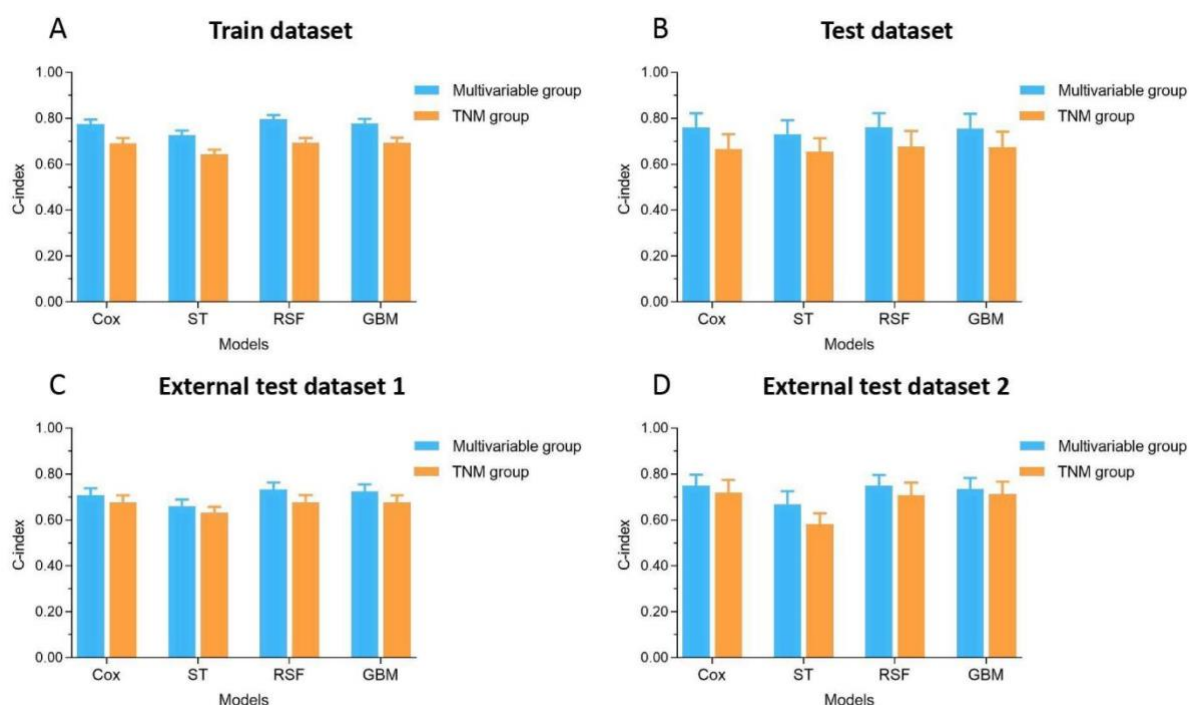| CSS	Development dataset	External test dataset 1	External test dataset 2
3-year CSS	72.71%	70.25%	55.25%
5-year CSS	65.92%	63.46%	52.49%

CSS: Cancer-specific survival.

Figure S1. Histogram of C-indices for various models for predicting 5-year CSS of osteosarcoma in multivariate group and TNM group. (A) Histogram of C-indices for train dataset; (B) Histogram of C-indices for test dataset; (C) Histogram of C-indices for external test dataset 1; (D) Histogram of C-indices for external test 2. C-index: Concordance index; CSS: Cancer-specific survival; TNM: Tumor, node, metastasis.

Table S2. C-index for various models for predicting 5-year CSS of osteosarcoma in multivariate group and TNM group

Models	Train dataset Mean (95% CI)	Test dataset Mean (95% CI)	External test dataset 1 Mean (95% CI)	External test dataset 2 Mean (95% CI)
Multivariable group				
Cox	0.77 (0.73-0.81)	0.76 (0.64-0.88)	0.71 (0.65-0.77)	0.75 (0.65-0.85)
ST	0.73 (0.68-0.77)	0.73 (0.61-0.85)	0.66 (0.60-0.72)	0.67 (0.55-0.79)
RSF	0.80 (0.76-0.83)	0.76 (0.64-0.88)	0.73 (0.67-0.79)	0.75 (0.65-0.85)
GBM	0.78 (0.74-0.82)	0.76 (0.63-0.88)	0.72 (0.67-0.78)	0.73 (0.63-0.83)
TNM group				
Cox	0.69 (0.65-0.73)	0.67 (0.54-0.79)	0.68 (0.62-0.74)	0.72 (0.62-0.82)
ST	0.64 (0.60-0.68)	0.65 (0.54-0.77)	0.63 (0.57-0.69)	0.58 (0.48-0.68)
RSF	0.69 (0.65-0.73)	0.68 (0.54-0.81)	0.68 (0.62-0.74)	0.71 (0.61-0.81)
GBM	0.69 (0.65-0.73)	0.67 (0.54-0.81)	0.68 (0.62-0.74)	0.71 (0.59-0.83)

C-index: Concordance index. CSS: Cancer-specific survival; ST: Survival tree; RSF: Random survival forest; GBM: Gradient boosting machine; TNM: Tumor, node, metastasis.
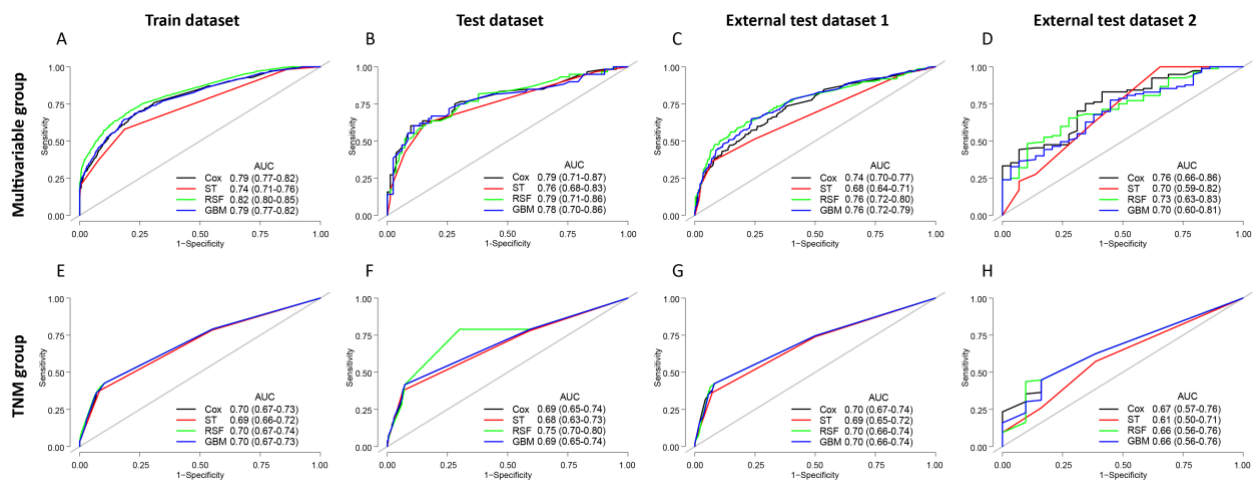

Figure S2. ROC curve for various models for predicting 5-year CSS of osteosarcoma in multivariate group and TNM group. (A) ROC curve for train dataset in multivariable group; (B) ROC for test dataset in multivariable group; (C) ROC curve for external test dataset 1 in multivariable group; (D) ROC curve for external test dataset 2 in multivariable group; (E) ROC curve for train dataset in TNM group; (F) ROC curve for test dataset in TNM group; (G) ROC curve for external test dataset 1 in TNM group; (H) ROC curve for external test dataset 2 in TNM group. ROC: Receiver operating characteristic curve; CSS: Cancer-specific survival; ST: Survival tree; RSF: Random survival forest; GBM: Gradient boosting machine; TNM: Tumor, node, metastasis.
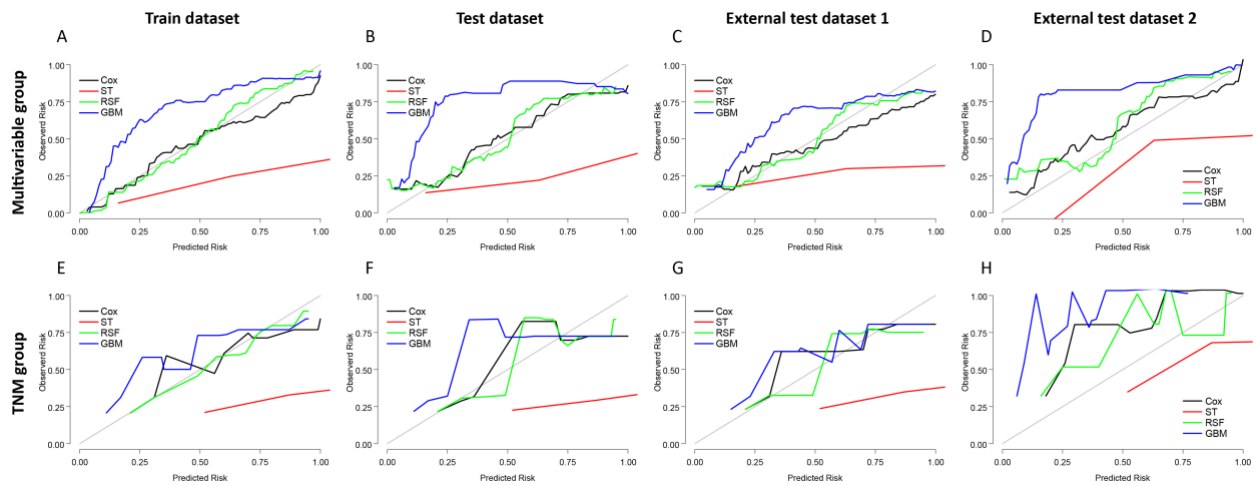
Figure S3. Calibration curves for various models for predicting 5-year CSS of osteosarcoma in multivariate group and TNM group. (A) Calibration curves for train dataset in multivariable group; (B) Calibration curves for test dataset in multivariable group; (C) Calibration curves for external test dataset 1 in multivariable group; (D) Calibration curves for external test dataset 2 in multivariable group; (E) Calibration curves for train dataset in TNM group; (F) Calibration curves for test dataset in TNM group; (G) Calibration curves for external test dataset 1 in TNM group; (H) Calibration curves for external test dataset 2 in TNM group. CSS: Cancer-specific survival; ST: Survival tree; RSF: Random survival forest; GBM: Gradient boosting machine; TNM: Tumor, node, metastasis.

Table S3. C-indices for various models for predicting 3- and 5-year CSS of osteosarcoma in same variables group

Models	Train dataset Mean (95% CI)	Test dataset Mean (95% CI)	External test dataset 1 Mean (95% CI)	External test dataset 2 Mean (95% CI)
Three-year survival cohort				
Cox	0.80 (0.75-0.84)	0.78 (0.65-0.91)	0.72 (0.65-0.78)	0.79 (0.71-0.87)
ST	0.77 (0.73-0.81)	0.76 (0.63-0.89)	0.70 (0.64-0.77)	0.66 (0.54-0.78)
RSF	0.82 (0.77-0.86)	0.77 (0.63-0.90)	0.74 (0.68-0.81)	0.75 (0.65-0.85)
GBM	0.80 (0.76-0.84)	0.78 (0.65-0.91)	0.74 (0.67-0.80)	0.76 (0.66-0.86)
Five-year survival cohort				
Cox	0.77 (0.73-0.81)	0.76 (0.64-0.88)	0.71 (0.65-0.77)	0.79 (-0.33-1.91)
ST	0.73 (0.69-0.77)	0.73 (0.61-0.85)	0.66 (0.60-0.72)	0.73 (-0.15-1.61)
RSF	0.79 (0.75-0.83)	0.76 (0.64-0.88)	0.73 (0.67-0.79)	0.75 (-0.21-1.71)
GBM	0.78 (0.74-0.82)	0.76 (0.64-0.88)	0.73 (0.67-0.79)	0.76 (-0.26-1.78)

C-index: Concordance index. CSS: Cancer-specific survival; ST: Survival tree; RSF: Random survival forest; GBM: Gradient boosting machine; TNM: Tumor, node, metastasis.
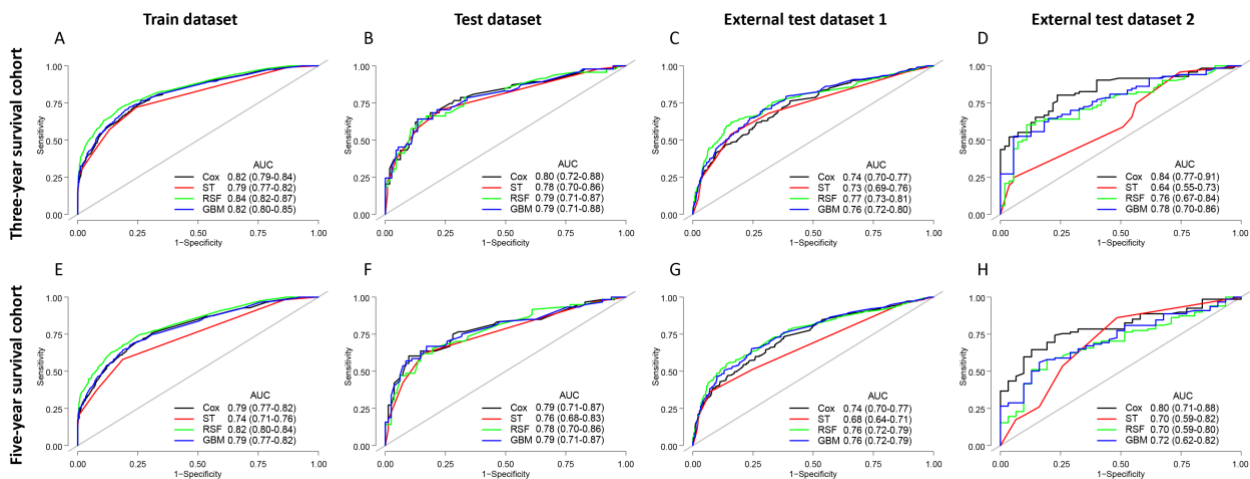

Figure S4. ROC curves for various models for predicting 3-year and 5-year CSS of osteosarcoma in same variables group. (A) ROC curves for train dataset in 3-year survival cohort; (B) ROC curves for test dataset in 3-year survival cohort; (C) ROC curves for external test dataset 1 in 3-year survival cohort; (D) ROC curves for external test dataset 2 in 3-year survival cohort; (E) ROC curves for train dataset in 5-year survival cohort; (F) ROC curves for test dataset in 5-year survival cohort; (G) ROC curves for external test dataset 1 in 5-year survival cohort; (H) ROC curves for external test dataset 2 in 5-year survival cohort. ROC: Receiver operating characteristic curve; CSS: Cancer-specific survival; ST: Survival tree; RSF: Random survival forest; GBM: Gradient boosting machine; TNM: Tumor, node, metastasis.
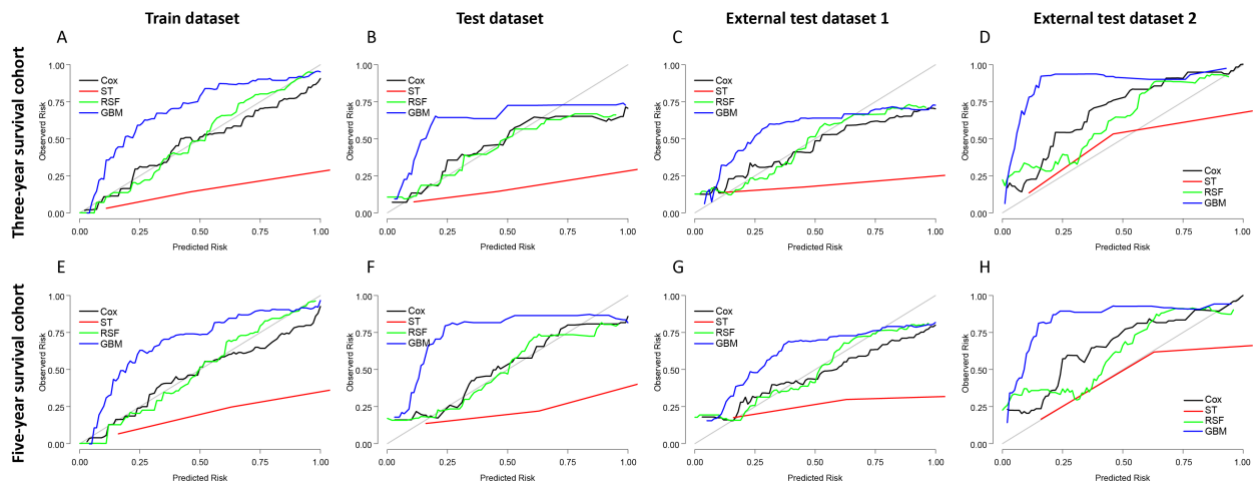
Figure S5. Calibration curves for various models for predicting 3-year and 5-year CSS of osteosarcoma in same variables group. (A) Calibration curves for train dataset in 3-year survival cohort; (B) Calibration curves for test dataset in 3-year survival cohort; (C) Calibration curves for external test dataset 1 in 3-year survival cohort; (D) Calibration curves for external test dataset 2 in 3-year survival cohort; (E) Calibration curves for train dataset in 5-year survival cohort; (F) Calibration curves for test dataset in 5-year survival cohort; (G) Calibration curves for external test dataset 1 in 5-year survival cohort; (H) Calibration curves for external test dataset 2 in 5-year survival cohort. ROC: Receiver operating characteristic curve; CSS: Cancer-specific survival; ST: Survival tree; RSF: Random survival forest; GBM: Gradient boosting machine; TNM: Tumor, node, metastasis.