Terahertz Time-Domain Spectroscopy of Glioma Patient Blood Plasma: Diagnosis and Treatment

Below, brief formulas mentioned in the manuscript are presented for convenience.

True positive (TP): glioma patients correctly diagnosed. False positive (FP): Healthy people incorrectly identified as patients with glioma. True negative (TN): Healthy people correctly identified as healthy. False negative (FN): Glioma patients incorrectly identified as healthy.

TPR (True Positive Rate) / Recall /Sensitivity

FPR (false positive rate) = 1- Specificity

Sensitivity = TP / (TP+FN)

Specificity = NT / (TN+FP)

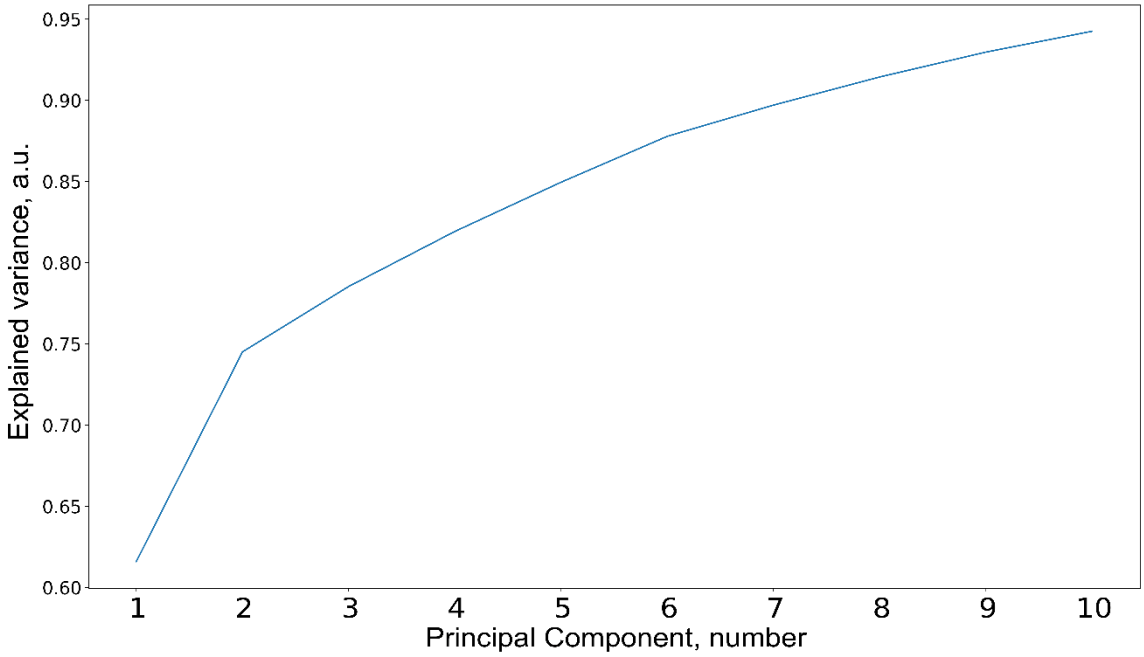
Accuracy = (TP+TN)/(TP+FN+TN+FN)

Precision = TP/(TP+FP)

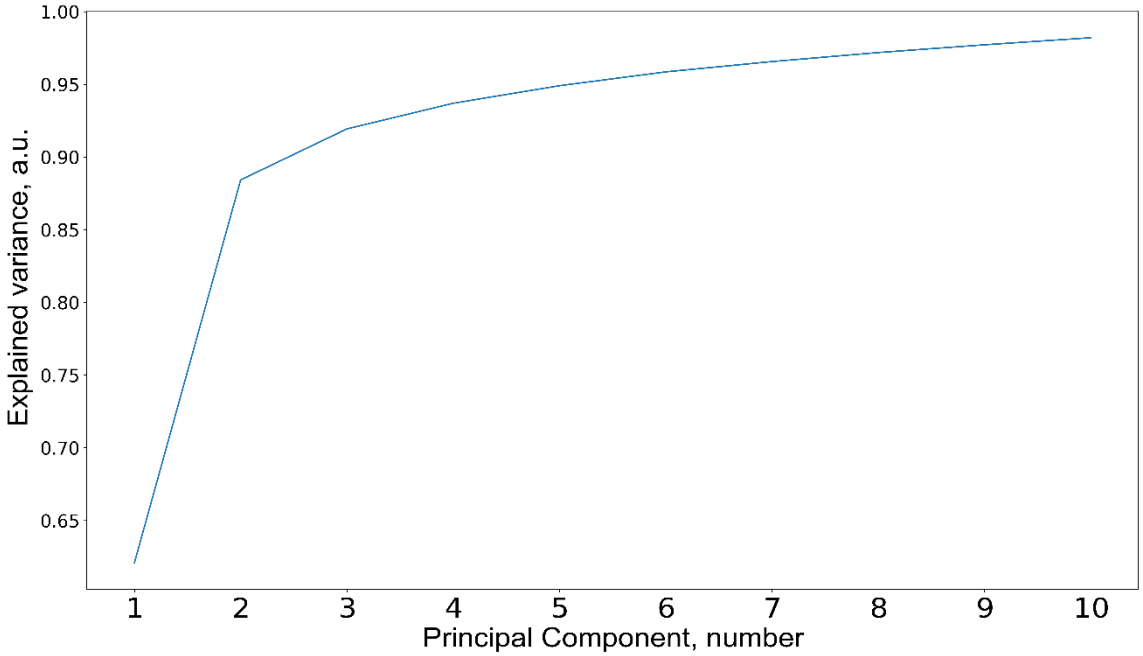
The Receiver Operating Characteristics (ROC) curve is plotted with TPR against the FPR where TPR is on the y-axis and FPR is on the x-axis.

ROC-AUC analysis is a commonly used technique in ML. ROC is a probability curve and AUC represents the degree or measure of separability as an area under the ROC. AUC value varies between 0 and 1, any of which is considered to be the best. AUC zero value means that class labels need to be swapped (model is reciprocating the classes). ROC (AUC = 0.5) is similar to chance curve and tells that classifier perform

random guessing which is not satisfactory. In general, an AUC of 0.5 suggests no discrimination (i.e., ability to diagnose patients with and without the disease or condition based on the test), 0.7 to 0.8 is considered acceptable, 0.8 to 0.9 is considered excellent, and more than 0.9 is considered outstanding.

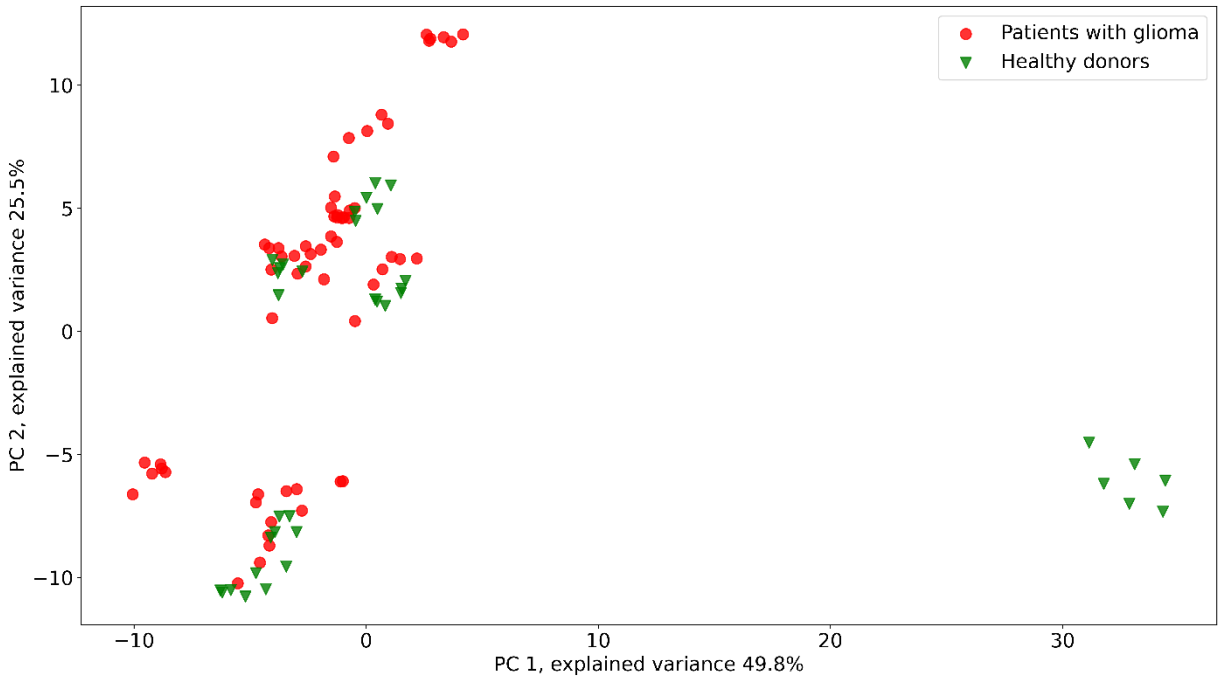


(a)

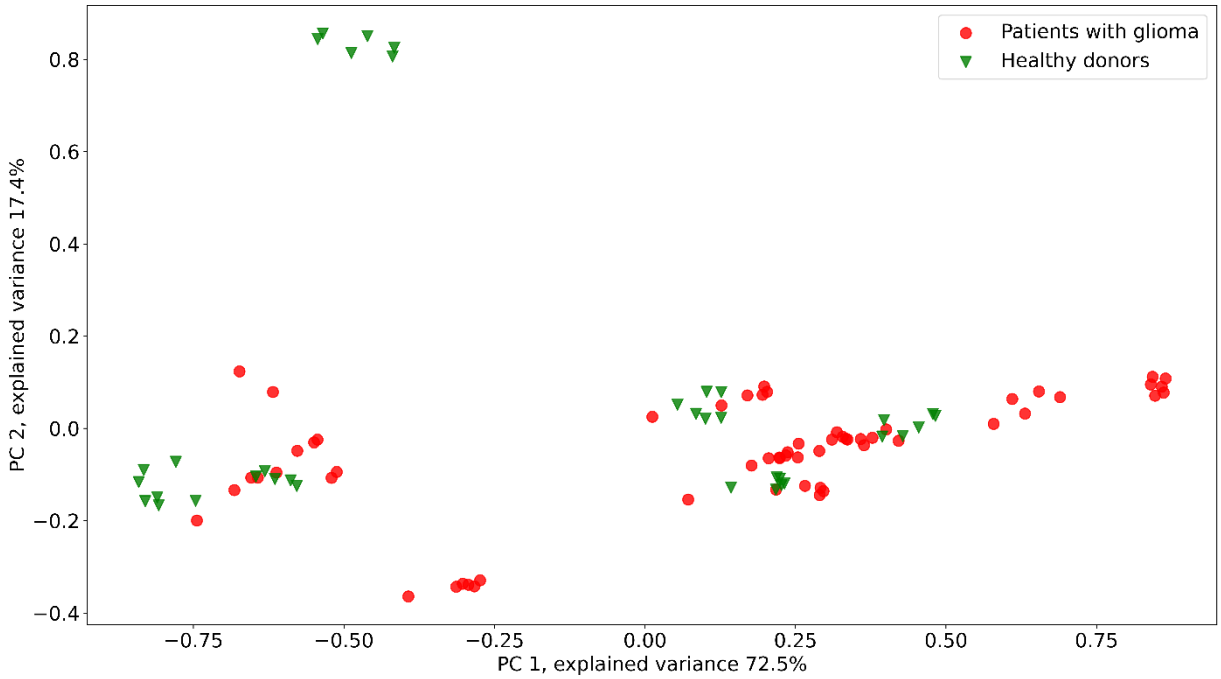


(b)

Figure S1. Graph of the cumulative explained variance to find the maximum allowed number of PCs for (a) blood plasma of glioma patients and healthy donors (case 1); (b) blood plasma of glioma patients and patients with skull craniectomy defects (case 2).

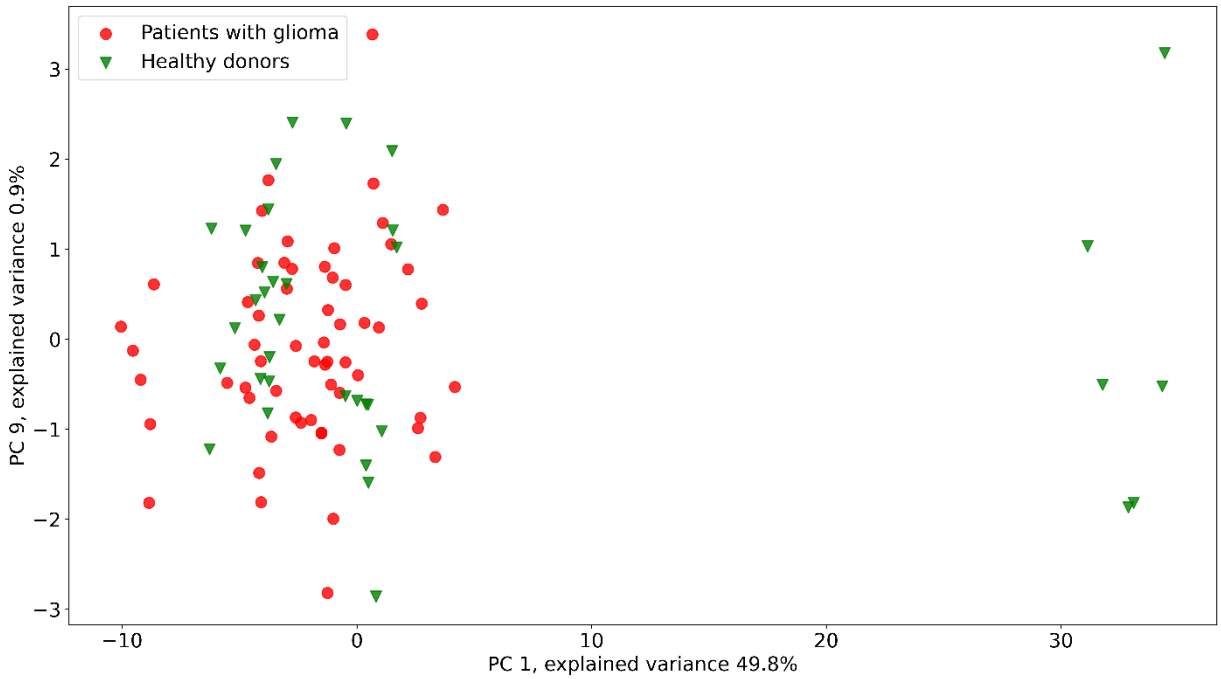


(a)

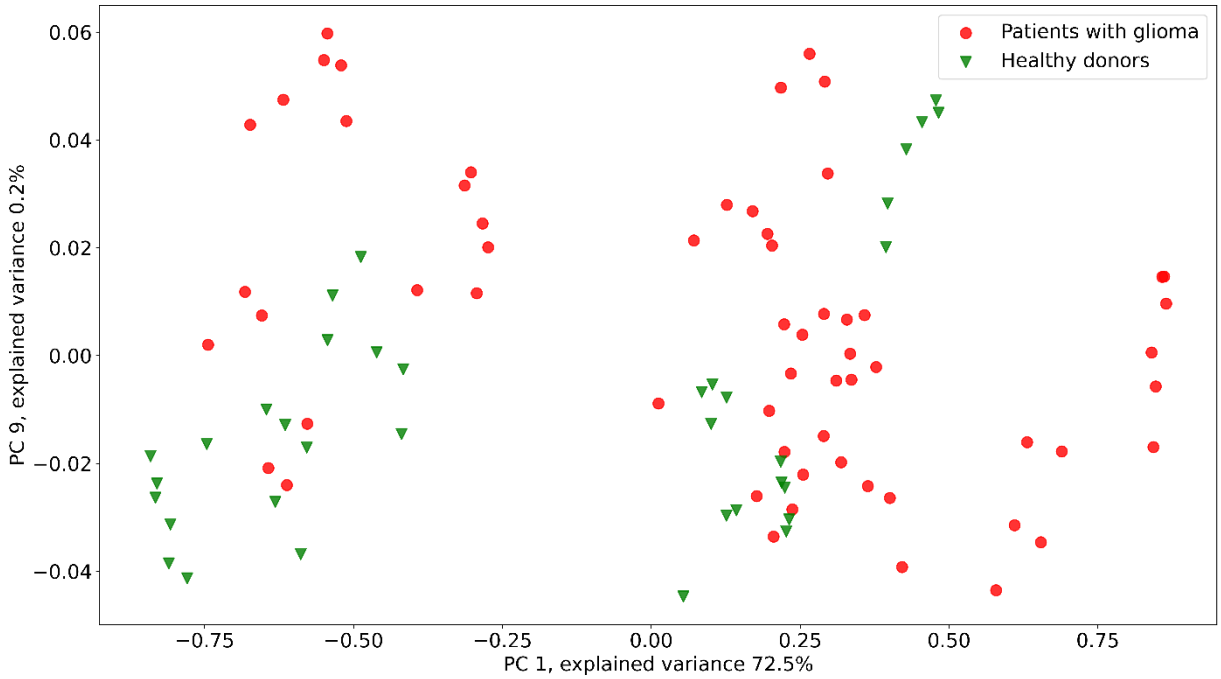


(b)

Figure S2. PCA score plot with data standardization (a) and without (b) for glioma patients and healthy donors, for PC1 and PC2.

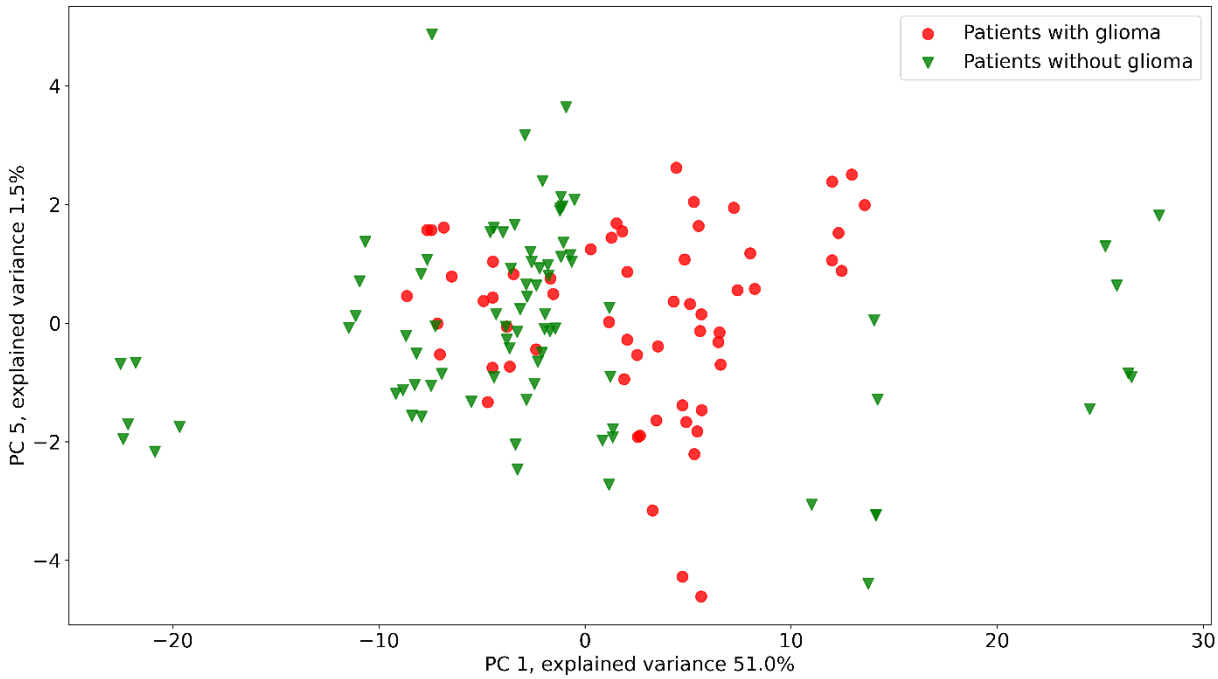


(a)



(b)

Figure S3. PCA score plot with data standardization (a) and without (b) for glioma patients and healthy donors, for PC1 and PC9.



(a)

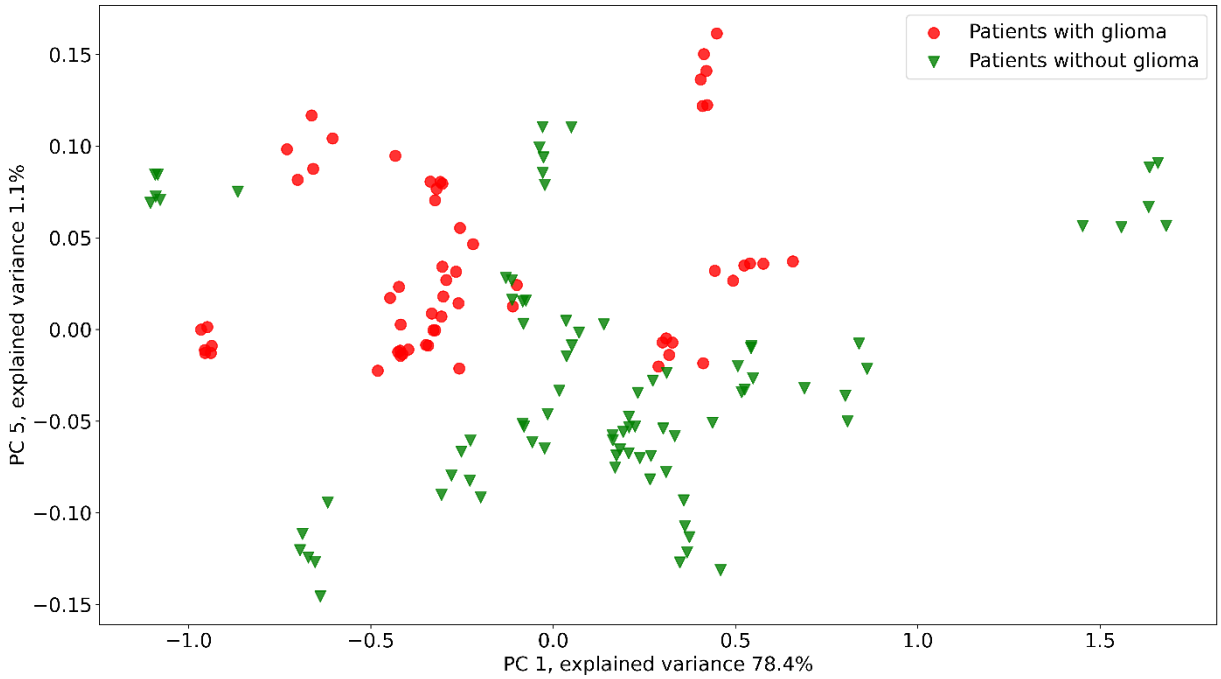
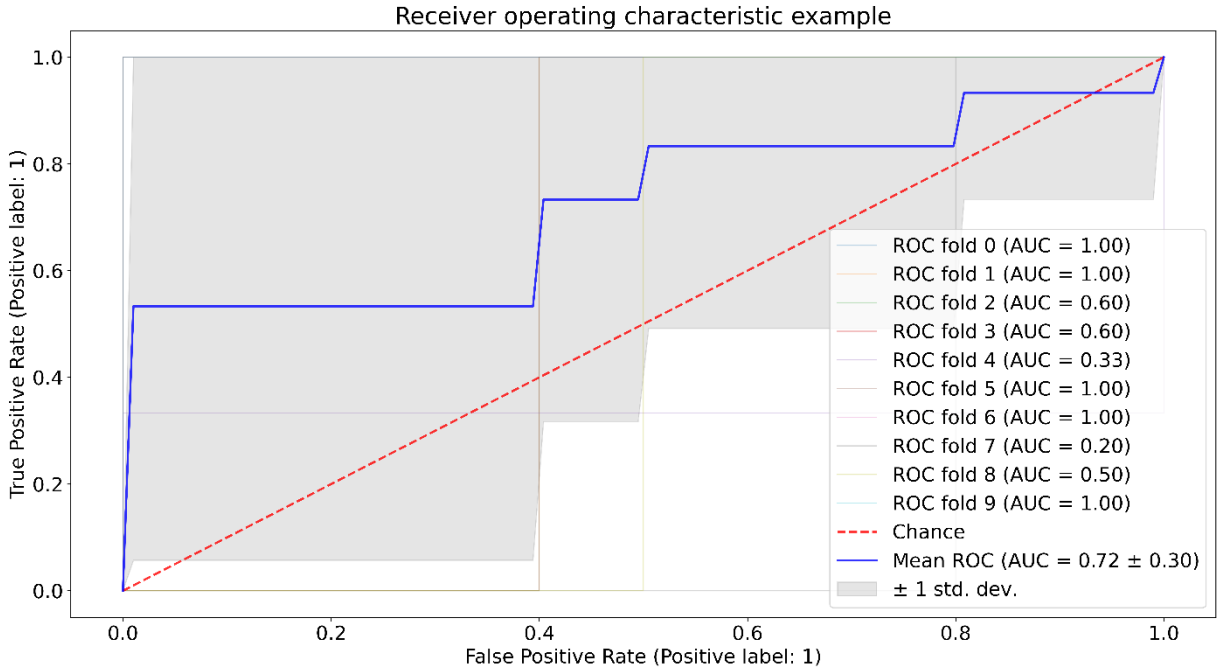
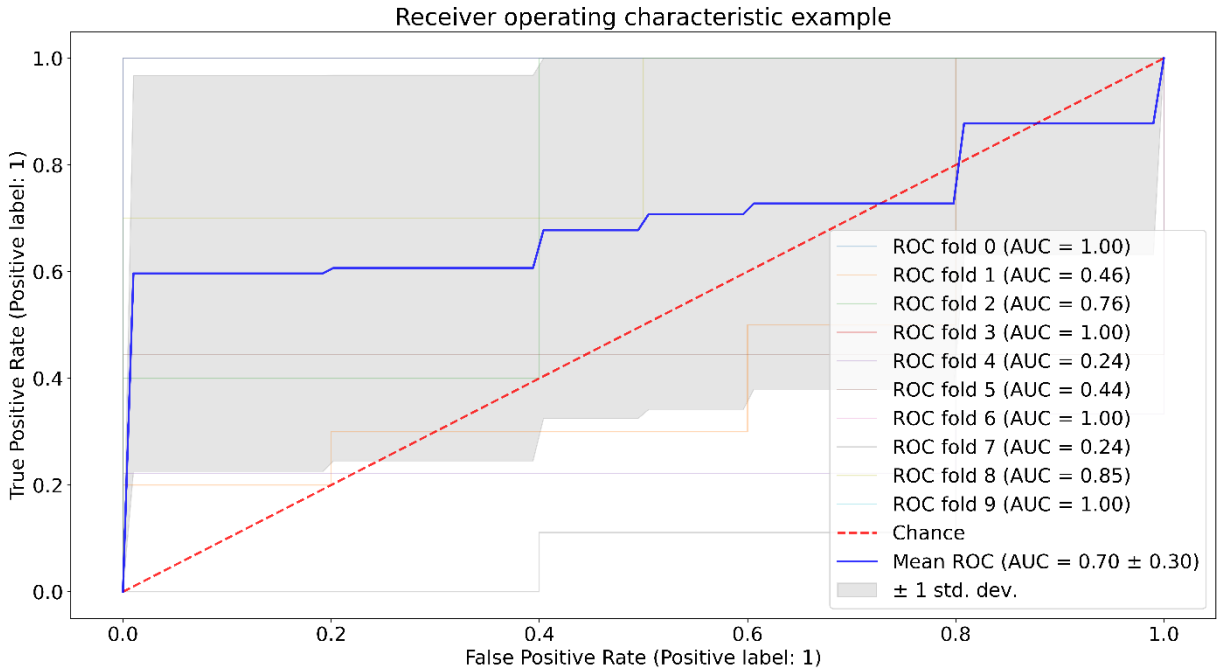


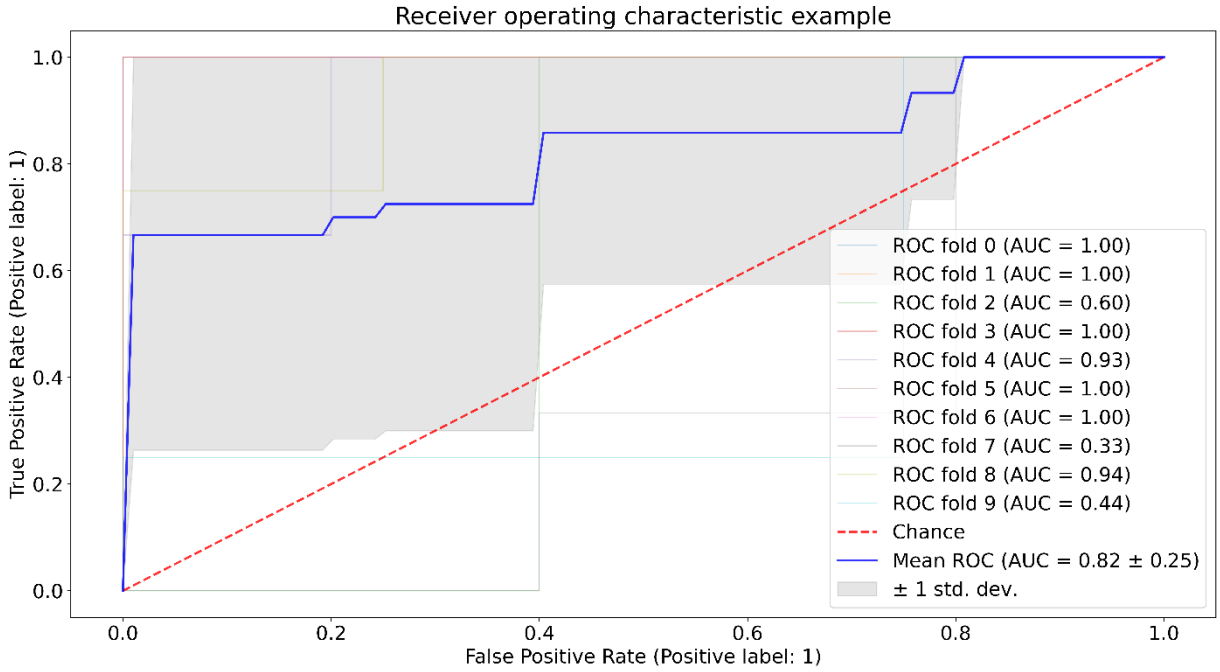
Figure S4. PCA score plot with data standardization (a) and without (b) for glioma patients and patients without glioma (with skull craniectomy defects), for PC1 and PC5.


Receiver operating characteristic example
True Positive Rate (Positive label: 1)
False Positive Rate (Positive label: 1)
0.0
0.2
0.4
0.6
0.8
1.0
ROC fold 0 (AUC = 1.00)
ROC fold 1 (AUC = 1.00)
ROC fold 2 (AUC = 0.60)
ROC fold 3 (AUC = 0.60)
ROC fold 4 (AUC = 0.33)
ROC fold 5 (AUC = 1.00)
ROC fold 6 (AUC = 1.00)
ROC fold 7 (AUC = 0.20)
ROC fold 8 (AUC = 0.50)
ROC fold 9 (AUC = 1.00)
Chance
Mean ROC (AUC = 0.72 ± 0.30)
± 1 std. dev.

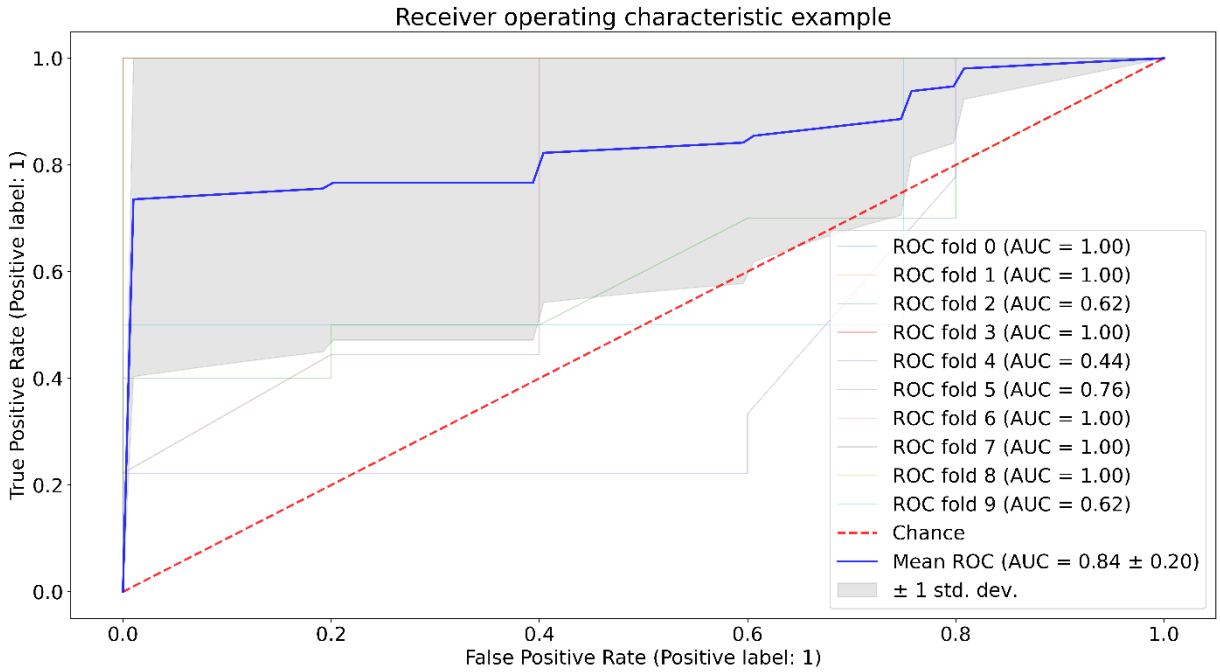
(a)


Receiver operating characteristic example
True Positive Rate (Positive label: 1)
False Positive Rate (Positive label: 1)
0.0
0.2
0.4
0.6
0.8
1.0
ROC fold 0 (AUC = 1.00)
ROC fold 1 (AUC = 0.46)
ROC fold 2 (AUC = 0.76)
ROC fold 3 (AUC = 1.00)
ROC fold 4 (AUC = 0.24)
ROC fold 5 (AUC = 0.44)
ROC fold 6 (AUC = 1.00)
ROC fold 7 (AUC = 0.24)
ROC fold 8 (AUC = 0.85)
ROC fold 9 (AUC = 1.00)
Chance
Mean ROC (AUC = 0.70 ± 0.30)
± 1 std. dev.

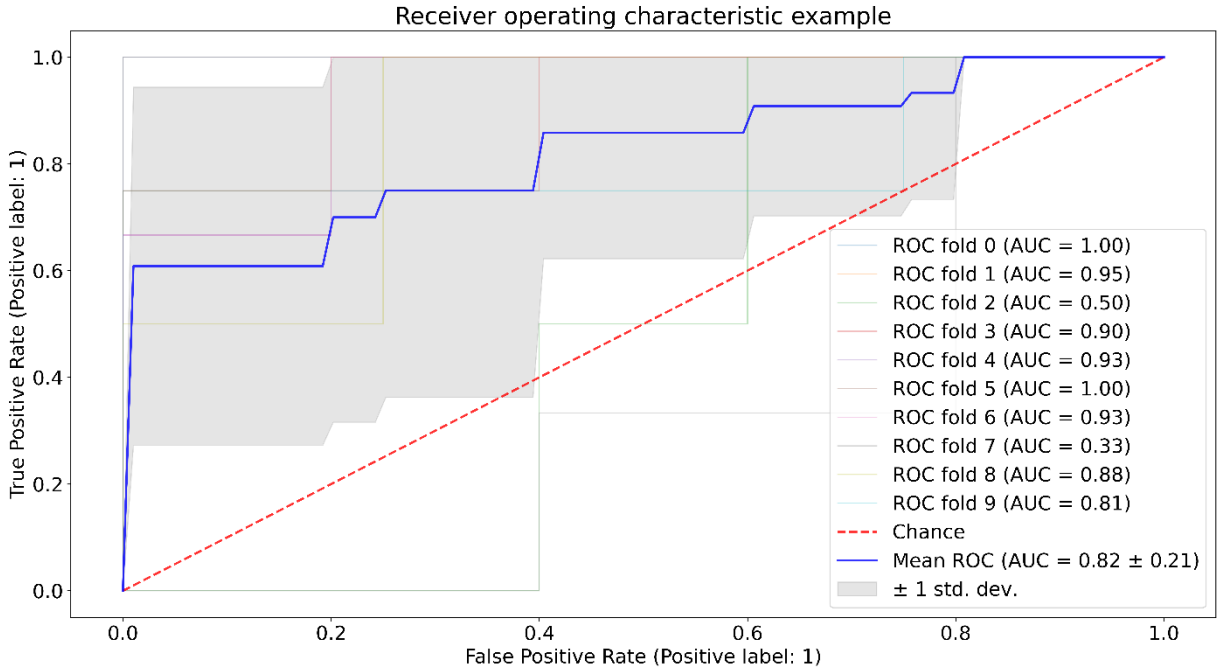
(b)


Receiver operating characteristic example
True Positive Rate (Positive label: 1)
False Positive Rate (Positive label: 1)
0.0
0.2
0.4
0.6
0.8
1.0
ROC fold 0 (AUC = 1.00)
ROC fold 1 (AUC = 1.00)
ROC fold 2 (AUC = 0.60)
ROC fold 3 (AUC = 1.00)
ROC fold 4 (AUC = 0.93)
ROC fold 5 (AUC = 1.00)
ROC fold 6 (AUC = 1.00)
ROC fold 7 (AUC = 0.33)
ROC fold 8 (AUC = 0.94)
ROC fold 9 (AUC = 0.44)
Chance
Mean ROC (AUC = 0.82 ± 0.25)
± 1 std. dev.

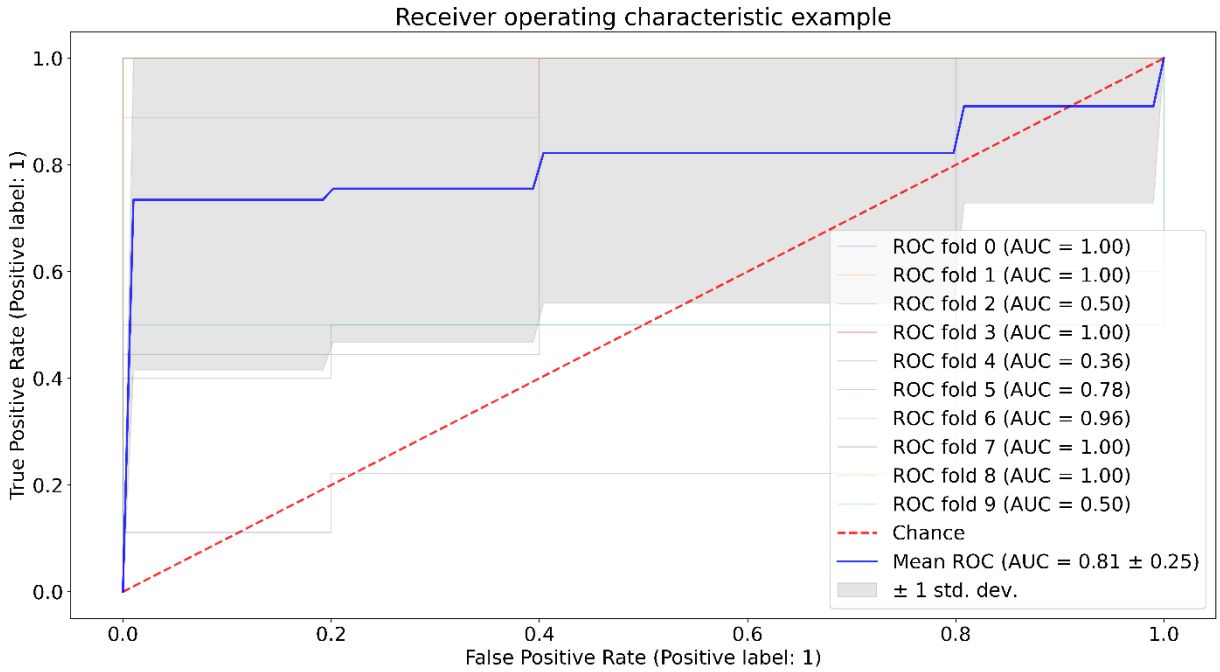
(c)


Receiver operating characteristic example
True Positive Rate (Positive label: 1)
False Positive Rate (Positive label: 1)
0.0
0.2
0.4
0.6
0.8
1.0
ROC fold 0 (AUC = 1.00)
ROC fold 1 (AUC = 1.00)
ROC fold 2 (AUC = 0.62)
ROC fold 3 (AUC = 1.00)
ROC fold 4 (AUC = 0.44)
ROC fold 5 (AUC = 0.76)
ROC fold 6 (AUC = 1.00)
ROC fold 7 (AUC = 1.00)
ROC fold 8 (AUC = 1.00)
ROC fold 9 (AUC = 0.62)
Chance
Mean ROC (AUC = 0.84 ± 0.20)
± 1 std. dev.

(d)



(e)



(f)

Figure S5. ROC AUC analysis: (a) – linear kernel SVM case 1; (b) linear kernel SVM case 2; (c)- RF case 1; (d)- RF case 2; (e) – XGBoost case 1; (f)- XGBoost case 2.

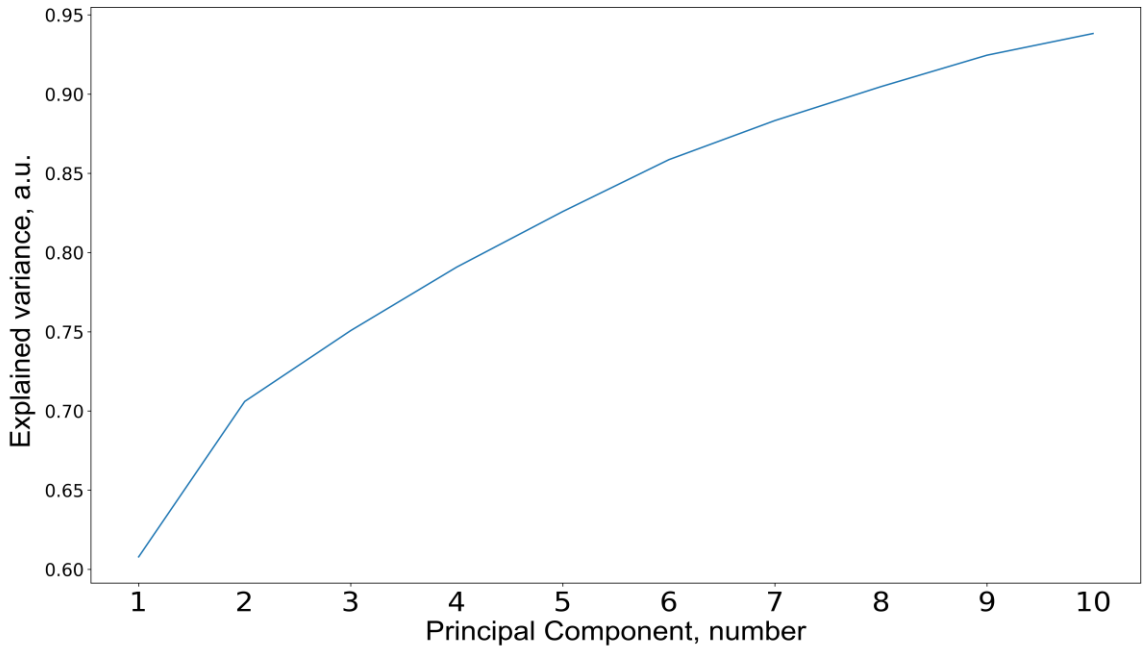
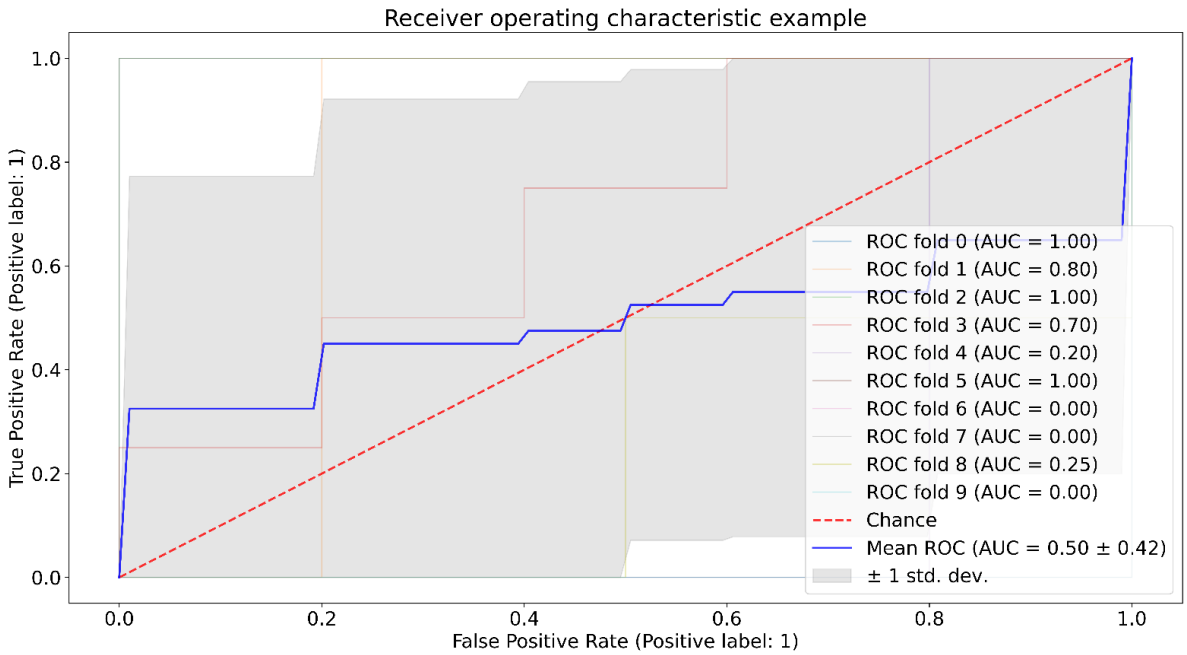
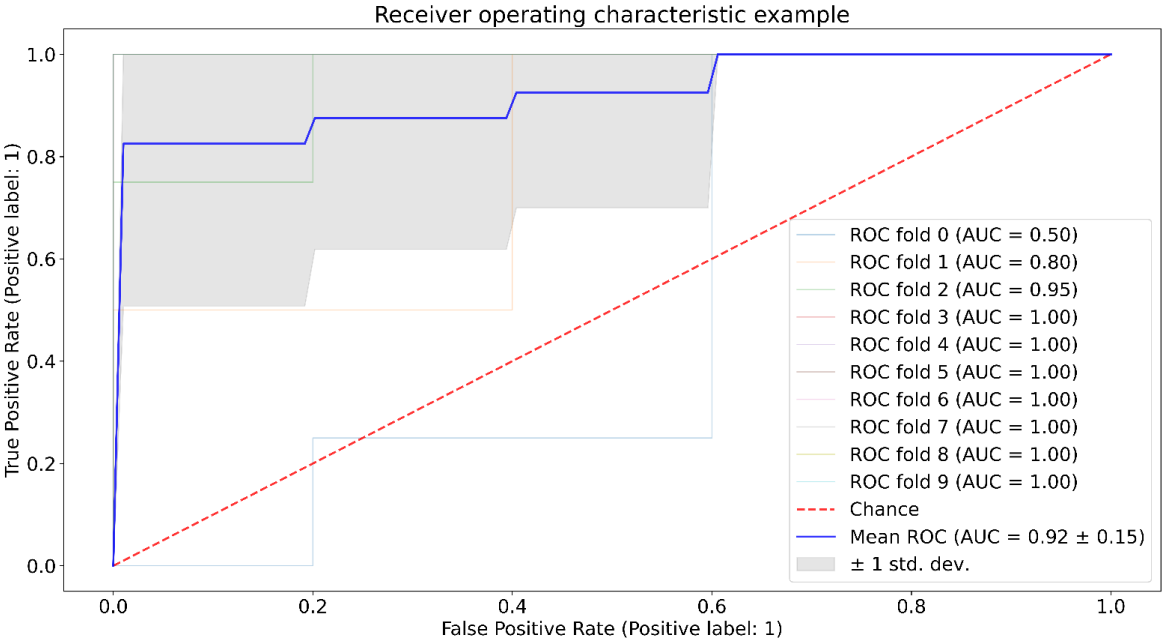


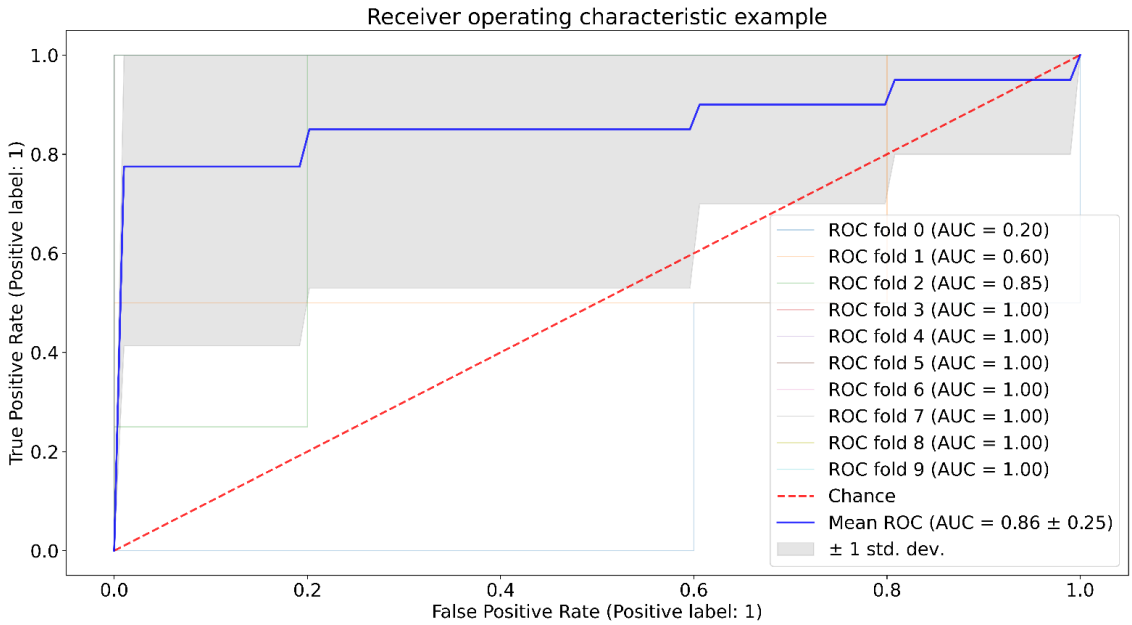
Figure S6. A graph of cumulative explained variance for determining the acceptable number of principal components.



(a)



(b)



(c)

Figure S7. Graph of ROC-AUC analysis of different ML methods for separation of blood plasma of glioma patients before and after surgery: (a) – linear kernel SVM; (b) - RF; (c) - XGBoost.