Ensemble Machine Learning Models for Tumor Sample Classification

TCGA: UCEC, KICH, LIHC, ESCA

Team Members

Tasnimul Hasnat
190041113

Nur Nasrum
1900411106

Khaja Abdus Sami
190041136

Introduction

Objective

Classification of UCEC, KICH, LIHC, ESCA tumor samples by tumor type using various types of machine learning models and ensembling them by max and avg voting.

Dataset Size

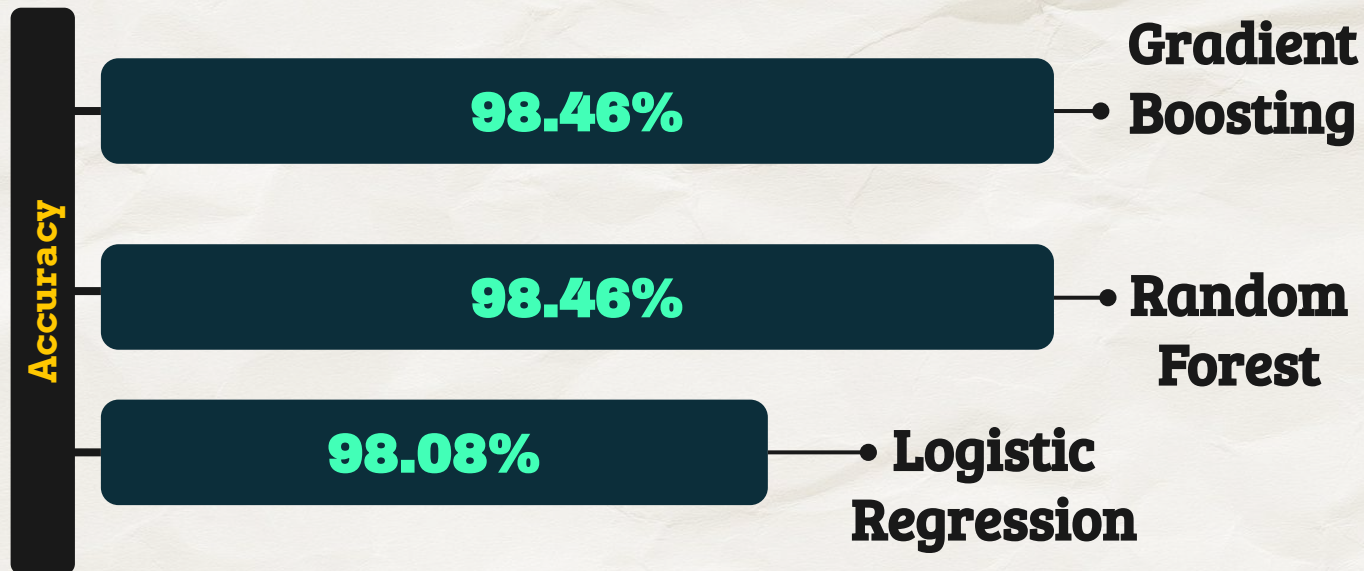**Number of Samples**
1297

Number of Genes
60660

Dataset Size

	Tumour	Normal	Total
TGCA-UCEC	553	35	588
TGCA-KICH	66	25	91
TGCA-LIHC	371	50	421
TGCA-ESCA	184	13	197
Total	1174	123	1297

Ensemble Learning

Max Voting

Each model in the ensemble makes a prediction for a given input, and the final prediction is determined by the mode of all the individual models.

Average Voting

Each model in the ensemble assigns a probability to each class for a given input. The final prediction is then determined by averaging these probabilities across all models.

Comparison
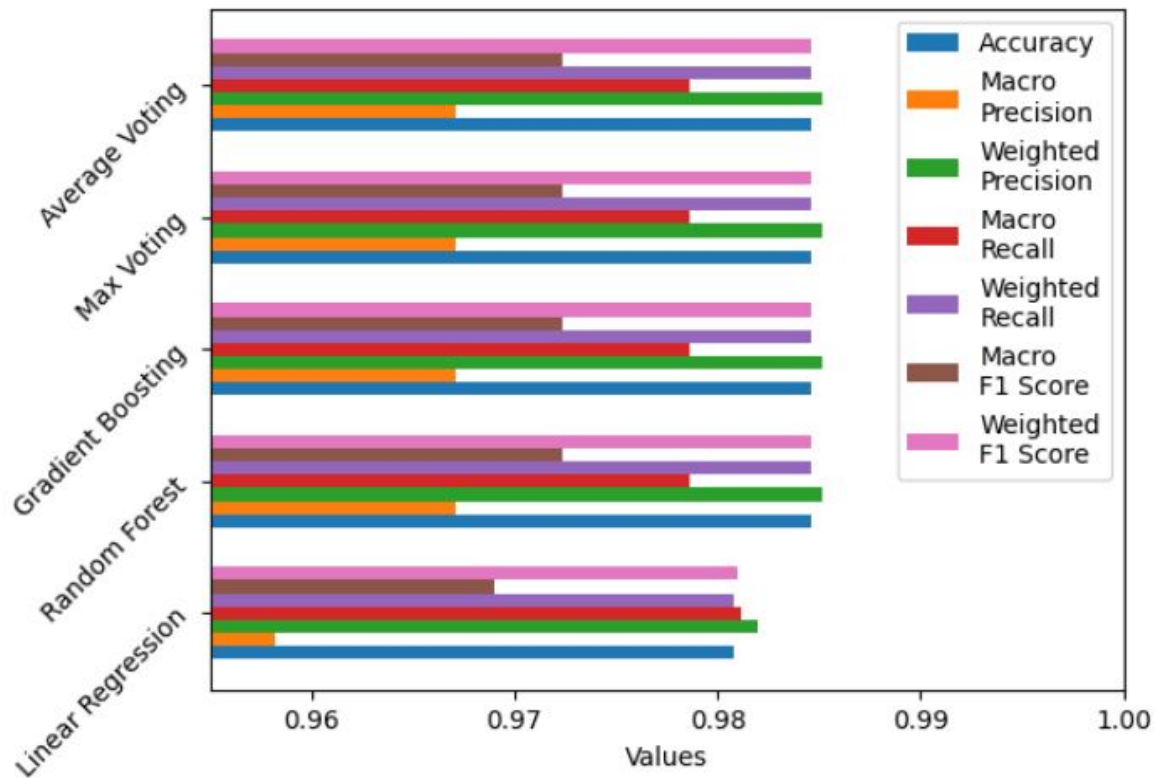
Comparative Analysis

Classifier	Accuracy	Macro Precision	Weighted Precision	Macro Recall	Weighted Recall	Macro F1 Score	Weighted F1 Score
Linear Regression	0.981	0.958	0.982	0.981	0.981	0.969	0.981
Random Forest	0.985	0.967	0.985	0.979	0.985	0.972	0.985
Gradient Boosting	0.985	0.967	0.985	0.979	0.985	0.972	0.985
Max Voting	0.985	0.967	0.985	0.979	0.985	0.972	0.985
Average Voting	0.985	0.967	0.985	0.979	0.985	0.972	0.985

Thank You!