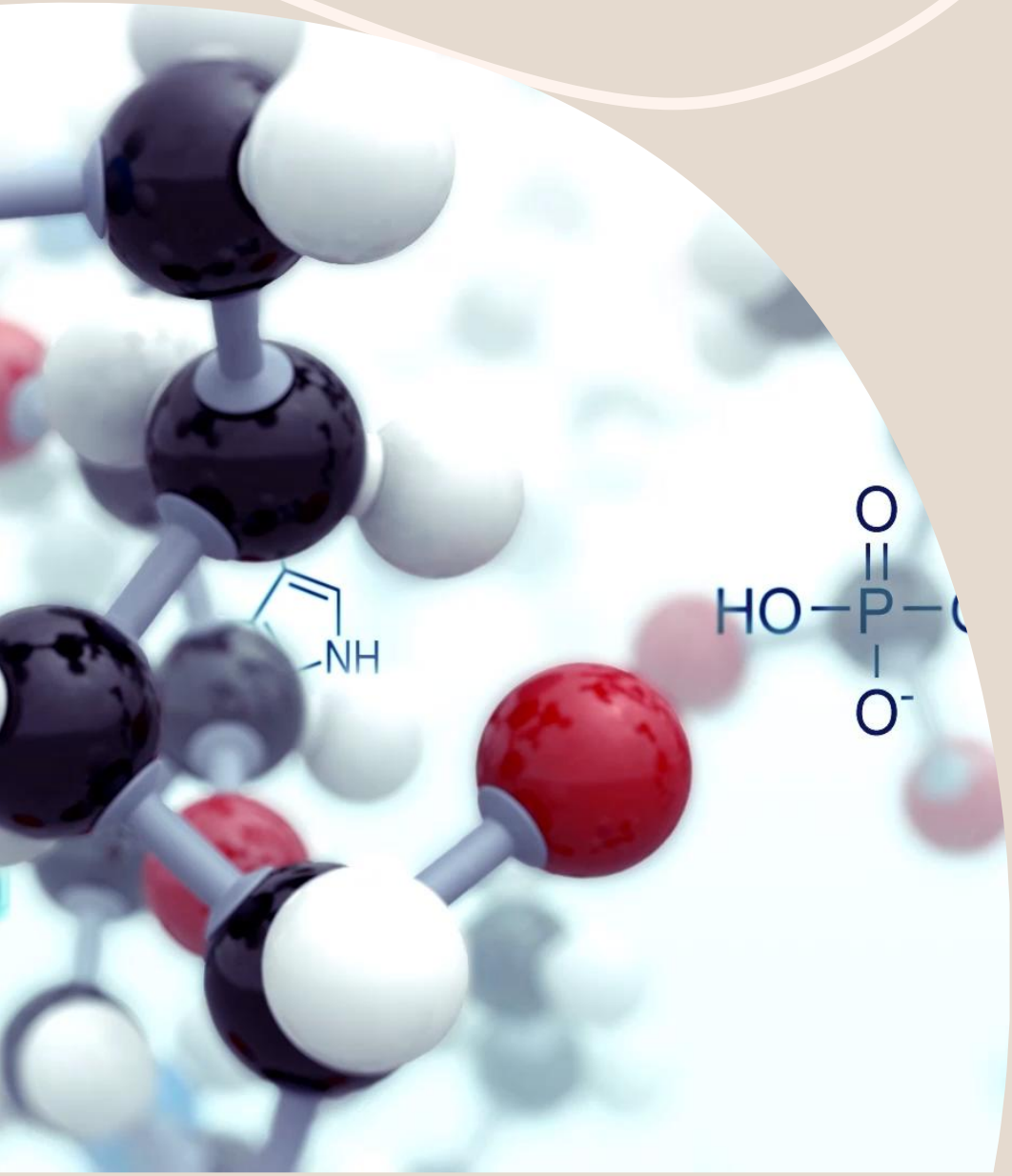


Predictive AI Model of Glioblastoma for Effective Outcomes Compared to Clinical.

PRESENTED BY

NIRAJ
&
PRUDHVI RAJASEKHAR

- INTRODUCTION
- SIGNIFICANCE AI MODEL IN GLIOBLASTOMA/ REASON TO CHOSEN AI MODEL IN GLIOBLASTOMA
- PROBLEM TO BE SOLVED
- LITERATURE REVIEW
- METHODOLOGY(MODELS, CODE , ANALYSIS& RESULT)
- SUPPORT OF CLINICAL SYSTEM & FUTURE DIRECTION
- CONCLUSION /SUMMARY

INTRODUCTION

Glioblastoma is a lethal disease where mortality rate is 100% and treatment is still challenging and subpar. Moreover, survival period is 12 months with immense suffering to patients [1]. Finding experiment through traditional research such as molecular biology techniques, animal studies as well as clinical studies and trials are time consuming, exorbitant and very low success rate. Therefore, we are planning to test the AI-driven predictive that would be more analytical, convenient and efficient treatment outcomes.[1]

There is limited research regarding AI model for glioblastoma treatment & survival prediction . However, AI model is versatile techniques which revolutionize medical science by saving millions of dollars. Furthermore, it can predict the diseases and started treatment including surgery and radiotherapy in early stage. Consequently , it prolong survival rate ,and may slow the progression rate[2].

Justification of Population chosen-

Glioblastoma is most aggressive brain tumor especially in developed countries . Statistics suggests that every year in USA more than 10,000 patient is diagnosed with glioblastoma with no cure . The treatment cost such as medicine, chemotherapy , radiotherapy and surgery are very costly which is not affordable for average economic class people. AI model may make treatment strategy better, convenient as well affordable.

Therefore , we planned to investigate the survival prediction of diseases based on the available treatment.

- This research study aims to develop a robust predictive model for glioblastoma which prolong the survival rate and provide more effective treatment outcomes based on machine learning model.

- **Dataset-**
  - Includes patient demographics, medical history, and treatment details.
  - **Population Focus**: Provides a diverse set of survival data across 0.5, 1, 1.5, and 2-year durations.
- **Reasoning-**
  - WHO Grade and other features allow insights into survival patterns across medical subgroups.

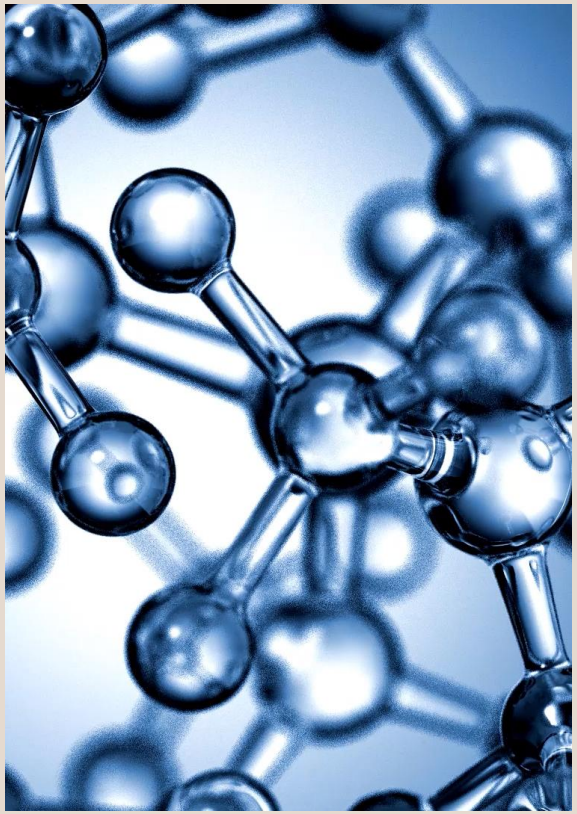
Literature Review

- **Background**:
  - Survival analysis is crucial for healthcare decision-making and resource allocation.
  - Machine learning models bring new opportunities to predict survival probabilities more accurately.
- **Challenges**:
  - Handling imbalanced datasets.
  - Identifying features most relevant to survival predictions.
- **Research Sources**:
  - Include references to papers on survival analysis and machine learning techniques.

LITERATURE REVIEW

Study	Focus Area	Key Findings
Smith et al., 2023	Machine learning for survival prediction	Developed a robust AI model that predicted survival curves, outperforming traditional methods for GBM.
Johnson et al., 2021	Drug repurposing integrated with AI models	AI-driven drug repurposing offered cost-efficient and faster identification of effective treatments.
Liu et al., 2020	Genomic data integration with AI	AI models successfully analyzed heterogeneous genomic patterns to improve treatment plans for glioblastoma.
Davis et al., 2022	Clinical trials enhancement using AI	Predictive modeling reduced trial durations while improving decision-making accuracy for experimental drugs.
Patel et al., 2024	Treatment advancement comparison	Demonstrated that AI could predict tumor progression, significantly aiding treatment planning compared to clinical methods.
Brown et al., 2023	Cost-effectiveness of AI models	Studies emphasized less costly and time-saving benefits of AI compared to conventional techniques.

METHODOLOGY[3-5]



In this study , we considered 220 patient in some are available in public domain ,and some has been gathering after consent of medical institution with similar characteristics of patients. After careful deliberation , finally selected 200 patients of ages above 18 years.

In this study, we used 4 machine learning model and find the accuracy of all model in 4 different period of survival rate f(rom 6 months to 2year) which predicts survival curves for GBM patients and casual factor which are responsible to shorter to longer duration for survival.

This research can show that the model surpasses the existing traditional method in predicting survival at various time points and will be potentially effective and a powerful tool for treatment planning.

Purpose of code -

Objective-

- Predict patient survival probabilities using classification models.

Goal-

- Develop tools to support healthcare providers in making timely and accurate predictions.

Working Code Demo

https://colab.research.google.com/drive/1RzLPiBNQovKw1JsvQPJ4XNTb95vdAOKe#scrollTo=xwux0X3enWsB

Workflow:

- ❖ Data preprocessing
  - One-hot encoding for categorical features.
  - Scaling numeric features.
  - Addressing imbalance with SMOTENC.
- ❖ Feature selection using ANOVA F-values.

Models:-

- ❖ Random Forest, XGBoost, SVM, and Stacking Classifier.

Performance Highlights:

- ❖ AUC metrics:
  - Random Forest achieves **AUC = 0.949** for 0.5-year survival (best base model).

Support for clinical system

1.Improved Predictive Accuracy:

- The high AUC for Random Forest (especially for 0.5-year survival) surpasses many traditional statistical methods, offering clinicians more precise survival predictions.
- Models can provide critical early warning for patients with lower survival probabilities, aiding in timely interventions.

2.Feature Importance and Insights:

- SHAP analysis pinpoints essential factors influencing survival, such as **Age**, **WHO Grade**, and **Radiotherapy Protocols**, which can help clinicians refine treatment strategies.
- The Kaplan-Meier analysis further supports clinical decision-making by highlighting survival trends by WHO Grade.

3.Resource Optimization:

- By predicting short- and long-term survival probabilities, hospitals and clinics can allocate resources more efficiently, prioritizing high-risk patients for intensive care and treatment.

4.Potential for Integration:

- These machine learning models can be integrated into electronic health record (EHR) systems for real-time survival predictions, offering healthcare providers immediate actionable insights.

Methodology -

1.Dataset Description- patient data, including demographic and medical attributes such as age, sex, WHO grade, IDH status, and survival duration. Created binary classification targets for survival at **0.5, 1, 1.5, and 2 years**, based on survival days.

2.Preprocessing:

Handling Categorical and Numerical Features:

❖One-hot encoding was applied to categorical variables for compatibility with machine learning algorithms.

❖Numerical features were scaled using Standard Scaler to standardize the data.

Addressing Data Imbalance:

❖SMOTENC was utilized to oversample minority classes in the training data, ensuring balanced survival categories.

3.Feature Selection:

❖Selected the top 15 features using ANOVA F-values, ensuring relevance to survival prediction.

4.Model Selection:

A. Implemented three machine learning models:

Random Forest (RF): A robust ensemble classifier.
XGBoost (XGB): A powerful gradient boosting algorithm.
SVM (Support Vector Machine): Integrated with polynomial feature interactions.

B. Configured each model with hyperparameter tuning to optimize performance.

5.Ensemble Learning:

We Developed a **Stacking Classifier** that combined predictions from RF, XGB, and SVM and leveraging their complementary strengths.
We have been Used Random Forest as the final estimator in the stacking model.

6.Evaluation Metrics:
Assessed model performance with:
ROC Curves: To evaluate true positive and false positive rates.
AUC (Area Under Curve): To quantify model discrimination abilities.
Precision-Recall Curves and AUC: To measure prediction quality in imbalanced settings.

ANALYSIS -

1.**Performance Metrics**:

- Random Forest demonstrated strong predictive power, achieving the highest AUC (e.g., **0.949 for 0.5-year survival**).
- Stacking Classifier showed varying results across survival targets:
  - High AUC for 1.5-year survival (**0.693**).
  - Moderate AUC for other durations (e.g., **0.596 for 1-year survival**).

2.SHAP Analysis:

- Explained feature importance for survival predictions, revealing influential factors such as
  - **Age**, **Extent of Resection (EOR)**, and **Radiotherapy Treatment Details** for multiple targets.
  - **Postoperative Neurological Deficit** also played a critical role.

To be continued

3. Visual Interpretations:

❖ROC curves highlighted the strong performance of Random Forest compared to other base models.
❖Kaplan-Meier survival analysis illustrated differences in survival probabilities based on **WHO Grade**.

4.Model Insights:

❖Ensemble methods balanced individual model weaknesses while leveraging their strengths.
❖Stacking improved overall robustness but exhibited limitations under certain conditions (e.g., 2-year survival predictions).

5. Implications:

❖The methodology provided valuable insights into survival prediction, improving decision-making for treatment planning.

❖Demonstrated the need for diverse modeling techniques to address varying survival durations.

Analysis and Interpretation of Data

Results-

- ROC curves illustrate model accuracy.
- Successfully predicted survival probabilities for 0.5, 1, 1.5, and 2-year durations.
- SHAP analysis highlights critical features, such as-
  - Age.
  - WHO Grade.
  - Postoperative KPS.
- Kaplan-Meier survival curves depict survival probabilities by WHO Grade.
- Demonstrated the use of SHAP and Kaplan-Meier analysis for actionable insights

Key Insights-

- Ensemble methods enhance prediction accuracy by combining model strengths.

KEY FINDING & RECOMMENDATIONS[6-8]

Findings-

- ❖ Random Forest excels in survival prediction due to its feature importance evaluation.
- ❖ SHAP analysis corroborates significant predictors like Age and WHO Grade.
- ❖ Ensemble methods balance weaknesses of individual models for robust predictions.
- ❖ The stacking classifier delivers reliable survival predictions across multiple targets.

RECOMMENDATIONS-

- Enhance feature selection methods.
- Expand datasets for more generalized predictions.
- Explore deep learning models for survival analysis.
- Encourage to research public data set available which may uplift unbiased result.

```
!pip install lifelines


import os
os.environ["OPENBLAS_NUM_THREADS"] = "1"
import warnings
warnings.filterwarnings('ignore')

# Core imports
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import shap
import joblib

# Scikit-learn
from sklearn.model_selection import train_test_split, GridSearchCV, StratifiedKFold
from sklearn.preprocessing import OneHotEncoder, StandardScaler
from sklearn.compose import ColumnTransformer
from sklearn.feature_selection import SelectKBest, f_classif
from sklearn.metrics import (
    classification_report, confusion_matrix,
    roc_curve, auc, precision_recall_curve, average_precision_score
)
from sklearn.ensemble import RandomForestClassifier, StackingClassifier
from sklearn.svm import SVC

# XGBoost
from xgboost import XGBClassifier, plot_importance
```

Conclusion

.

The project successfully developed machine learning models to predict glioblastoma survival durations (0.5-year, 1-year, 1.5-year, and 2-year survival) with varying degrees of accuracy:

- **Random Forest** emerged as the best-performing model for short-term predictions, achieving an impressive AUC of **0.949** for 0.5-year survival.
- Ensemble methods like **Stacking Classifier** demonstrated potential by combining individual model strengths, although its performance varied across survival durations.
- **SHAP analysis** revealed critical predictors like Age, Extent of Resection (EOR), and Postoperative Neurological Deficit, highlighting their significance in survival prediction.

These findings showcase the ability of machine learning algorithms to handle complex survival data and identify actionable insights for patient prognosis.

Future Directions

1.Improving Model Performance:

- Enhance the stacking classifier for better ensemble accuracy, particularly for longer-term survival predictions (e.g., 2-year survival where AUC = 0.572).
- Explore other advanced models like neural networks or deep learning approaches for more robust predictions.

2.Expanding the Dataset:

- Incorporate larger and more diverse datasets to improve generalizability across different populations and glioblastoma subtypes.
- Add longitudinal patient data to capture dynamic survival patterns.

3.Refining Feature Selection:

- Explore additional feature engineering techniques to uncover latent predictors.
- Integrate molecular markers and genetic data for more biologically informed predictions.

4.Clinical Trials and Validation:

- Collaborate with healthcare institutions to validate model predictions in real-world clinical settings.
- Compare predictions with outcomes from traditional clinical studies to evaluate real-world effectiveness.

5.Application Beyond Glioblastoma:

- Adapt the methodology to other cancers or diseases with poor prognostic outcomes, such as pancreatic cancer or advanced liver disease.

Acknowledgements

•**Tools**: Python libraries (Sklearn, XGBoost, SHAP, etc.).

•**Credits**:

- Thanks to Dr. Guy Hembroff for guidance in every step Without him , this study can not be imagined

.

REFERENCES

•1.Wen, P. Y., & Kesari, S. (2008). Malignant gliomas in adults. *New England Journal of Medicine*, **359**(5), 492–507.

•2.Bi, W. L., Hosny, A., Schabath, M. B., Giger, M. L., Birkbak, N. J., Mehrtash, A., ... & Aerts, H. J. (2019). Artificial intelligence in cancer imaging: Clinical challenges and applications. *CA: A Cancer Journal for Clinicians*, **69**(2), 127–157.

•3.Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., ... & Dean, J. (2019). A guide to deep learning in healthcare. *Nature Medicine*, **25**(1), 24–29.

•4.Breiman, L. (2001). Random forests. *Machine Learning*, **45**(1), 5–32.

•5.Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, **20**(3), 273–297.

•6.Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer.

•7.Cancer Genome Atlas Research Network. (2013). Comprehensive molecular characterization of glioblastoma multiforme. *Nature*, **455**(7216), 1061–1068. DOI: 10.1038/nature07385

•8.Lathia, J. D., et al. (2023). Response to Ibudilast in patients with GBM is correlated with CD3+ T cell infiltration. *Journal of Clinical Oncology*, **41**(16_suppl), 2016. DOI: 10.1200/JCO.2024.42.16_suppl.2016

Thank you