Group Project 2- NIRAJ NIRAJ & RAJ SABBAVARAPU

(Work equally done by both the candidates such as reviewing literature, methodology, clinical trial selection, manuscript writing and citing and references)

5114 progress report

Title-

Predictive AI Model for Glioblastoma Treatment for Effective Outcomes Compared to Clinical Outcomes Result.

1) **strong Introduction of the problem you are working to solve –**

   Glioblastoma is a highly lethal disease, and treatment is challenging and still subpar in which survival period is 12-18 months as well it has almost 100% fatality rate [1]. Finding treatment through traditional research such as molecular biology techniques, animal studies as well as clinical trials are time consuming, exorbitant as well very low success rate. Therefore, we are planning to test the AI-driven predictive that would be more analytical, should have more efficient treatment outcomes and should be quick, cost-effective and much potential.

   There is limited research regarding AI model for glioblastoma treatment prediction [2]. If successful, our model revolutionized medical science by saving millions of dollars and precious time for health organizations in terms of patient outcomes.

2) **contain a methodology of how you plan to solve the problem using AI models (be specific)**

Methodology

1. Data Collection and Preprocessing

To analyze Glioblastoma treatment outcomes, we will utilize data from two clinical trials number **NCT02780024 and** NCT**03782415** that contain patient demographics, treatment protocols and survival outcomes. The steps include:

- Data Extraction: Collect structured and unstructured data, including patient response to treatments, biomarkers.

- Data Normalization: Handle missing values, standardize features (e.g., age, dosage, genetic markers).[3]

2. AI Models for Prediction & Analysis

To extract meaningful insights, we will deploy multiple AI models:

Predicting Treatment Response Using Machine Learning

- **Model**: Random Forest and Support Vector Machines **[4,5]**
- **Purpose:**
  - Predict patient survival and treatment response based on clinical trial data.
- **Process:**
  - Use structured data (age, tumor size, genetic mutations, treatment type).
  - Train models on labeled survival data (e.g., 6-month, 12-month survival).
  - Evaluate using ROC-AUC score, Precision, Recall, and F1-score.

3. Model Validation

- **Cross-validation**: Use k-fold cross-validation to test model performance.[6]

Evaluation-

By combining machine learning for treatment response prediction, and therapy recommendations, this AI-driven approach aims to enhance Glioblastoma treatment strategies based on real-world clinical trial data.

3) discuss the datasets (see below for more details) you plan on using within your models

I. Clinical trial data (NCT02780024 and NCT**03782415)** Datasets will be taken into the account. [7-8]

A. Clinical Trials of NCT02780024 Dataset [7]

We will utilize real-world data from publicly available that focuses on Glioblastoma treatment and outcomes:

- **Study Objective**: This dataset will be evaluated and used for training predicting models (such as random forest) to analyze treatment response and survival prediction. This dataset provides genomic, clinical and imaging data on glioblastoma patients.
- **Key Features from the Dataset:**
    - Patient demographics (age, sex, genetic mutations).
    - Tumor characteristics (size, location, progression over time).
    - Treatment regimen (immunotherapy dosage, frequency).
    - Patient outcomes (progression-free survival, overall survival rates).

Relevance to AI Models:

- This dataset will be used for training predictive models like Random Forest and neural network (optional) to analyze treatment response and survival prediction.

- **AI Model Usage:** This dataset will be used to train **Random Forest and Neural Networks** models for analyzing treatment response and survival predictions.

B. Clinical trials of NCT03782415 [8]-

- **Study Objective**: Provides MRI scans for glioblastoma segmentation, aiding in tumor characterization and growth prediction
- **Key Features from the Dataset:**
    - Patient inclusion criteria (newly diagnosed vs. recurrent GBM).
    - Treatment protocol (TTFields intensity, chemotherapy combination).
    - Adverse effects and overall survival rates.
    - Imaging data (MRI scans with tumor segmentations)
    - Tumor subregion annotations (enhancing AI-driven detection)
    - Survival outcome records for patients

Relevance to AI Models:

- Comparison of TTFields response vs. standard treatments can enhance ML-based survival prediction models. And deep learning models (CNNs) are useful for tumor detection and progression prediction.

II. Supplementary Public Datasets for AI Model Training

To enhance the performance of AI models, we will incorporate publicly available datasets(optional): [7-9]

- BraTS (Brain Tumor Segmentation) Dataset:
    - Helps in understanding glioblastoma tumor shape, size, and growth patterns.
- TCGA-GBM (The Cancer Genome Atlas - Glioblastoma Multiforme Dataset):
    - Useful for training predictive models to analyze genetic markers affecting treatment outcomes.

III. Data Preprocessing Considerations-

- Normalization: Patient demographic and lab values will be normalized to ensure consistency across datasets.
- Feature Selection: Key biomarkers, tumor measurements, and clinical parameters will be extracted for model training.

Predictive outcomes-

By integrating clinical trial data with our AI models, our aim will be able to predict treatment responses accurately, analyze tumor progression, and Enhanced Glioblastoma therapy.

If our model has positive results, it could revolutionize Glioblastoma treatment strategies by providing a fast, cost-effective AI-driven approach to predict patient outcomes more effectively than traditional clinical methods

4. 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 Science & Business Media.

7. Cancer Genome Atlas Research Network. (2013). Comprehensive molecular characterization of glioblastoma multiforme. *Nature*, 455(7216), 1061–1068. https://doi.org/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. https://ascopubs.org/doi/10.1200/JCO.2024.42.16_suppl.2016

9.. Menze, B. H., et al. (2015). The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). *IEEE Transactions on Medical Imaging*, 34(10), 1993–2004. https://doi.org/10.1109/TMI.2014.2377694