Fall 2024 Cycle B, F01E, Project I

Report

12/19/2024

Medical Tech Analysts

Tenzin Micky
Denzel Green
Kionna Taylor
Michael Valenzuela

Table of Contents:

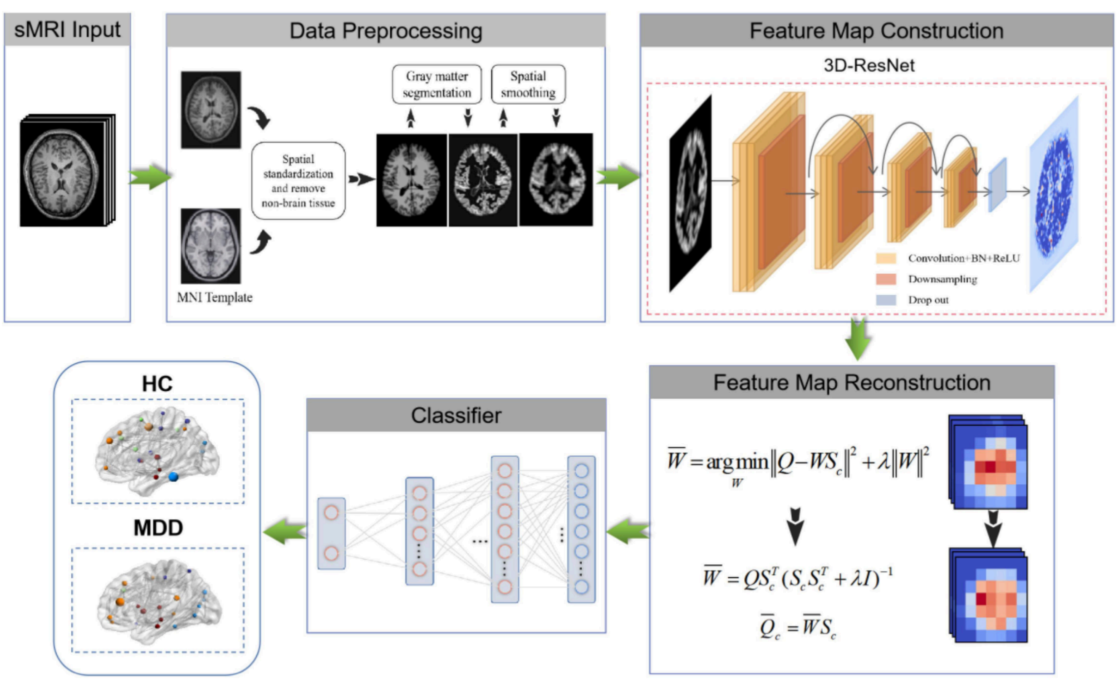
Synopsis

This project focused on creating an interactive visualization dashboard to analyze brain tumor data. Initially, the goal was to build a robust detection and classification system using machine learning techniques; however, due to constraints, the project pivoted toward a visualization-based approach. The final dashboard provides actionable insights for medical professionals, enabling efficient analysis of tumor data and patient outcomes.

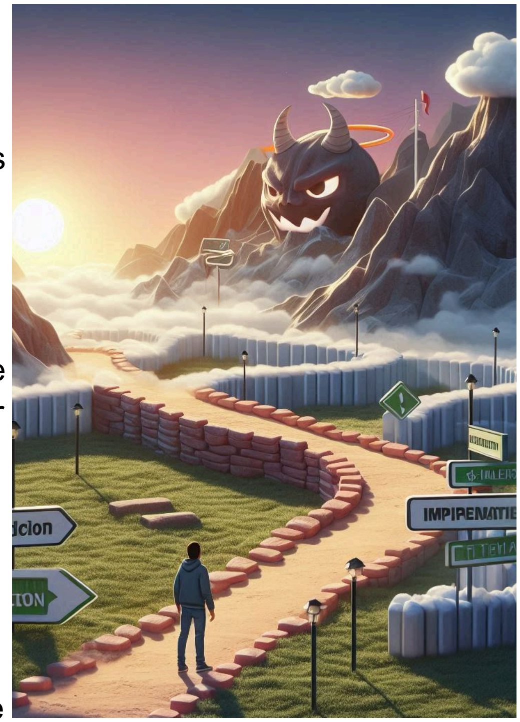
Original Proposal

Our initial proposal aimed to develop a brain tumor detection and classification system utilizing convolutional neural networks (CNNs). The system is intended to classify tumor types and provide statistical insights into tumor characteristics, paired with a user-friendly dashboard for stakeholders to interpret model outputs effectively.

We envisioned the dashboard as a tool that would integrate the outputs from the CNN models, enabling stakeholders to seamlessly interpret diagnostic results and access relevant statistical insights. However, as development progressed, we encountered significant challenges that made this goal unattainable within the given timeframe.



Reasons for Pivoting



1. **Ambitious Scope**: The original project aimed to implement multiple complex processes, including CNN model design, training, and integration, which proved to be overly ambitious given the constraints.
2. **Hardware and Software Constraints**:
    - Lack of GPU support for TensorFlow hindered model training and testing.
    - The sklearn library required a compatible C++ compiler that was unavailable in our setup.
3. **Knowledge Gap**: Despite extensive research, we underestimated the advanced expertise required to implement the machine learning components effectively within the project timeline. Although the professor gave useful advice on using cloud GPUs, the learning curve was too steep for the remaining time. The lack of prior experience with CNNs further compounded these challenges.

Team Hurdles

- **Limited Technical Resources**: We lacked the computational power necessary for deep learning model training, which stalled progress in the early stages.
- **Software Incompatibility**: The integration of required machine learning libraries presented compatibility challenges that consumed significant time.
- **Time Constraints**: The demanding timeline required us to continuously reevaluate our goals and adjust our scope to ensure a deliverable outcome.

Explanation of Errors & Compatibility

Several technical errors and compatibility issues surfaced during the project's initial phases:

- **Hardware Limitations**: Attempting to train CNN models without GPU support led to excessive processing times, making iterative testing impractical.
- **Software Conflicts**: The sklearn library required a compatible compiler, but setup issues prevented successful installations.
- **Overlooked Complexity**: The intricate preprocessing, model tuning, and evaluation processes demanded far more time and resources than initially anticipated.

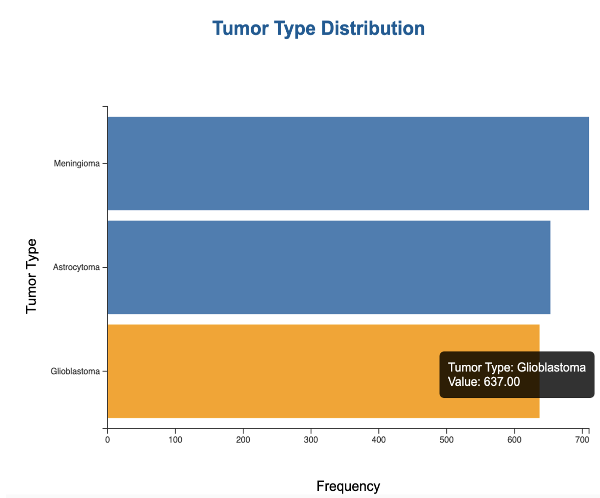
These hurdles ultimately steered us toward pivoting to a more feasible and impactful approach—the development of an interactive visualization dashboard.

New Idea and Dataset

Following the pivot, we shifted our focus to creating an **interactive visualization dashboard** to explore brain tumor recurrence patterns. This new approach leveraged a brain tumor dataset that was structured in a CSV format, allowing us to apply data preprocessing and visualization techniques effectively.

Key features of the dashboard include:

- **Interactive Charts**: Enable users to dynamically explore tumor types, treatment outcomes, and survival trends.
- **Statistical Insights**: Provide an intuitive overview of key metrics, helping researchers and clinicians identify actionable patterns.
- **Tooltips and Hover Effects**: Enhance user experience by delivering detailed insights on specific data points.



Purpose

The dashboard was designed to bridge the gap between raw medical data and actionable insights. Brain tumors are complex and multifaceted, requiring detailed analysis to identify crucial patterns that can aid in diagnosis and treatment. However, traditional data interpretation methods are often time-consuming and inaccessible to many medical professionals, hindering timely decision-making.

This dashboard addresses these challenges by transforming large, unstructured datasets into intuitive, interactive visualizations. By visualizing tumor progression and recurrence data, researchers and clinicians can:

- **Identify Patterns and Trends**: Quickly detect relationships between tumor characteristics, treatment outcomes, and patient demographics to better understand disease behavior.
- **Support Personalized Treatment Approaches**: Highlight correlations between tumor features and survival times, enabling clinicians to tailor treatments to individual patient needs.
- **Enhance Decision-Making Processes**: Provide a centralized platform for data exploration, offering stakeholders the ability to draw insights without requiring advanced technical expertise.
- **Facilitate Collaboration**: Create a shared tool for researchers, doctors, and other stakeholders to communicate findings and develop unified strategies for patient care.

Furthermore, the dashboard empowers medical professionals by incorporating advanced filtering options, dynamic visualizations, and statistical summaries. These features allow users to analyze the dataset from multiple perspectives, ensuring a comprehensive understanding of tumor dynamics and treatment outcomes. Ultimately, this tool serves as a bridge between data and action, helping stakeholders make informed, evidence-based decisions.

Explanation of New Dataset

The dataset, obtained in CSV format, comprises detailed records for 2,000 patients, including the following features:

1. **Patient Demographics**: Unique identifiers, age, and gender.
2. **Tumor Characteristics**: Type, grade, and location.
3. **Treatment Details**: Types of treatment, outcomes, and time to recurrence.
4. **Survival Metrics**: Survival time and recurrence patterns.

This dataset highlights the temporal dynamics of tumor progression and recurrence, offering a rich resource for analysis and visualization. Cleaning and preprocessing ensured the data was reliable and ready for integration into the dashboard.

BrainTumor.csv (194.63 kB)

Detail Compact Column 10 of 11 columns

Patient ID	Age	Gender	Tumor Type	Tumor Gra...	Tumor Loc...	
1	45	Male	Glioblastoma	IV	Frontal lobe	Su
2	55	Female	Meningioma	I	Parietal lobe	Su
3	60	Male	Astrocytoma	III	Occipital lobe	Su Ch
4	50	Female	Glioblastoma	IV	Temporal lobe	Su

Key Features

1. **Demographics Analysis**:
   - Charts displaying the distribution of patients by age and gender.
   - Insights into demographic trends within the dataset.
2. **Tumor Characteristics**:
   - Visualizations of tumor type and grade distributions.
   - Tumor location analysis to highlight common sites of occurrence.
3. **Treatment Pathways**:
   - Charts linking tumor type to treatment outcomes and survival times.
   - Analysis of treatment effectiveness based on recurrence and survival metrics.
4. **Survival Insights**:
   - Correlation analyses between age, tumor type, and survival times.

Problem Statement

Brain tumors present a significant challenge for medical professionals due to their complexity and variability. The vast amount of available data often lacks accessibility and structure, making it difficult for stakeholders to extract meaningful insights.

This project addresses these challenges by:

- Simplifying data interpretation through interactive visualizations.
- Highlighting key correlations between tumor characteristics, treatments, and patient outcomes.
- Providing a user-friendly interface that supports efficient decision-making and research.

Milestones & Timeline

Week 1: Project Setup and Initial Development

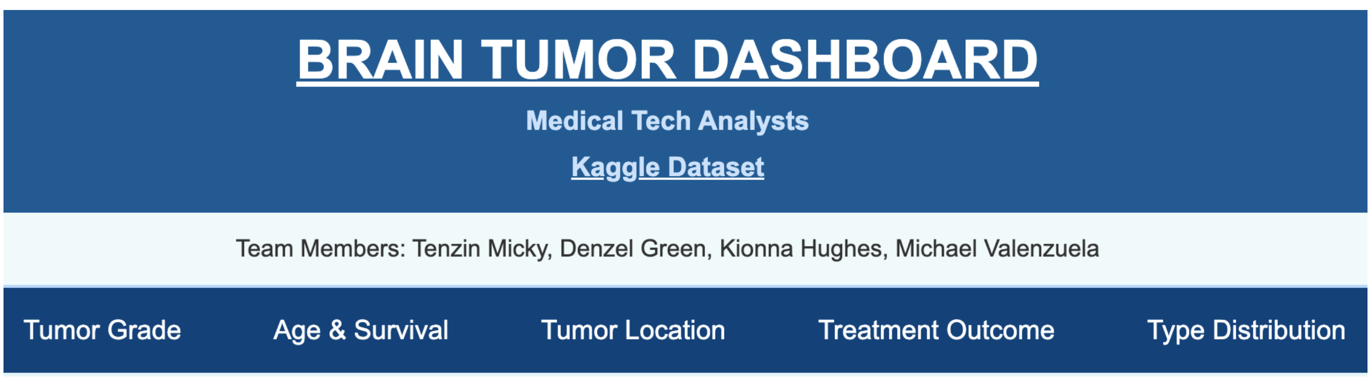
- Finalized the project plan, established milestones, and ensured dataset readiness.
- Set up the development environment and designed the dashboard’s initial layout.
- Conducted preliminary data analysis to identify trends and refine goals.

Week 2: Development and Data Visualization

- Implemented key visualizations, including bar charts, scatter plots, and pie charts.
- Integrated statistical findings into the dashboard.
- Enhanced interactivity through filters and dynamic tooltips.

Week 3: Finalization and Presentation

- Conducted testing for bugs, performance, and cross-device compatibility.
- Finalized the dashboard’s design and functionality.
- Prepared the final presentation and project report.



Data preprocessing involved:

- **Loading and Cleaning**: Imported the dataset from a CSV file and handled missing values by replacing NAs with zeros.
- **Data Transformation**: Converted string fields (e.g., tumor type, grade) to numerical values for analysis.
- **Filtering**: Removed irrelevant or inconsistent records to ensure data quality.
- **Validation**: Cross-checked processed data for accuracy and consistency before visualization.

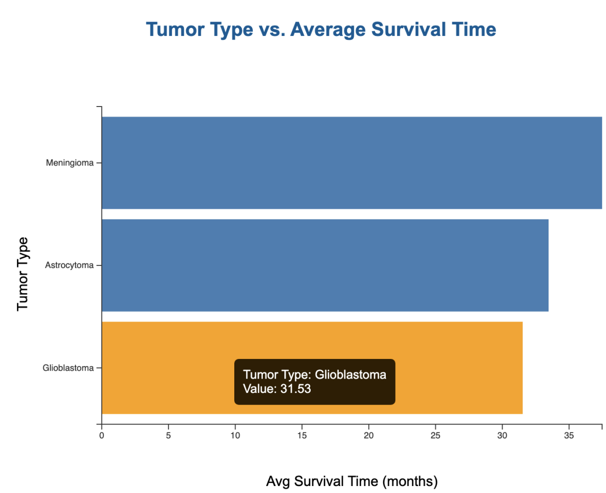
BrainTumor_Cleaned.csv > data

```
Patient ID,Age,Gender,Tumor Type,Tumor Grade,Tumor Location,Treatment,Treatment Outcome,Time to Recurre
1,45,0,Glioblastoma,IV,Frontal lobe,Surgery,Partial response,10.0,Temporal lobe,18
2,55,1,Meningioma,I,Parietal lobe,Surgery,Complete response,0.0,0,36
3,60,0,Astrocytoma,III,Occipital lobe,Surgery + Chemotherapy,Progressive disease,14.0,Frontal lobe,22
4,50,1,Glioblastoma,IV,Temporal lobe,Surgery + Radiation therapy,Complete response,0.0,0,12
5,65,0,Astrocytoma,II,Frontal lobe,Surgery + Radiation therapy,Partial response,24.0,Frontal lobe,48
6,45,0,Glioblastoma,IV,Frontal lobe,Surgery,Partial response,10.0,Temporal lobe,18
7,55,1,Meningioma,I,Parietal lobe,Surgery,Complete response,0.0,0,36
8,60,0,Astrocytoma,III,Occipital lobe,Surgery + Chemotherapy,Progressive disease,14.0,Frontal lobe,22
9,50,1,Glioblastoma,IV,Temporal lobe,Surgery + Radiation,Complete response,0.0,0,12
```

Statistical Analysis

Our statistical analysis focused on:

1. **Tumor Type Distribution**:
    - Highlighted the prevalence of glioblastoma, meningioma, and other tumor types.
2. **Survival Analysis**:
    - Explored the relationship between tumor types and survival times.
    - Identified glioblastoma as having the shortest average survival time.
3. **Demographic Trends**:
    - Examined age distributions and their correlations with treatment outcomes.
4. **Treatment Effectiveness**:
    - Assessed the success rates of different treatments, with surgery showing the highest success for meningioma.



Correlation Analysis

Key correlations analyzed include:

- **Tumor Type vs. Tumor Grade**: Understanding severity distributions.
- **Treatment Type vs. Treatment Outcome**: Evaluating treatment efficacy.
- **Age vs. Survival Time**: Examining the impact of age on prognosis.
- **Time to Recurrence vs. Survival Time**: Identifying patterns in recurrence and overall outcomes

	Age	**Gender**	**Survival Time (months)**	**Time to Recurrence (months)**
Age	1	-0.139558279	0.066281516	-0.050582495
Gender	-0.139558279	1	-0.096686649	-0.042887679
Survival Time (months)	0.066281516	-0.096686649	1	-0.274673949
Time to Recurrence (months)	-0.050582495	-0.042887679	-0.274673949	1

Correlation analysis showed minimal influence of demographic factors like age and gender on survival and recurrence. **Age** showed a weak positive correlation with survival time (**0.066**) and a weak negative correlation with time to recurrence (**-0.051**). **Gender** had slight negative correlations with survival (**-0.097**) and recurrence timing (**-0.043**). A moderate negative correlation between survival time and time to recurrence (**-0.275**) suggests that patients who survive longer may experience earlier tumor recurrence, possibly due to closer monitoring or tumor aggressiveness. These findings highlight the complex dynamics of survival and recurrence, warranting further investigation into additional contributing factors.

Treatment Analysis

The treatment pathway analysis revealed:

- **Meningioma treated with surgery and radiation shows the most success,** with a complete response and average survival time of **40.4 months**.
- **Astrocytoma has the longest survival time with surgery-only treatments** (66 months), highlighting its potential for recovery under aggressive care.
- **Glioblastoma is the most difficult tumor type,** with the shortest survival times and higher rates of disease progression.
- **Recurrence patterns** indicate quicker recurrence for Glioblastoma and Astrocytoma compared to Meningioma.

Example of the mathematical model used to get the first treatment pathway analysis.

Steps:

1. **Filtering**:
   - Adjusted treatment naming conventions to ensure "Surgery + Radiation" and similar variations were included.
   - Selected rows with "Meningioma" tumor type, treatments containing both "surgery" and "radiation," and outcomes marked as "Complete response."
2. **Computation**:
   - Used the mean formula: $\overline{S} = \frac{\Sigma S_i}{n}$
   - Where S_i are survival times for the selected patients, and n is the total number of patients in this subset.

Analysis of Brain Tumors and Survival Factors

Tumor Types

Our analysis focused on three major brain tumors: **Meningioma, Astrocytoma, and Glioblastoma.** Meningioma emerged as the most common, with **710 cases**, followed closely by **653 Astrocytoma cases** and **637 Glioblastoma cases.** In terms of survival, Meningioma showed the best outcomes, with an average survival time of **37.47 months.** Astrocytoma patients had an average survival of **33.46 months**, while Glioblastoma patients faced the most challenging prognosis, with an average survival of just **31.53 months.**

Survival by Tumor Grade

Survival outcomes were also examined based on tumor grades. **Grade I tumors** had the best prognosis, with an average survival time of **37.47 months.** Grades II and III followed, with survival times averaging **34.45 months** and **32.35 months,** respectively. **Grade IV tumors** had the poorest survival outcomes, averaging just **31.51 months.** These findings demonstrate a clear trend: as tumor grade increases, survival time decreases.

Age and Tumor Outcomes

The impact of age on tumor distribution and survival was also explored. The majority of patients were in the **50 to 60 age range,** accounting for over **1,086 cases.** Interestingly, survival rates were highest among patients aged **70 to 80,** with an average survival time of **41 months.** By tumor type, **Astrocytoma** was more common in patients aged **40 to 50,** while **Glioblastoma** was also prevalent in this group but associated with shorter survival times. **Meningioma** was most frequently observed in patients aged **50 to 60** and was linked to better survival outcomes.

Gender and Survival Outcomes

Gender also played a role in survival. For **Astrocytoma,** male patients had a higher average survival time (**35.3 months**) compared to females (**31.7 months**). Similarly, in **Glioblastoma,** males averaged **32.5 months,** while females averaged **30.4 months.** For **Meningioma,** survival times were nearly identical, with males at **37.6 months** and females at **37.3 months.** These results indicate that gender differences have a greater impact on **Astrocytoma** and **Glioblastoma** than on **Meningioma.**

Recurrence Patterns

Recurrence patterns varied significantly between tumor types. **Astrocytoma** recurred after an average of **15.3 months,** most commonly in the **occipital lobe. Glioblastoma** recurred more quickly, averaging **11.7 months** and primarily affecting the **parietal lobe. Meningioma** had the slowest recurrence, with an average of **8 months** and fewer documented recurrence sites. These findings highlight the aggressive nature of **Glioblastoma** compared to other tumors.

Design and Visualization

For our Brain Tumor Dashboard project, my primary focus was on designing an intuitive and visually appealing interface that enhances data interpretation and user interactivity. The dashboard's design integrates key visualizations that provide actionable insights into tumor types, treatment outcomes, and patient survival probabilities. Below is an overview of my contributions to the project.

Dashboard Layout and User Interface Design

- **Purposeful Color Scheme**:
  I implemented a blue and white color palette, as blue conveys professionalism, trust, and reliability—essential qualities for a healthcare-focused platform. The colors also ensure a clean, modern look that is easy on the eyes, making it suitable for users of all age groups.
- **Logical Chart Arrangement**:
  The charts were strategically placed to mimic clinical workflows:
  - Demographics and tumor types appear at the top to provide a high-level overview.
  - Treatment outcomes and survival probabilities follow, guiding users through a logical progression of the patient care journey.
  - The placement ensures seamless data exploration, reducing cognitive load for users.

Visualization Implementation

- **Bar Charts for Demographics and Tumor Types**:
  - These charts display the distribution of tumor types across various demographic groups, providing insights into how factors such as age, gender, and ethnicity influence tumor prevalence.
  - The bars are color-coded for clarity, with **hover tooltips** offering additional details on specific data points (e.g., counts and percentages).
- **Pie Charts for Treatment Outcomes**:
  - Treatment outcomes are represented in pie charts, allowing users to quickly understand the proportion of successful treatments versus adverse outcomes.
  - The interactivity enables users to click on specific segments to drill down into more granular data.
- **Kaplan-Meier Curves for Survival Probability**:
  - These curves illustrate patient survival probabilities over time, helping clinicians and researchers evaluate the effectiveness of treatments.

- Filters for tumor types and treatments dynamically adjust the curves, providing tailored insights.

Interactivity and Usability

- **Dynamic Filters**:
  I integrated filters that allow users to explore data by **tumor type, treatment, and demographic attributes**. This dynamic exploration ensures the dashboard remains adaptable to diverse user needs.
- **Hover Tooltips**:
  Detailed **hover tooltips** provide immediate access to underlying data for each visualization. For example, hovering over a bar or pie segment displays exact counts, percentages, or survival probabilities. This feature minimizes the need for users to switch contexts or reference external documentation.

Technical Considerations

- **Responsive Design**:
  The dashboard was designed to be responsive, ensuring usability across different devices and screen sizes, from desktop computers to tablets.
- **Consistent Visual Standards**:
  I ensured that font sizes, color codes, and chart dimensions adhered to a consistent standard, enhancing the dashboard's professional appearance and readability.
- **Scalability**:
  The charts and filters were implemented using D3.js, enabling smooth transitions and real-time updates as users interact with the dashboard. The design also allows for future integration of additional datasets and visualizations.

The design and visualizations I contributed to the Brain Tumor Dashboard project aim to balance aesthetic appeal with functionality. By emphasizing interactivity, clarity, and user-centric design, the dashboard provides stakeholders with a powerful tool for exploring critical data on brain tumors, treatments, and outcomes. My role focused on ensuring the visualizations were not only accurate but also actionable and accessible, aligning with the project’s goal of delivering insights for better decision-making in healthcare.

Physician Assistant Consultation

Following a consultation with a Physician Assistant (P.A.), we refined our focus to concentrate on key aspects of the dataset that provide the most meaningful insights for medical professionals. These aspects include:

1. **Tumor Type Analysis**:
   - **Tumor Type Distribution**: Understanding the prevalence of different tumor types within the dataset.
   - **Tumor Type vs. Survival Time (Prognosis)**: Analyzing the relationship between tumor type and patient survival time to assess prognosis trends.
2. **Age Analysis**:
   - **Age Distribution**: Exploring the age demographics of the patients.
   - **Age vs. Treatment Outcome**: Evaluating how treatment outcomes vary across different age groups.
   - **Age vs. Survival Time (Prognosis)**: Investigating survival trends based on age.
   - **Age vs. Tumor Type**: Identifying correlations between age and the types of tumors diagnosed.
3. **Treatment Pathway Analysis**:
   - Examining the progression from **Tumor Type** → **Treatment** → **Treatment Outcome** → **Time to Recurrence & Survival Time** to uncover patterns and assess the effectiveness of treatments.

This focused approach, guided by the expertise of the P.A., ensures that our analysis and visualizations align with the needs of medical professionals and provide actionable insights into patient outcomes.

Results

Key Findings from the Dashboard:

1. **Tumor Type Insights**:
    - Glioblastoma had the shortest average survival time across all tumor types.
    - Meningioma, when treated with surgery, showed the highest success rate (87%).
2. **Age-Related Trends**:
    - Younger patients (<40 years) generally had better treatment outcomes and longer survival times compared to older age groups.
    - Age was positively correlated with more aggressive tumor grades (Grade III and IV).
3. **Treatment Effectiveness**:
    - Surgery was the most effective standalone treatment across multiple tumor types.
    - Radiation therapy showed varying results depending on tumor type but was generally effective in delaying recurrence.
4. **Recurrence and Survival Correlation**:
    - Recurrence often occurred earlier in glioblastoma cases, leading to shorter overall survival times.
    - Patients with slower recurrence times tended to have higher survival rates.

Dashboard Impact:

- **Clinical Support**: According to our consultation. The dashboard offers medical professionals a streamlined way to identify critical insights and trends within seconds, aiding in faster, evidence-based decision-making. It needs more adjustments and improvement but is very much on it's way.
- **Research Facilitation**: Provides researchers with a detailed yet user-friendly tool to explore tumor characteristics, treatment outcomes, and patient demographics in a single platform.

Future Improvements and Enhancements

1. **Scalability**:
   - Incorporate additional datasets to expand the scope of analysis, including data on genetic markers or comorbidities.
   - Enable the dashboard to handle real-time data updates for live monitoring.
2. **Predictive Analytics**:
   - Integrate machine learning models to predict treatment outcomes and survival probabilities based on patient and tumor characteristics.
   - Provide risk scores for recurrence to support clinical decision-making.
3. **Advanced Features**:
   - Introduce more granular filtering options, such as specific tumor grades or combinations of treatments.
   - Include a heatmap to visualize tumor location distributions across different demographics.
   - Include a survival curve to illustrate the connection between a population's survival probability and elapsed time.
4. **Addressing Limitations**:
   - Improve data consistency by sourcing larger and more diverse datasets.
   - Incorporate data validation steps to ensure the accuracy and reliability of insights.

Conclusion

This project successfully transitioned from an overambitious detection and classification model to a focused visualization dashboard. By leveraging interactive charts and statistical analyses, the dashboard enables stakeholders to understand complex brain tumor data effectively, facilitating better clinical and research outcomes.

Team Reflection:

This experience underscored the importance of adaptability and collaboration. Despite initial setbacks, the team effectively pivoted to a new approach, resulting in a deliverable that meets stakeholder needs. Key takeaways include:

- The value of stakeholder consultation in refining project goals.
- The critical role of preprocessing and visualization in presenting complex data.
- Managing and working with a team remotely, in different time zones, with different backgrounds, technical skills, and experience.
- Accepting failure effectively and pivoting efficiently.

The team is proud of the outcome and looks forward to future opportunities to build upon this foundation.

Reference:

[1] Dessai, Prathamesh Pradeep. “Brain Tumor Stage-Based Recurrence Patterns.” *Kaggle*, 22 May 2024, www.kaggle.com/datasets/thegoanpanda/brain-tumor-stage-based-recurrence-patterns.

[2] Hong, Jialin, et al. “3D FRN-ResNet: An Automated Major Depressive Disorder Structural Magnetic Resonance Imaging Data Identification Framework.” *Frontiers*, Frontiers, 19 Apr. 2022, www.frontiersin.org/journals/aging-neuroscience/articles/10.3389/fnagi.2022.912283/full

[3] “Image generated by Microsoft Copilot illustrating medical tech and data science.”

[4] “Image generated by Microsoft Copilot illustrating the concept of realizing the difficulty of a goal during implementation.”

[5] Nickparvar, Masoud. “Brain Tumor MRI Dataset.” *Kaggle*, 24 Sept. 2021, www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset/code.