A Brain Metastasis Study with Machine Learning

The team

- Adrián Aguado García
- Arturo Linares Muñiz
- Mateo Rodríguez Suárez
- Olumayowa Onabanjo
- Raquel Martínez Martíne

What is brain metastasis?

It happens when cancer cells travel from other parts of the body and form tumors in the brain. It is currently one of the most complicated forms of cancer to treat. But there is hope, early detection and effective treatment go a long way to have better odds.

Our goals

We sought to resolve one of the most important questions patients have: "*How much time do I have left to enjoy life*?". Other goals such as helping doctors predict when a person is likely to relapse, and effective treatment methods were also important.

What's our dataset like?

We worked on the MOLAB Brain Metastasis Dataset, a structured and anonymized dataset covering segmented Magnetic Resonance Imaging (MRI) scans, clinical data and morphological measurements of 75 patients.

There are 3 groups of files with structured and unstructured data.

- *Clinical data*: an excel spreadsheet with patient data and treatment history.
- *Morphological measures*: an excel file with tumor dimensions and MRI info.
- The *MRI scans* are in formats such as nifti and DICOM.

📁 Clinical data

Variable	What it really means
Patient Data	Section includes data such as age at 1st scan and sex
Lesion	Tumors numbered in the order they were discovered
Primary Tumor	Numbered 1 - 12 with subtype for breast & lung cancer
Whole Brain Radiation Therapy (WBRT)	Includes dosage and period administered
Stereotactic Radiosurgery (SRS)	A non-surgical radiation therapy.
Radiation Necrosis	Tissue breakdown due to radiation
Surgery	Describes type (full, partial, unknown) and time
Systemic Treatment	Drugs given and the time period
Death	Measured in days. Cause is also given.
MRI Follow up dates	Minimum 4 appointments, Maximum 19

🗒️ Morphological measures

Variable	What it really means
General data	Patient ID number, info about image and equipment used
Other measurement data (Don't worry too much about this unless you're an MRI expert)	• SBS (Space between Slices) in millimeters • SCS (Slice Thickness) in millimeters • REPTIME (Repetition Time) in milliseconds • CEVOLUME (Contrast-Enhancing Volume) VCE • NECVOLUME (Necrotic (Non-Enhancing) Volume) Vn • TOTALVOLUME V=VCE+VN • CERIMWIDTH (Contrast-Enhancing (CE) Spherical Rim Width): based on CE areas' average width. • SURFACEREGULARITY: 0 to 1 for roundness.

	• MAXDIAMETER3D: tumor's max. diameter.

🧠MRI scans

The dataset came with images that had already been processed (Figure 1). One of the benefits of working with this dataset were the clearly annotated images with dimensions**.**

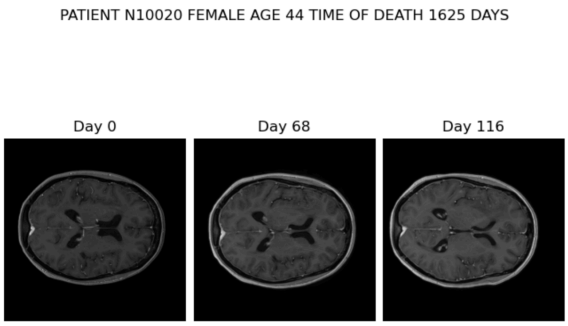


Figure 1: Sample MRI scan over time period

Our strategy

We'll run you through how we explored our data, decided on a Machine Learning (ML) approach and our results.

Data preparation (pre-processing & cleaning)

Clinical data and morphological measures were structured differently in separate files. Organizing the data in one dataset made it easier for our models to understand. Also there were empty cells and data format issues to resolve.

Data analysis

A whole lot of plots, graphs and code were needed to understand the relationship between our variables and know what we could predict with the most certainty. While working our data, we were made aware of factors that could affect the accuracy of our model, such as size, class imbalance and outliers (Figure 2 & 3).

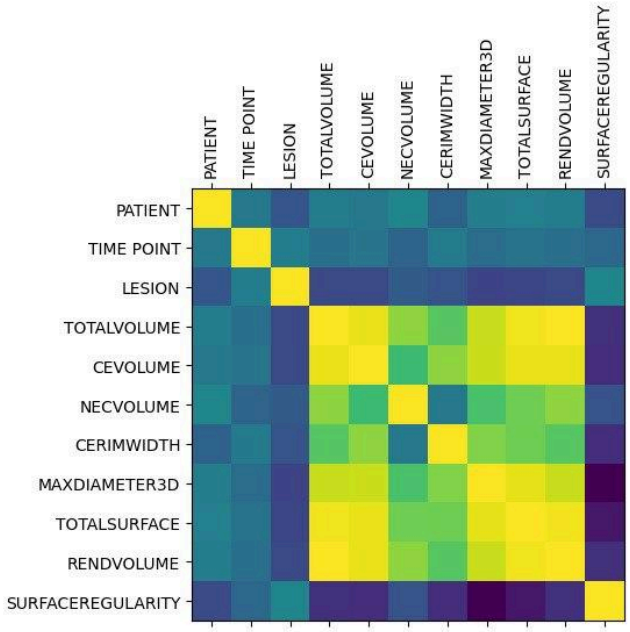


Figure 2: Correlation matrix for Morphological Measures. High correlation between variables derived from each other. Aided selection of most pertinent variables for model training.

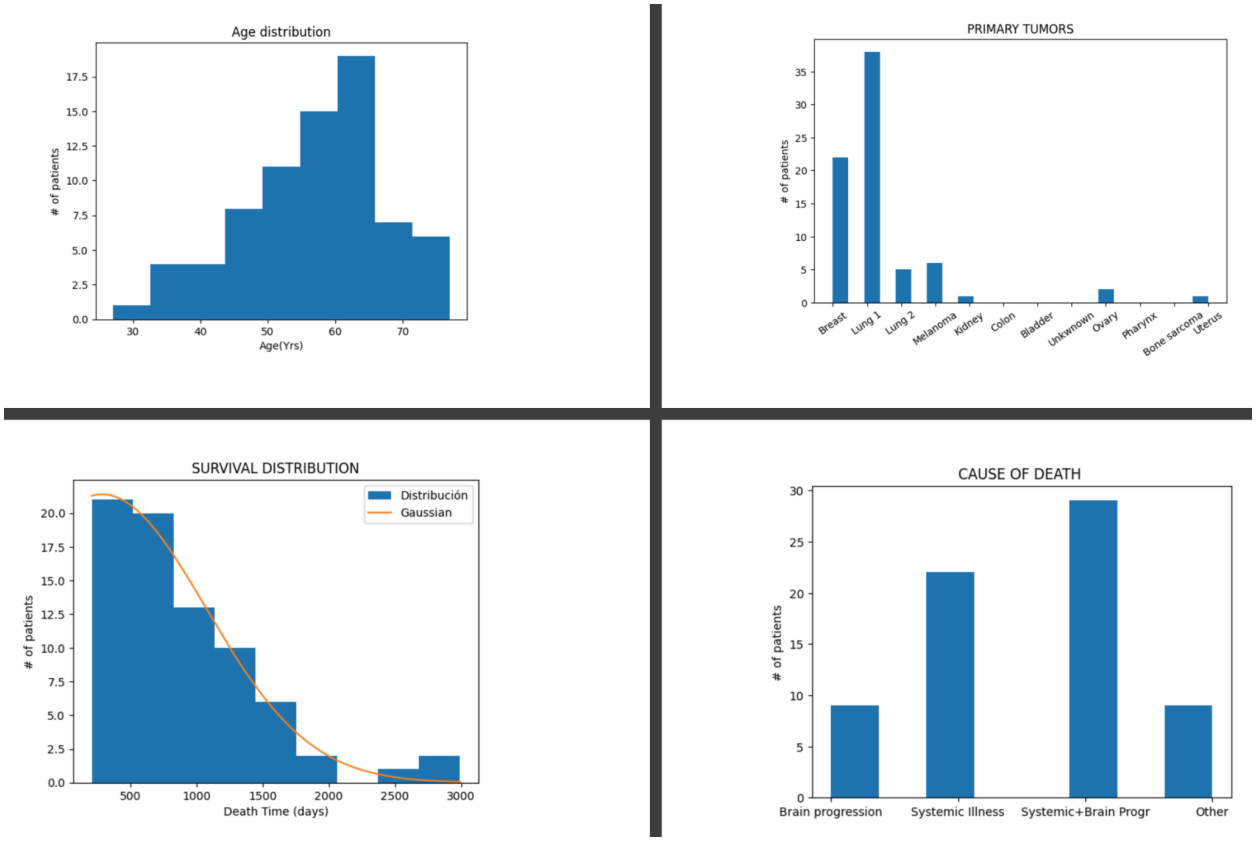


Figure 3: Clinical Data Distribution (Age, primary tumor, survival time, and cause of death).

Prediction approach

It was challenging selecting which level to apply our model (patient, tumor, evolution over time or a combination). Figure 4 is an example for lesion volume. We saw it would be difficult to do a time series prediction as tumor evolution varied widely.

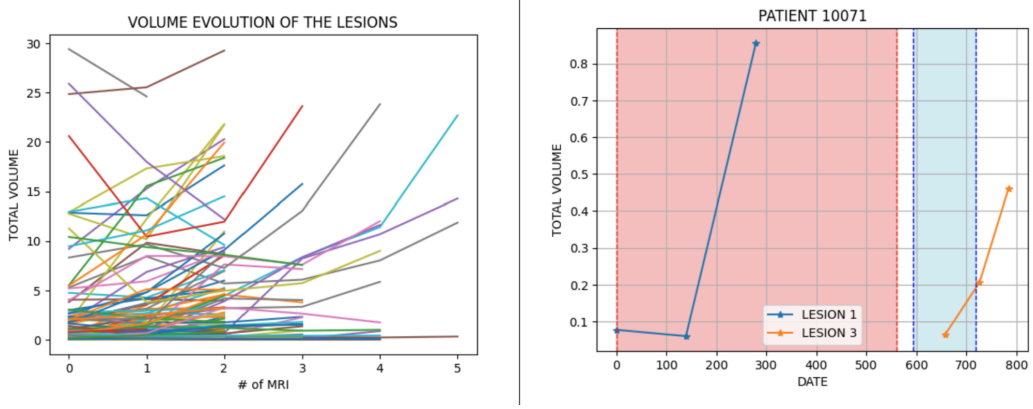


Figure 4: Temporal evolution of tumor total volume viewed by tumor and by patient

Model Selection

Our small dataset size limited model selection. Figure 5 shows the main ML method groups with the ones we explored highlighted in red (Decision trees, Logistic Regression, Neural Networks, Random Forests, and XGBoost).

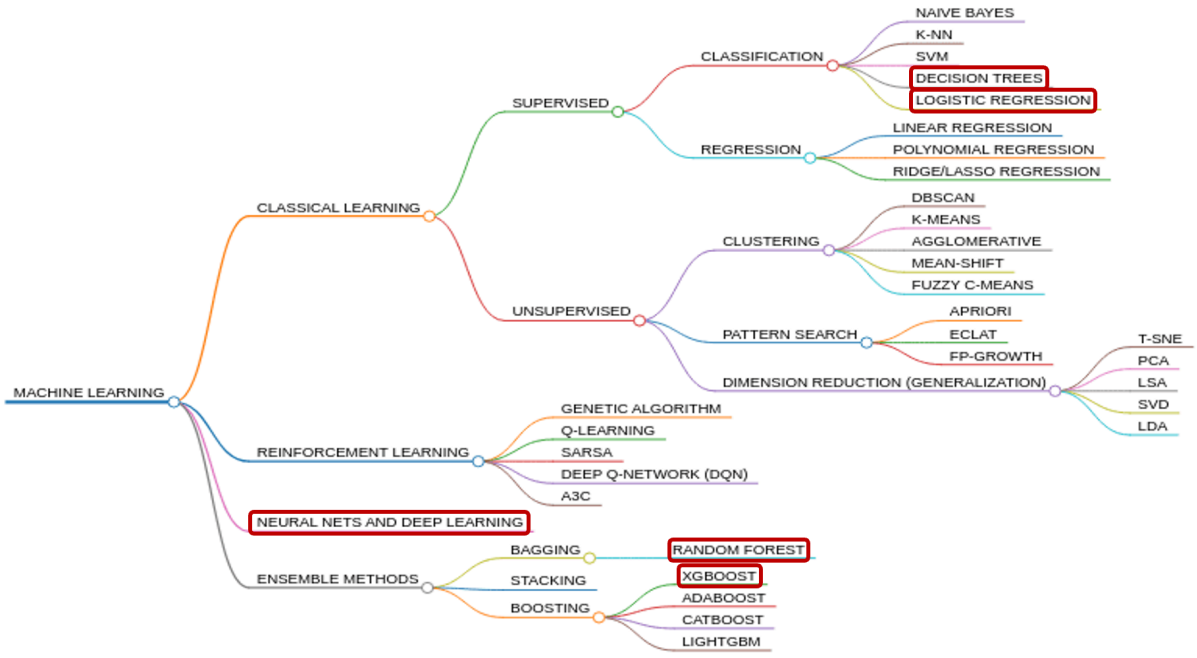


Figure 5: Machine learning methods concept map showing selected models from the main ML method groups (Classical, Reinforcement, Neural Networks/Deep Learning & Ensemble methods)

Model training

Our main objectives were translated into variables for our models to predict.

- **Objective 1**: Predict survival time in days
- **Objective 2a**: Predict lesion relapse (for all patients)
- **Objective 2b**: Predict lesion relapse (only patients who had radiosurgery)

We chose a binary and multiclass approach to compare model performance. Various model parameters were used to arrive at optimal results.

Binary Classes

Objective 1: Survival time

- Survival <500 days
- 500 days or more

Objective 2a & 2b: Relapse (all patients & patients with radiosurgery)

- Relapse: tumor growth > 30%
- No relapse: tumor growth < 30%

Multiclasses

Objective 1: Survival time

- Survival <500 days
- >= 500 or <1500 days
- >= 1500 days

Objective 2a & 2b: Relapse (all patients & patients with radiosurgery)

- Relapse: tumor growth > 30%
- No relapse: tumor growth 0 - 30%
- Improvement/tumor shrinkage: tumor growth < 0%

Input data summary

Objective	Size	Input	Predicted variable
1	75 rows 5 columns	sex (categorical) age (numerical discrete) gpa (numerical discrete) nlesions (categorical ordinal) type (categorical ordinal)	survival days (numerical discrete)
2a & 2b	345 rows (2a) 105 rows (2b) 13 columns	sex (categorical) age (numerical discrete) gpa (numerical discrete) cermwidth (numerical continuous) surfaceregularity (numerical continuous) totalvolume (numerical continuous) srs (numerical continuous) days_since_srs (numerical discrete) wbrt (numerical continuous) days_since_wbrt (numerical discrete)	relapse (categorical boolean)

Results

We summarized results of different trials. More information is available on GitHub.

Objective 1 (Predict Survival Time) - Model Training Results (Binary)

Model	Accuracy	AUC*	Comments
Logistic regression	71	-	Seeds: 10000
Decision Tree	72	67	Tree depth of 4 performed best
Random Forest	73	-	Estimators: 15, Iterations: 5000
XGBoost	73	75	Estimators: 4, depth: 4, iterations: 10000

Neural networks	75	-	Epochs: 5, Seeds: 1000 (higher uncertainty than previous trials)

*AUC (**A**rea **U**nder ROC **C**urve)

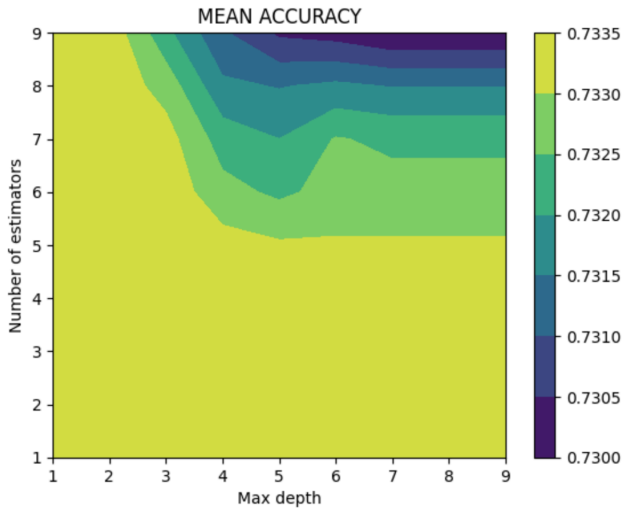


Figure 6: Selection of parameters for best performing model (XGBoost)

Objective 1 (Predict Survival Time) - Model Training Results (Multiclass)

Objective	Class	Results
1 (survival time)	Multiclass	Model accuracy varied between 35 - 46%. XGBoost showed the best performance

Objective 2a Summary Predict lesion relapse (all patients)

Objective	Class	Results
2a (relapse, all patients)	Binary	Model accuracy varied between 47 - 58%. Decision trees showed the best performance
2a (relapse, all patients)	Multiclass	Model accuracy varied between 43 - 50%. Best performance with XGBoost.

Objective 2b Summary Predict lesion relapse (patients with radiosurgery)

Objective	Class	Results
2b (relapse, patients with srs)	Binary	Model accuracy varied between 52 - 59%. Best performance with Logistic Regression.
2b (relapse, patients with srs)	Multiclass	Model accuracy varied between 33 - 41%. XGBoost showed the best performance

Conclusion

Our model performed best for predicting survival days, we expect results can be improved with more data. Predicting relapse gave poorer results given higher variance. Also there was generally poorer model performance on multiclass compared to binary classifications. Model performance depended on the task, emphasizing the importance of experimentation with several models to compare results.

We recognize that there are factors that impact both outcomes that are not captured in our dataset. Still it serves as a good starting point for similar studies on diseases with relatively small populations.

We must add that these are promising results. With only 75 patients we have been able to predict with an accuracy of 75% their survival time. In the case of predicting relapses, it is obvious that we are just beginning and accuracy can be improved with additional data and more details. We also expect advanced approaches for model adjustment to yield more favorable outcomes.

Explainable ML

Our model is designed to make predictions that are critical to people's lives and their decisions for the future. So we made sure it would be transparent both to us and patients. To that end, we've added the SHAP (SHapely Additive exPlanations) values,

links to the data and our code. Why? Because we know it's harder to trust AI when it's shrouded in mystery.

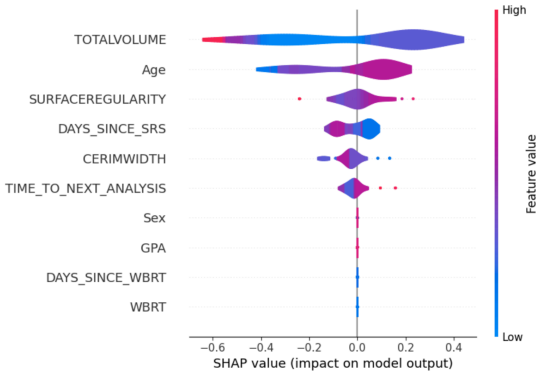
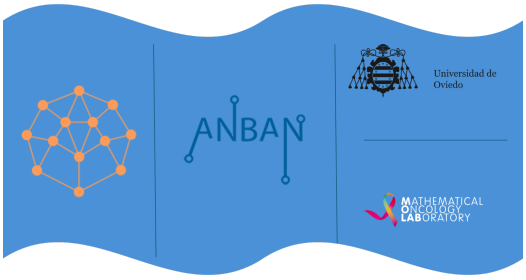


Figure 7: SHAP value (impact on model output)

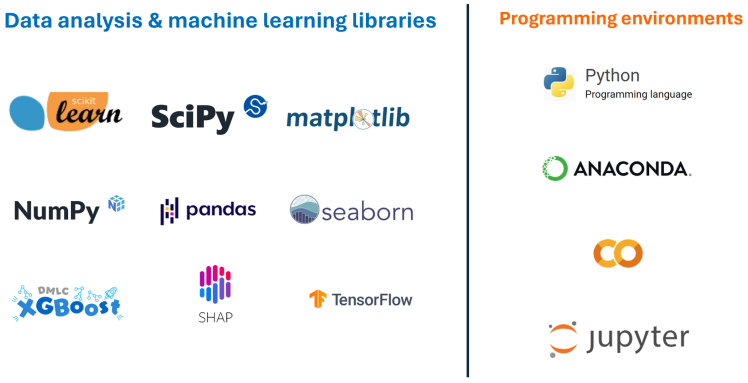
Disclaimer: models are fallible and should NEVER be used without professional advice.

How we did it

With teamwork and a lot of help from our coordinator, instructors, mentors, the organizers of SaturdaysAI Asturias, our hosts, and the wonderful people at MOLAB.



Some of us knew little about Python and in just 12 weeks together, we trained ML models. Hopefully, our story helps you on your journey to making lives better with artificial intelligence. We look forward to your feedback and comments on our ML project.



We acknowledge ideation and visualization support from Generative AI - ChatGPT, Bing Copilot, and DALL-E.

Next steps

For anyone looking to join our fight against brain metastases, here are some avenues we would have liked to explore.

- We discovered the cause of death was a variable that could be predicted during our data exploration. However, we were not able to advance this option due to time constraints.
- Varying the threshold for percentage tumor growth to define as a relapse. We used 30% growth as our criteria.
- Augment the dataset size with Generative Adversarial Networks.
- We did not achieve better results by separating patients that had surgery from the rest. Notwithstanding, it is worth exploring whether splitting the dataset by other criteria will yield better results.

Resources

- MOLAB
- MOLAB brain metastasis dataset
- Brainlab.org
- Brain metastases: A documentary