



OPEN Systematic review and epistemic meta-analysis to advance binomial AI-radiomics integration for predicting high-grade glioma progression and enhancing patient management

María Fátima Chilaca-Rosas^{1,4}, Manuel Tadeo Contreras-Aguilar^{1,4}, Federico Pallach-Loose^{2,4}, Nelly F. Altamirano-Bustamante^{3,4}, David Rafael Salazar-Calderon^{1,4}, Cristina Revilla-Monsalve², Juan Carlos Heredia-Gutiérrez¹, Benjamin Conde-Castro¹, Rafael Medrano-Guzmán¹ & Myriam M. Altamirano-Bustamante^{2,4}✉

High-grade gliomas, particularly glioblastoma (MeSH:Glioblastoma), are among the most aggressive and lethal central nervous system tumors, necessitating advanced diagnostic and prognostic strategies. This systematic review and epistemic meta-analysis explore the integration of Artificial Intelligence (AI) and Radiomics Inter-field (AIRI) to enhance predictive modeling for tumor progression. A comprehensive literature search identified 19 high-quality studies, which were analyzed to evaluate radiomic features and machine learning models in predicting overall survival (OS) and progression-free survival (PFS). Key findings highlight the predictive strength of specific MRI-derived radiomic features such as log-filter and Gabor textures and the superior performance of Support Vector Machines (SVM) and Random Forest (RF) models, achieving high accuracy and AUC scores (e.g., 98% AUC and 98.7% accuracy for OS). This research demonstrates the current state of the AIRI field and shows that current articles report their results with different performance indicators and metrics, making outcomes heterogenous and hard to integrate knowledge. Additionally, it was explored that today some articles use biased methodologies. This study proposes a structured AIRI development roadmap and guidelines, to avoid bias and make results comparable, emphasizing standardized feature extraction and AI model training to improve reproducibility across clinical settings. By advancing precision medicine, AIRI integration has the potential to refine clinical decision-making and enhance patient outcomes.

Keywords Artificial Intelligence, Glioblastomas, Machine learning, Radiomics, Artificial Intelligence Radiomics Interfield (AIRI), Segmentation, Progression, Prognosis, Overall survival (OS), Progression-free survival (PFS), Roadmap, Mind map

Gliomas are the most common malignant tumors of the central nervous system (CNS), originating from glial cells such as astrocytes, oligodendrocytes, and ependymal cells^{1,2}. Among these, glioblastomas represent the most aggressive subtype, with high recurrence rates and poor overall survival (OS). Despite advancements in treatment strategies, including surgery, radiation, and chemotherapy, the prognosis for glioblastoma patients remains dire,

¹Hospital de Oncología, Centro Médico Nacional Siglo XXI, Instituto Mexicano del Seguro Social, 06720 Mexico City, Mexico. ²Unidad de Investigación en Enfermedades Metabólicas, Centro Médico Nacional Siglo XXI, Instituto Mexicano del Seguro Social, 06720 Mexico City, Mexico. ³Servicio de Endocrinología, Instituto Nacional de Pediatría, 04530 Mexico City, Mexico. ⁴María Fátima Chilaca-Rosas, Manuel Tadeo Contreras-Aguilar, Federico PallachLoose, Nelly F. Altamirano-Bustamante, David Rafael Salazar-Calderon and Myriam M. Altamirano-Bustamante contributed equally to this work. ✉email: myriamab@unam.mx

with a median OS of approximately 14.6 month^{3,4}. The ability to accurately assess tumor progression and predict patient outcomes is essential for optimizing therapeutic approaches and advancing precision medicine.

Medical imaging plays a pivotal role in the diagnosis and management of gliomas, with magnetic resonance imaging (MRI) considered the gold standard for tumor visualization and treatment response assessment^{4,5}. Advanced imaging techniques, including perfusion-weighted imaging, diffusion-weighted imaging, and positron emission tomography (PET), have further enhanced our understanding of tumor biology². However, conventional radiological assessments often fail to capture the full complexity of gliomas, necessitating more advanced computational approaches.

Radiomics, an emerging field within precision oncology, leverages machine learning (ML) and artificial intelligence (AI) to extract high-dimensional qualitative features from medical images⁶. These features capture tumor heterogeneity, spatial distribution, and texture patterns that may not be discernible through traditional imaging analysis⁷. When integrated with clinical and molecular data, radiomics models can improve prognostic accuracy, facilitate early tumor progression detection, and support personalized treatment planning.

Despite the potential of AI-driven radiomics, challenges remain in standardized feature extraction, validating predictive models, and ensuring their reproducibility across diverse patient populations. The lack of consensus regarding optimal imaging protocols, segmentation methodologies, and model interpretability has hindered the widespread clinical implementation of these techniques. Addressing these limitations requires a comprehensive evaluation of existing AI-radiomics methodologies and a structured approach to model development and validation.

This study conducts a systematic review and epistemic meta-analysis to assess the current state of AI-radiomics integration in glioma prognosis. By analyzing methodologies, predictive features, and machine learning algorithms employed in previous research, we aim to identify best practices and propose a standardized roadmap for developing robust and clinically applicable AI-Radiomics models. Through this approach, we aim to bridge the gap between research and clinical application, ensuring that AI-Radiomics models are not only precise and reliable but also seamlessly integrable into existing glioma prognosis workflows.

Materials and methods

Article scaffolding

The global structure of the article is the following: The first step was formulating the research question, and the research objectives followed by separating the question into its components according to the PIO (Participants/Intervention/Outcome) strategy. This was continued by searching the keywords and MeSH (Medical Subject Headings) terms in three databases following PRISMA AND PICO approach^{8–10}. Secondly, the resulting articles were assessed and reviewed. Thirdly, a epistemic meta-analysis was performed.

Search strategies

A systematic review was conducted until November 2023 using a modified PIO (Participants/Intervention/Outcome) approach, adhering to the Principles for Systematic Reviews and Meta-Analyses (PRISMA) guidelines to validate searches across various databases^{8–10}.

PIO strategy

The search strategy was developed using MeSH terms and Boolean operators to address the research question. The primary objective was to explore the latest findings on the application of AI, particularly machine learning (ML), in predicting high-grade glioma progression using a radiomics approach. The PIO strategy encompassed Participants, Interventions, and Outcomes, focusing on the following research question: Which radiomic features are associated with progression in patients with high-grade gliomas assessed by MRI using AI methods such as ML?

The databases used were PubMed, BIREME, and Web of Science with the following configuration:

Participants Articles related to high-grade gliomas diagnosed by histopathology and MRI studies were included. Relevant terms included: Glioma, high-grade glioma, brain tumor, tumor microenvironment, magnetic resonance image, MRI, texture analysis, semiautomated, semiautomated analysis, semiautomated platform, biomarkers, precision medicine, and virtual biopsy.

Intervention Articles involving software and algorithms used to extract features from images were included, with terms such as: artificial intelligence, machine learning, Adaboost, Xgboost, Lightgbm, gaussian process, decision trees, gradient boosted decision trees, neural Networks, k-nearest neighbor, deep convolutional neural network, logistic regression, partial least squares discriminant analysis, quadratic discriminant analysis, random forest, stochastic gradient descent, support vector classification, fully connected network, Lifex, PyRadiomics, Olea sphere, 3DSlicer, Python, Python PyTorch, supervised learning and unsupervised learning.

Outcome Articles relating first and second-order radiomic features to the prognosis of high-grade gliomas were included.

First-order features Properties of individual voxels without considering their spatial relationships, such as intensity, histogram, conventional, and discretized.

Second-order features Analyzing the texture and spatial relationships between voxels within the region of interest:

- Gray Level Co-occurrence Matrix (GLCM)
- Gray Level Run Length Matrix (GLRLM)
- Gray Level Size Zone Matrix (GLSZM)
- Neighboring Gray-Tone Difference Matrix (NGTDM)
- Gray Level Dependence Matrix (GLDM)

Shape and volume Geometric properties of the ROI.

- Sphericity
- Compactness
- Elongation
- Surface area
- Size of ROI (volume)

Terms related to progression

- Progression
- Progression disease

To define the proportion of semiautomated or automated platforms used, we assessed the tumoral radiomic features in high-grade gliomas and their association with the tumor microenvironment (edema). Additionally, we evaluated the models proposed to predict progression using radiomic analysis through AI.

The following questions were proposed to guide the study; what proportion of semi-automated platforms assess radiomic features in high-grade gliomas? What percentage of automated platforms assess radiomics features in high-grade gliomas? Which radiomic features have been associated with the tumor microenvironment in high-grade gliomas? and, what are the main methods of modeling and validation of proposed AI-based algorithms?

Databases and searches

The electronic databases used were PubMed, BIREME, and Web of Science. The search strategy was replicated for each database using the combination of several Booleans appropriated for our research question (Fig. 2). This figure describes in full detail the search strategy in the PubMed database as well as all keywords used. Subsequently, the results obtained from these searches were recorded in the Mendeley database.

IA application to predict the progression of disease

This systematic review followed the PRISMA strategy to assess the articles pertinent to the research question.

Quality control assessment

A quality control has been carried out taking into consideration the methodological aspects of the articles obtained for the analysis.

We used the criteria assessment proposed by Juarez et al.⁷ that briefly encompasses:

1. Research questions and objectives. A clear question, a clear objective.
2. Methodology: Complete, robust, and adequate; Suitable type of study, suitable instrument, reproducibility, and bias control.
3. Results: Congruent with the objective, congruent to the question.
4. Terms: Properly defined.
5. Conclusions: That answers the question and achieves objectives.

The quality control score ranges from 0 to 100, with articles scoring 75 or higher being accepted for the meta-analysis. A total of 19 articles were assessed for quality, resulting in their inclusion in the final meta-analysis.

Integral epistemic meta-analysis: exploring the application of AI in the progression of high-grade gliomas

Epistemic meta-analysis is an approach used in research to analyze and synthesize existing knowledge, theories, or findings. On this research, the term ‘epistemic’ refers to the examination of knowledge or understanding by evaluating the underlying assumptions, methodologies, and theoretical frameworks of AIRI studies. Unlike a traditional meta-analysis which is only quantitative and requires data that is homogeneous, consistent and unified between studies, an epistemic meta-analysis allows to gather knowledge from articles that are not easily comparable, which is the case of the AIRI field to this date because each paper reports results with different metrics and the studies had to be divided into four topic groups to be comparable. The precedent situation requires to first create a puzzled construction by a epistemic meta-analysis, which is the goal of this study. Additionally, this approach aims to provide a deeper understanding of the current state of knowledge within the AIRI glioma domain, highlighting its strengths, limitations, and areas for further investigation. The objective is to create a roadmap and guidelines that help to homogenize and integrate the results of the field, so that in the future, the field can be more robust for traditional meta-analysis to be viable.

This study was conducted by a cross-functional group consisting of medical doctors, oncologists, radiomics experts, a physical engineer with experience in AI, and bioinformaticians.

Integral AI meta-analysis algorithm (IAIMA)

The 19 studies selected in the systematic review that passed the quality control assessment were explored with an IAIM. The Integral AI meta-analysis algorithm is a process where the common elements of the article scaffold are identified, analyzed, and classified. It is a process that encompasses 5 steps. First, the studies were classified by application topics. This classification is shown in Fig. 4. The topics are prognosis of OS or PFS, progression after treatment, segmentation, and other stratifications. Second, all the relevant data (clinical, radiomics, processes, components, etc.) for building the model was registered in mind maps, starting from the common components

of the scaffold and identifying the variable components (Fig. 2). Third, the data gathered led to building two organized tables consisting of what was done in the studies (Table 2) and how it was done (Table 3). Fourth, the studies were analyzed by two methods (mixed method, and by application topic) based on Tables 2 and 3. Fifth, a proposal of a pipeline to follow for the development of an enhanced model for the main application topics that could be implemented in hospitals was done (Fig. 8).

This approach made it possible to compare the studies in relatively equal terms. This meta-analysis served the purpose of gathering the key elements for building an AI radiomics model for the prognosis of OS, prognosis of PFS, and progression after treatment. The combination of the best scaffolding of each article is the starting point to make a pipeline for building a model that would help oncologists and doctors perform a kind of triage for patients' care centers.

Expanded algorithm of the meta-analysis

To visualize all the processes, we describe each step mentioned above.

Application topic classification

The studies were classified by application topic. The classification of topics helps visualize the different models that can be developed to help the innovation and treatment of cancer patients with high-grade gliomas. This was done by a thorough reading of the papers. Some studies had two topics, so a Venn diagram was used to represent the classification. There are 4 topics: prognosis of OS or PFS, progression after treatment, segmentation, and other classifications. The prognosis of OS or PFS means that the article's model stratifies patients regarding their prognosis of OS or PFS. The topic of progression after treatment consists either of articles that diagnose the patients' progression condition (pseudo-progression or true progression) some weeks after the treatment, or articles that predict the future progression areas outside the tumor from the pre-operative MRI sequences. Segmentation articles focus only on the division of the regions of the tumor and report that performance. The classification topic means that a model from the article classified something else than prognosis, progression after treatment, or segmentation. These articles classified either the REP condition, the GBM molecular subtype, or the molecular biomarkers, glioma grade, and identified progression from no progression.

Studies' roadmaps

The steps of the roadmap of each article were identified as computable objects. This means that the input of the system is complex and integrated with clinical, molecular, and image data. Followed by pre-processing, segmentation, and feature extraction of the images that are then converted into numerical values named radiomic features. The roadmap is the series of steps that each study did to develop their model. According to the scaffolding of the articles, a general roadmap of what most studies did was identified.

Mind maps

The mind maps served as a tool for summarizing the key information of the development process of each model and its performance. The data of both the best-performing model and additional models built on each article were identified, registered, and classified. A baseline scaffolding-template mind map was used for all the studies to standardize both the structure and the data summarized in this step (Fig. 1). The mind maps allow a deep understanding of each of the nineteen articles in a glimpse, and they are also effective for looking up and overlapping specific and relevant data of the study. They are also helpful in identifying the elements of the complete model development process.

The mind maps consist of two parts: what was done (right), and how it was done (left). The right side has the main information of the article, while the left side has the information that supports the results of the article, it summarizes how it was done (Fig. 1).

In Fig. 1, the right side has the following branches: (i) identifying the application topic or type of article, (ii) the objective, (iii) the cancer type, (iv) the model and its performance, and (v) additional specific results of the article. (i) The applications of binomial AIRI of the articles have four main categories: prognosis of OS or PFS, progression after treatment, segmentation, and other classifications. Some articles have two application topics. (ii) The objective of the article is the purpose of why it was written. (iii) The cancer type was divided into three: GBM, LGG, and DMG. (iv) The performance of the model was reported with different variables, depending on which variables were used in the article. (v) Some articles performed additional studies that diagnose: REP condition, GBM molecular subtype, molecular biomarkers, or glioma grade.

In Fig. 1 the left side, which consists of how the study was carried on, has two proper branches, which are (i) data and (ii) approach, and one branch that is common with the right side, (iii) the model branch. The following elements were included on the (i) data branch: imaging types, number of patients, source, availability of the parameters used for the MRI sequences, and the pre-processing information. The (ii) approach branch has two sub-branches: features and segmentation. The features sub-branch has registered the number and type of features used, the feature selection process, and the software used to extract the radiomic features. The segmentation sub-branch has the segmented regions, the level of automatization for this process, and the software used.

AI Table: information for the creation of the model

The information gathered in the mind maps was then organized on a table. It was placed in such a way that it gives the possibility to compare any article with the others on any aspect. The data of the best-performing model of each application topic was the one registered on Tables 2 and 3. In Chilaca's study, rather than training an AIRI model, the researchers focused on evaluating the predictability of individual features without creating a comprehensive model. This alternative approach is particularly appealing for the oncologist, for it emphasizes the relevance of each feature. The features' performances reported for that study on Table 2 were CONV_RIM_

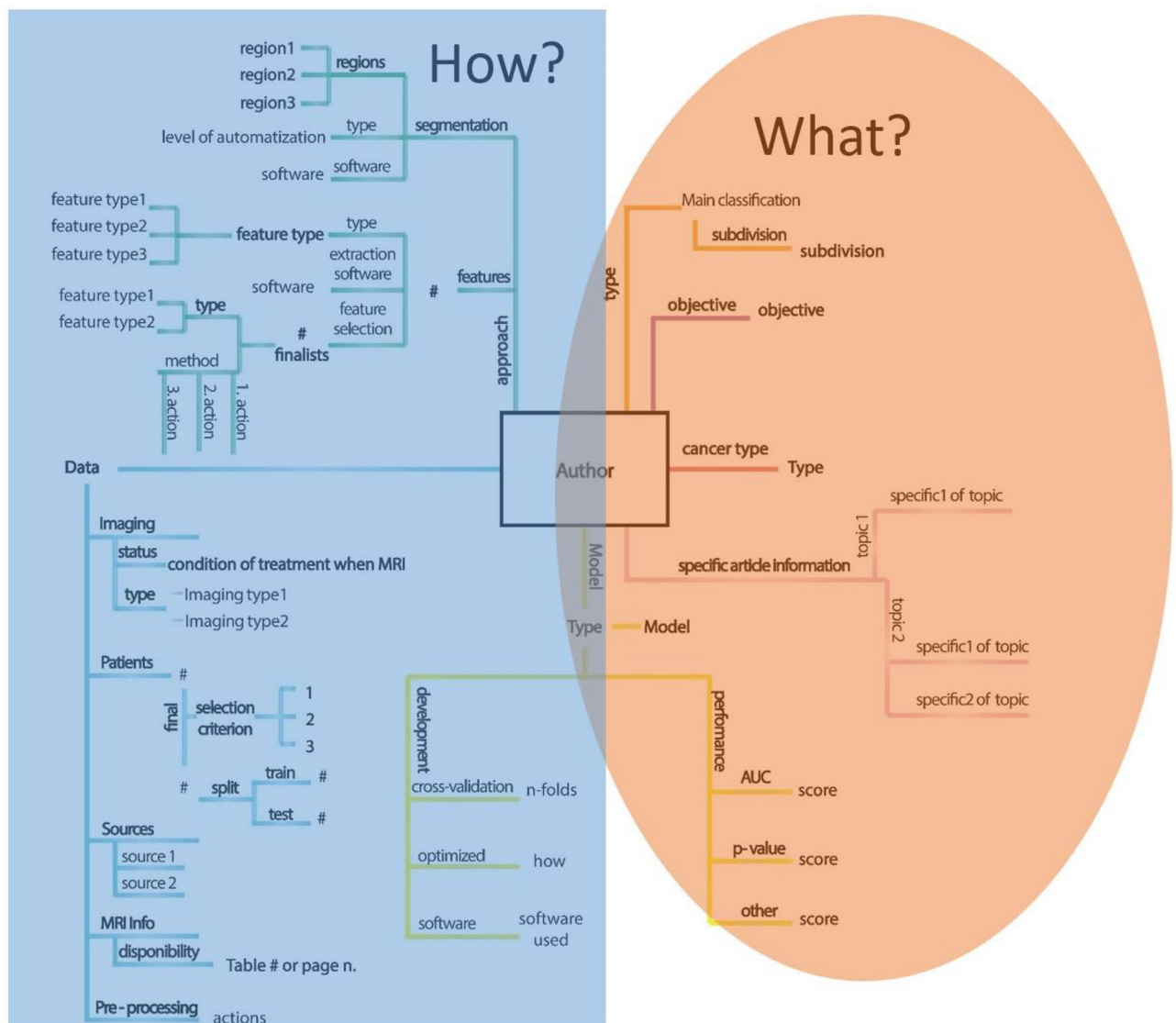


Fig. 1. Base-line mind map for step 2 of the meta-analysis.

stdev for OS and GLCM contrVar for PFS¹². The AI table has six subheadings, which are: Main Information, Model Information, Model Performance, Segmentation, Features, and Data. The AI table is divided into Tables 2 and 3. The division criterion is the same as on the mind maps (what, how).

Table 2 has three headings; main information, model information, and model performance. The main information heading has general information about the study, the topics covered in it, the clinical data, and the study's objective. The model information subheading has the specifications of the machine learning model, it consists of 7 columns. The model performance heading contains how well the model did and it has both the training and testing performances registered. The performance variables used were AUC, accuracy, sensitivity, specificity, and others (e.x. *p* values). If any of these variables was not reported, it was specified as blank (-).

Table 3 has three headings as well: segmentation, features, and data. The segmentation heading consists of segmented regions, the number of regions, the level of automatization, and the software used for the segmentation process. The features subheading has 7 columns. It registers which features were used, with which software they were extracted, how the feature selection process was done, and which ones were the most predictive features. Last of all, the data branch has the MRI sequences used, the status of the patient's treatment upon imaging, the number of patients, the split of training and testing, and the sources of the data.

Model comparison

Two types of analysis were carried on, a mixed method analysis (quantitative and qualitative), and an analysis by application topic. The mixed method consisted of analyzing together all the articles on a specific aspect. There were three aspects analyzed: clinical patient data, radiomic features and segmentation, and ML models. This led to finding the different proportions that answer some of the research questions. On the other hand, the analysis by application topic consisted of first grouping the articles in the topics previously identified in step 1 of the

meta-analysis and then comparing the articles within each application topic separately. This analysis made it possible to compare the performance of the models against each other. The application topics analyzed were two; prognosis of OS or PFS, and diagnosis progression after treatment. The AI table served as the source material for making the comparisons. The models with the highest performance scores were identified and highlighted. The features from these models were considered to be the most predictive ones, taking into account that the features used to train a model are the key elements for the model's performance. Furthermore, the best algorithm to use and a robust methodology to develop a model were explored.

Model pipeline proposal

A proposal for a pipeline to follow to develop an AIRI prognosis model and its implementation in a clinical setting was done. The development approach combines the methods, features, and ML algorithms used on the best-performing models. The development proposal is a roadmap with eight steps with specific suggestions to follow.

Results

State of the art of progression in high-grade gliomas using AIRI

The research question of this systematic review: Which are the radiomic features associated with progression in patients with high-grade glial tumors assessed by MRI using AI methods such as ML. The PIO dissected the research question in these parts: Participants with histopathology diagnosis of high-grade glioma; Intervention assessed with an MRI analysis by semiautomated radiomic platform; Outcome features associated with progression disease. We can follow the PIO strategy in Fig. 2.

PRISMA assessment results

This systematic review began with 210 references obtained from three electronic databases: 79 were from PubMed, 69 were from BIREME, and 63 were from Web of Science. In the first screening, double references were excluded, removing 105 articles.

A second screening (eligibility) was performed using the following set of quality criteria:

1. Original articles (not review articles, letters, abstracts, or viewpoints).
2. Published in 10 years.
3. Written in English.

These criteria were applied, leaving with 70 articles.

A third screening was conducted to exclude articles that were irrelevant to the review, resulting in 51 exclusions and leaving 19 articles for the review. See Fig. 3 and Table 1.

Quality control results

The results of the quality control are described in Table 1, which integrates all the data on the criteria used to qualify the methodology of the studies included for the systematic review.

Meta-analysis results: AIRI binomial as a route to improve diagnosis of patients' progression

Application topics of studies

As for the application topics of the nineteen selected articles, eleven were prognosis of OS or PFS, six were progressions after treatment, and two were segmentation. Three articles whose topic was the prognosis of OS or PFS also did another type of classification. A diagram with the classification of the study's application topics is shown in Fig. 4.

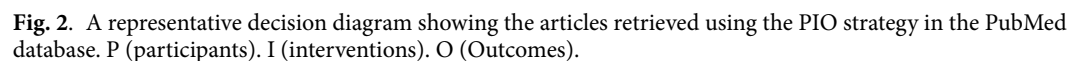
General AIRI roadmap

Most of the articles share common steps in their AIRI roadmaps. First, the patient selection is done, consisting of MRI, clinical, and molecular data acquisition. Then the MRI is pre-processed followed by the segmentation of the tumor, where the regions of interest (ROI) are identified for each image. Afterwards, the radiomic features are extracted from the ROI of the images. Then feature selection is performed, where the most predictive features are chosen (radiomic, clinical, and molecular). This is followed up by training the machine learning or statistical model, where a mathematical model is built with the most predictive features and a particular algorithm. Finally, the performance of the model is assessed for both the patients with whom the model was developed (training set) and for patients who were left out of the model's development process (testing set). Sometimes feature selection, training the model, and assessing its' performance work as a single cyclic step (example: Recursive Feature Elimination (RFE)^{15,21}, Backward Feature Elimination (BFE)¹³). This general AIRI roadmap is shown in Fig. 5.

Mind maps of the AIRI epistemic meta-analysis

The baseline scaffolding-template mind map proved useful to analyze all the articles disregarding the differences in application topics. Despite that, new sub-branches were added or branches were trimmed to adequate the baseline mind map for each article. When no information was found on a specific topic, the letters N.S. (not specified) were used. Figures 6 and 7 show two example mind maps of articles with different application topics.

Chang et al.²⁵ study was taken as an example of the mind map analysis (Fig. 6). On the first visualization, one can see that the right side corresponds to what Chang study did, and the left side about how he did it. Subsequently, the mind map is divided into four quadrants. In the upper right quadrant are all the study characteristics (orange, purple, and pink). In the upper left quadrant are all the clinical features and the process



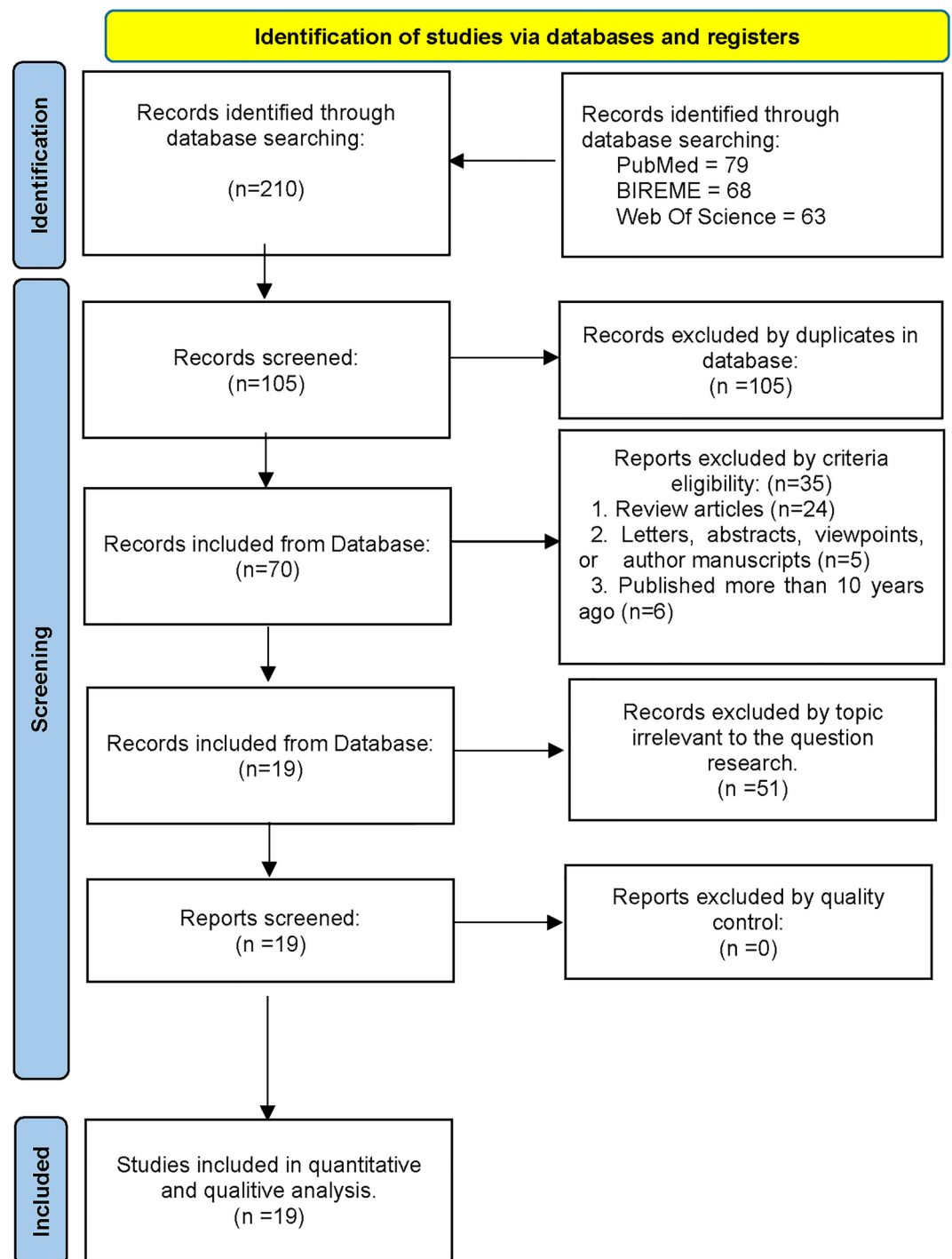


Fig. 3. Identification of studies via databases and registers with the PRISMA approach.

to obtain them (green). In the lower left quadrant is the base data of the model (blue). In the lower right quadrant is the data related to the machine learning algorithm and its' performance (yellow). This mind map structure is common to all studies. Overlapping the minds maps allows to identify on the fly the constant parts of the studies and find their differences based on their application topics.

Patel et al.¹⁵ study was taken as another example of the mind map analysis (Fig. 7). The constant parts of this mental map compared with Chang's study, are: That the right side corresponds to what they did, and the left side to how they did it. The upper right quadrant shares most of the structure, but the content adapts to the application topic. The upper left quadrant stays the same. In the lower right quadrant, both pink branches are of statistical analysis, and the yellow branch basic structure remains relatively constant. The lower left quadrant stays the same. The parts that differ from Chang's study are the type of article on the upper right, the specific approach of the statistical analysis of the pink branch, and the way the model's performance is reported on

Articles		Research questions and objectives			Complete, robust, and adequate methodology					Results		Properly defined terms (20 pts)		Conclusions		Total points
Author, year	Title	Clear question (10 pts)	Clear objective (10 pts)	Suitable type of study? (5 pts)	Suitable instrument? (5 pts)	Is it reproducible? (5 pts)	Bias control (5 pts)	Results congruent with the objectives (10 pts)	Results congruent to question (10 pts)	Conclusion that answers the question (10 pts)	Conclusion that achieves objectives (10 pts)					
Chilaca et al., 2023 ¹²	Diagnostic Performance of Selected MRI-Derived Radiomics Able to Discriminate Progression-Free and Overall Survival in Patients with Midline Glioma and the H3F3AK27M Mutation	10	10	5	5	5	5	10	10	10	10	20	10	10	100	Yes
Farzana et al., 2023 ³²	Prediction of Rapid Early Progression and Survival Risk with Pre-Radiation MRI in WHO Grade 4 Glioma Patients	10	10	5	5	0	5	10	10	10	10	20	10	10	95	Yes
Zhang et al., 2023 ³¹	Diffusion-weighted imaging and arterial spin labelling radiomics features may improve differentiation between radiation-induced brain injury and glioma recurrence	10	10	5	5	5	5	10	10	10	10	20	10	10	100	Yes
Borges et al., 2021 ¹¹	Kurtosis is An MRI Radiomics Feature Predictor of Poor Prognosis in Patients with GBM	10	10	5	5	5	5	10	10	10	10	20	10	10	100	Yes
Su et al., 2021 ¹³	Identification of glioblastoma molecular subtype and prognosis based on deep MRI features	9	9	5	5	4	3	9	9	5	5	15	5	5	78	Yes
Sun et al., 2021 ¹⁴	Differentiation of pseudo progression from true progression in Glioblastoma Patients after Standard Treatment: A Machine Learning Strategy Combined with Radiomics Features from T1-weighted Contrast-enhanced Imaging	10	8	5	5	3	5	7	9	7	5	17	7	5	81	Yes
Patel et al., 2021 ¹⁵	Machine learning-based radiomic evaluation of treatment response prediction in glioblastoma	8	8	5	4	4	5	7	7	6	6	15	6	6	75	Yes
Pak et al., 2021 ¹⁶	Prediction of Prognosis in Glioblastoma Using Radiomics Features of Dynamic Contrast-Enhanced MRI	8	10	5	5	5	5	10	8?	9	10	18	9	10	93	Yes
Verma et al., 2020 ¹⁷	Tumor Habitat-derived Radiomic Features at Pretreatment MRI That Are Prognostic for Progression-free Survival in Glioblastoma Are Associated with Key Morphologic Attributes at Histopathologic Examination: A Feasibility Study	9	8	5	4	4	3	8	7	8	8	17	8	8	81	Yes
Yan et al., 2020 ¹⁸	A Neural Network Approach to Identify the Peritumoral Invasive Areas in Glioblastoma Patients by Using MR Radiomics	8	8	4	4	3	3	7	7	7	6	18	7	6	75	Yes
Kim et al., 2019 ¹⁹	Incorporating diffusion- and perfusion-weighted MRI into a radiomics model improves diagnostic performance for pseudoprogression in glioblastoma patients	9	8	5	4	4	2	8	8	7	7	13	7	7	75	Yes
Chaddad et al., 2018 ³⁰	Prediction of survival with multi-scale radiomic analysis in Glioblastoma Patients	10	0	5	5	5	0	10	10	10	10	20	10	10	85	Yes
Sanghani et al., 2018 ²¹	Overall survival prediction in glioblastoma multiforme patients from volumetric, shape, and texture features using machine learning	3	10	5	4	5	0	10	10	10	10	20	10	10	87	Yes
Kebir et al., 2020 ²²	A Preliminary Study on Machine Learning-Based Evaluation of Static and Dynamic PET-PET for the Detection of Pseudo progression in patients with IDH-Wildtype Glioblastoma	7	9	5	5	4	5	9	9	9	10	18	9	10	90	Yes
Zhou et al., 2017 ²³	MRI features predict survival and molecular markers in diffuse lower-grade gliomas	10	10	5	5	5	2	10	10	10	10	18	10	10	95	Yes
Zhuge et al., 2017 ²⁴	Brain tumor segmentation using holistically nested neural networks in MRI images	8	10	5	4	3	4	10	8	9	10	18	9	10	89	Yes
Chang et al., 2016 ²⁵	Multimodal imaging patterns predict survival in recurrent glioblastoma patients treated with bevacizumab	10	10	5	5	5	0	10	10	10	10	10	10	10	85	Yes
Continued																

Articles		Research questions and objectives		Complete, robust, and adequate methodology				Results		Conclusions				
Author, year	Title	Clear question (10 pts)	Clear objective (10 pts)	Suitable type of study? (5 pts)	Suitable instrument? (5 pts)	Is it reproducible? (5 pts)	Bias control (5 pts)	Results congruent with the objectives (10 pts)	Results congruent to question (10 pts)	Properly defined terms (20 pts)	Conclusion that answers the question (10 pts)	Conclusion that achieves objectives (10 pts)	Total	
Chaddad et al., 2016 ²⁶	A quantitative study of shape descriptors from glioblastoma multiforme phenotypes for predicting survival outcome	10	10	5	5	5	5	10	10	10	10	10	90	Yes
Artzi et al., 2013 ²⁷	FLAIR lesion segmentation: Application in patients with brain tumors and acute ischemic stroke	8	10	5	5	5	4	10	8	20	9	10	94	Yes

Table 1. Quality control assessment. The articles included in our systematic review were assessed by a quality control evaluation to select those who achieved a score equal to or higher than 75 points for a meta-analysis.

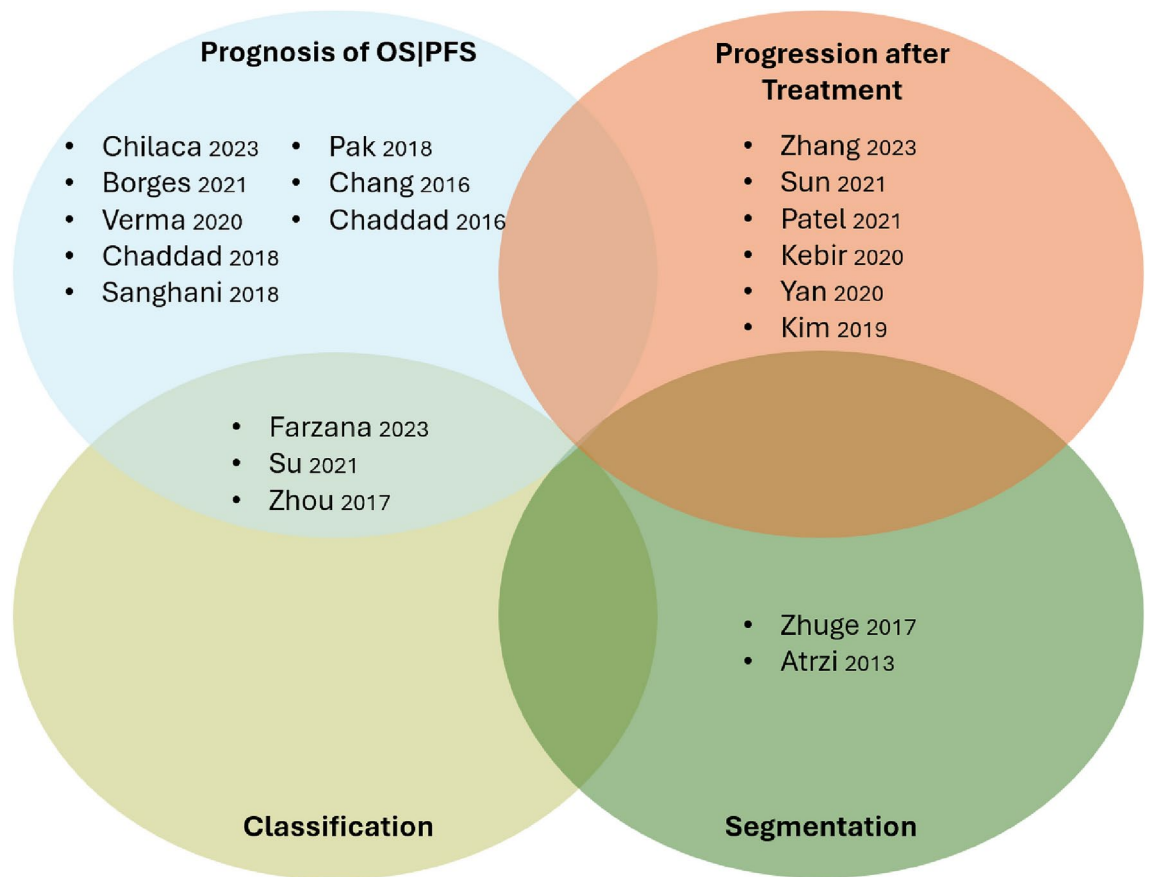


Fig. 4. Application topics of systematic review.

the yellow branch due to Chang developing four different models. This comparison shows that this mind map analysis allows to identify on the fly each studies' strengths, methods, and specific approach, recuperating the essential parts of each research.

AIRI meta-analysis table

In the tables (Tables 2, 3), when a particular article created models for two different application topics, both were reported. Sometimes camel case convention was used to optimize space. To organize information, the following convention was used: category:{element1,element2,..., element n}. On some cases, the total number was specified inside brackets.

Finding the cornerstone of AIRI: comparison between studies, a mixed method approach.

Clinical characteristics in high-grade glioma From the evaluated studies, a striking 94.4% are centered on adult patients (Table 2), while there has been less investigation on the pediatric population. This is reasonable because incidence of malignant glial tumors in the pediatric population is lower, ranging from 3 to 15%²⁸. However, the prognosis for primary central nervous system tumors in children remains equally grim as in adults, underscoring the importance of exploring this field. Some variances have been noted in imaging assessments, particularly in differentiating characteristics observed in T2 multiparametric resonance sequences²⁹. Thus, evaluating radiomic features across T1 sequences with contrast, T2, FLAIR pre- and post-treatment becomes imperative, shedding light on the distinct tumor habitat behaviors in pediatric glial tumors (Table 3).

While most studies concentrate on hemispheric lesions, our previous publication¹² uniquely addresses midline tumors, recognized as the most aggressive subgroup. This pathology merits greater global attention due to its particularly grim prognosis. Thus, emphasizing efforts in understanding and addressing midline tumors could significantly impact overall outcomes in glioma management.

Segmentation and radiomics characteristics The first finding was the heterogeneity of the methodology applied in radiomics processes (identification of volumes of interest, segmentation, and analysis) with AI support in neuro-oncology.

In radiomics analysis, most studies extracted first and second-order features, with a mean of 2,087 initial features per patient (Table 3). These features were obtained based on various analysis parameters, including regions of interest, number of segmentations, and image sequences. Given the large volume of data, we emphasize the importance of utilizing AI and big data analysis platforms to manage and interpret the results effectively.

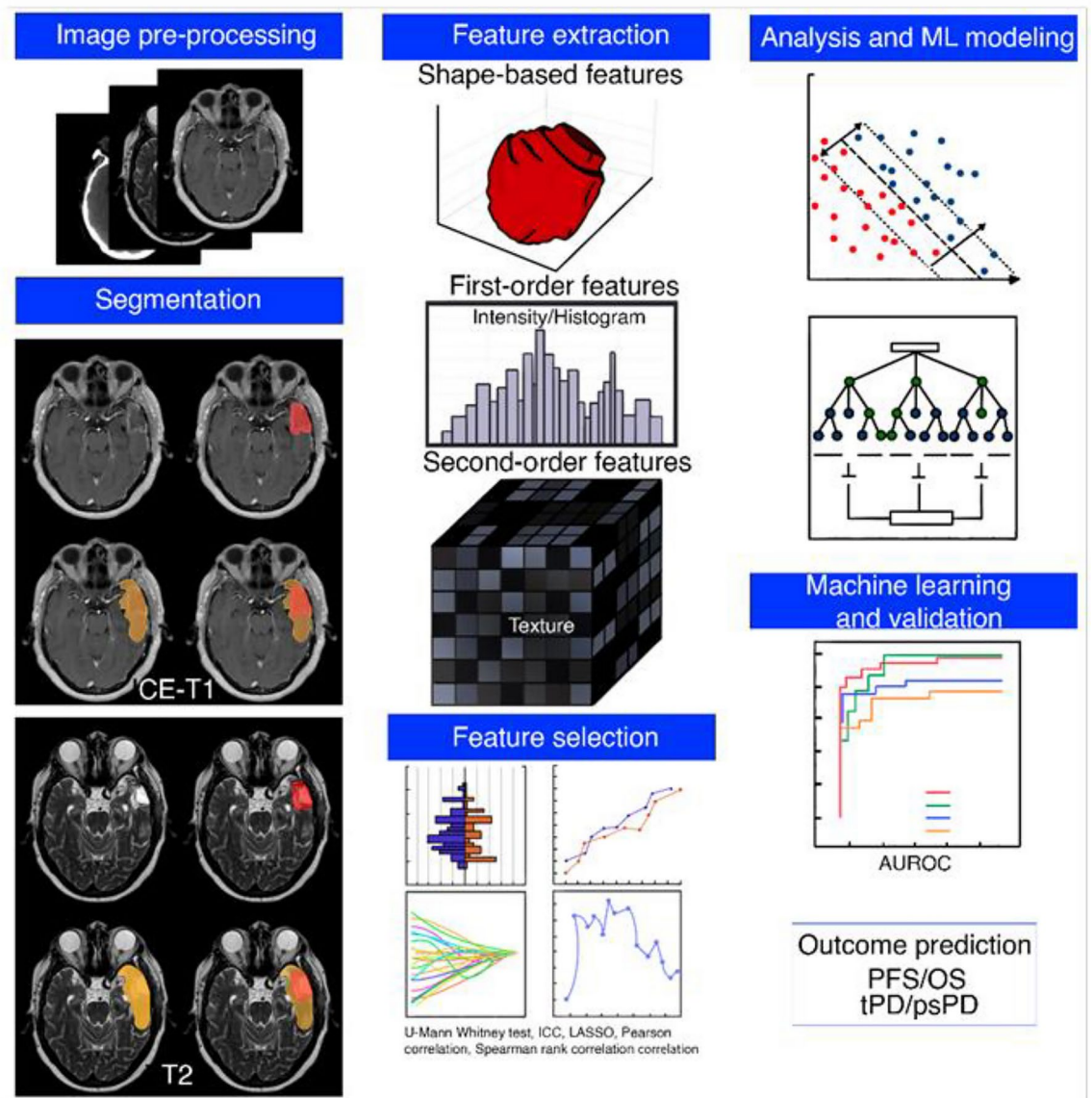


Fig. 5. AIRI roadmap (pipeline). The steps involved in the radiomics analysis workflow start with image acquisition and pre-processing and end with the application and evaluation of the outcome. After pre-processing, segmentation of the tumor regions (ROI) in MRI sequences is performed to then extract data from the segmented regions (feature extraction). Afterward, the most predictive features are selected (feature selection), and the ML model is trained with a specific algorithm (model training). AI has been incorporated to perform predictions of outcomes as the final step.

Many studies did not include clinical features, while some incorporated molecular features (58%) to help avoid bias (Tables 2, 3).

Second-order radiomic characteristics were used in 68% of the studies. The alterations on these features provide detailed metrics on the tumor microhabitat, enhancing our understanding of the diagnosis and impacting surgical treatment plans as well as other therapies such as radiotherapy. This allows for precision and personalized treatment tailored to the specific phase of the oncological disease in the central nervous system.

In the segmentation step, 52.6% of studies focused on both the tumor and other regions, while 36.8% focused solely on the tumor. Managing multiple regions can complicate the analysis of the extensive data obtained from each segment. Therefore, it is imperative to reach a consensus on the regions of interest (ROI) to be segmented and to establish the minimum necessary volumes for ROI, such as tumor, edema, and necrosis for prognosis models, and only tumor for progression after treatment models, which are the number used on the best performing models (Tables 2, 3). Particularly, tumor and edema are crucial for clinical and therapeutic planning, including surgery and radiotherapy. The lack of consensus on the number of regions to segment, the methodology to follow, and the accuracy of results presents significant drawbacks, leading to discrepancies in

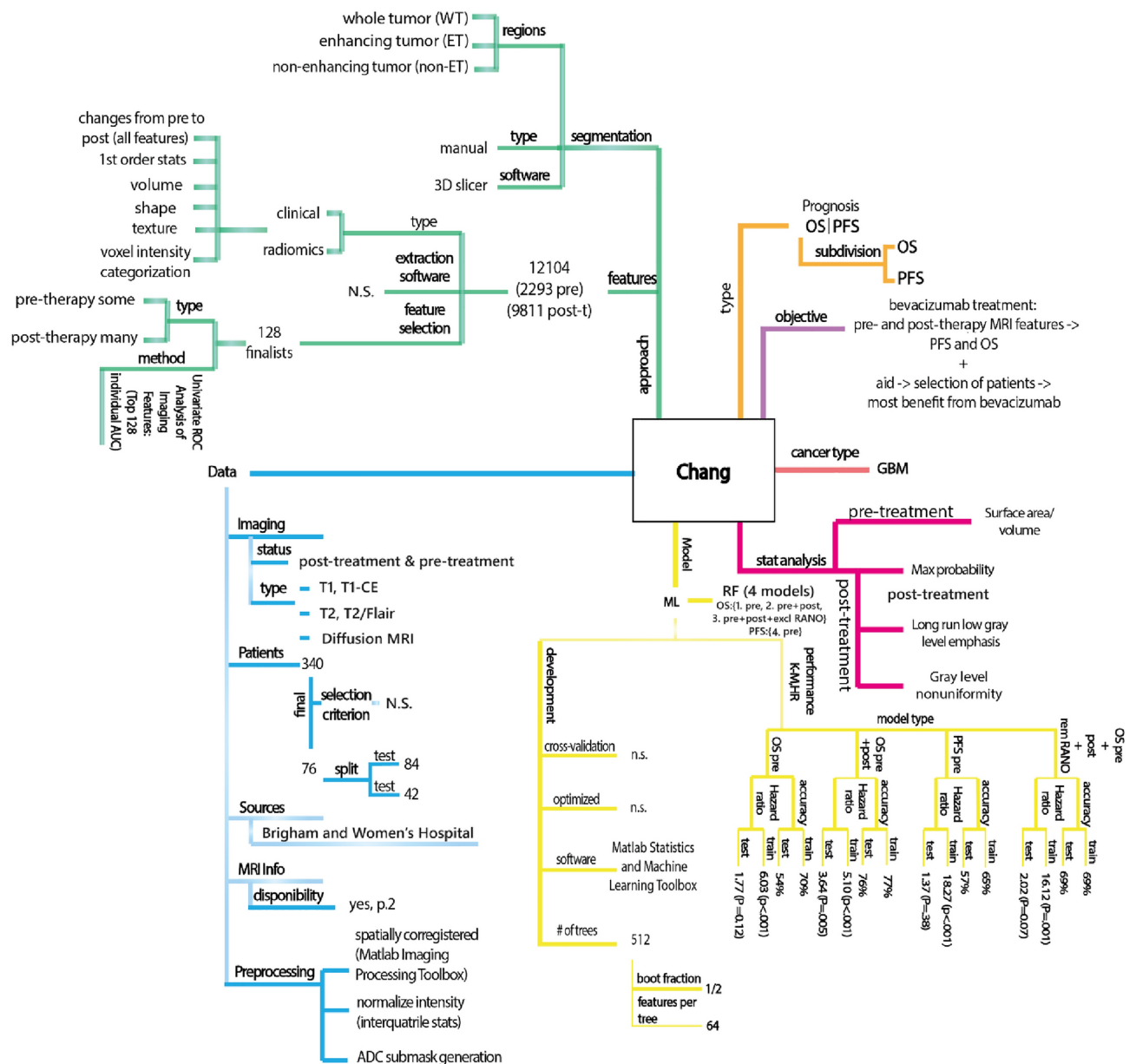


Fig. 6. Example mind map of an article on the prognosis of the overall survival (OS) or progression-free survival (PFS).

radiomic acquisitions and analysis. Additionally, the software used to extract features varied between studies, with Pyradiomics and MATLAB being the most commonly used (on four studies each) (Tables 2, 3).

Characteristics of the artificial intelligence models proposed in the studies Most studies (76%) employed machine learning algorithms, while only a smaller portion (26%) utilized statistical algorithms (Table 2). The most commonly used machine learning algorithms were Random Forest (RF), featured in six studies, and Support Vector Machine (SVM), implemented in three studies. LASSO, combined with either linear regression or logistic regression, was also used in three studies. Other algorithms included Linear Discriminant Analysis (LDA), Bayesian Networks (BN), CatBoost (CB), and two types of neural networks (NN) (Table 2).

Finding the cornerstone of AIRI: comparison between studies, following the footsteps of the best binomial AIRI models—application topics analysis

Prognosis of OS or PFS studies Regarding the studies that did prognosis of OS or PFS, the ones that had the best performance were: Su-2021, Chaddad-2018, Pak-2018, and Sanghani-2018. For predicting two class OS, the k-fold cross-validation test scores of the top articles were an AUC of 85.5%²⁰, and an accuracy of 98.7%²¹. The mean AUC and accuracy of the studies that predicted two class OS were 83.2% with a standard deviation of 11.6% and 87.3% with a standard deviation of 11.2% respectively. For the three class OS classification, the top

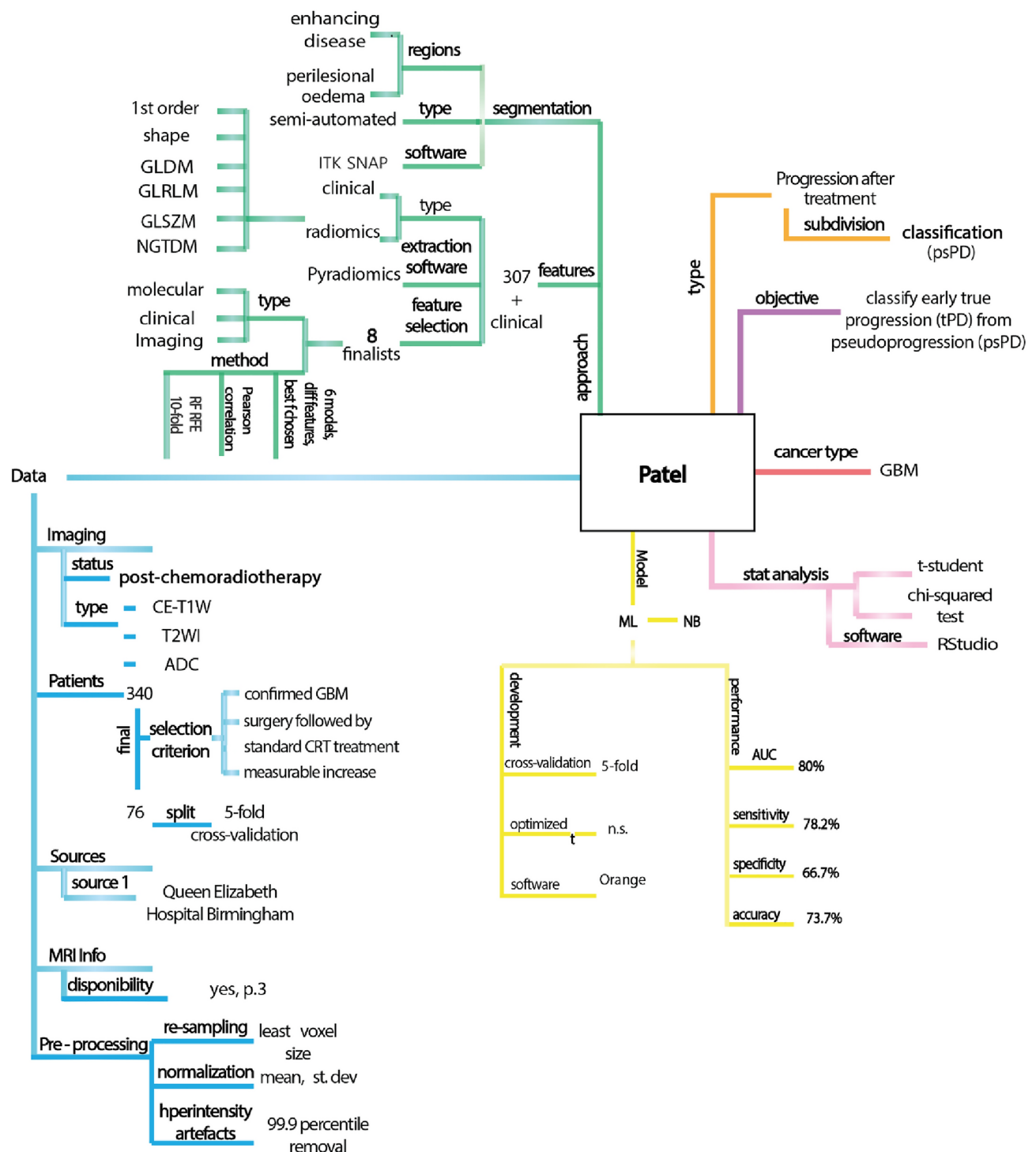


Fig. 7. Example mind map of an article on progression after treatment.

test scores for the k-fold cross-validation were an AUC of 98%¹³ with an accuracy of 94.8%¹³ and an accuracy of 89.0%²¹. For two class PFS, the best k-fold validation test scores were an AUC of 85.4%²⁰ and a *p* value on the long rank test of <0.001¹⁶. All this information can be found in Tables 2, 3.

The features that proved to be predictive for the prognosis of OS or PFS were many. Because the initial features varied from study to study, the features found to be the most predictive after feature selection differed from author to author. The standard features (features that were included in most studies) that proved predictive are 1st order (kurtosis, energy necrotic Ve map), standard deviation, as well as others), 2nd order (GLCM, GLDM, NGTDM, GLRLM, GLSZM), shape and volume, and clinical and molecular features (age, IDH, and MGMT). There were also innovative features that proved predictive in specific articles. For example, logFilter texture features were the only features used in Chaddad's model, which had a good performance. Gabor texture

Main information										
General information		Focus	Objective	Cancer type	Molecular feature	Patients	Clinical scales	Tumor location		
Chilaca, 2023, Mexico City, Mexico		Prognosis OS PFS	Feature Predictibility for PFS and OS: radiomics 1st and 2nd order	DMG	H3K27M mutation	Paediatric, young adults	N.S	Brainstem		
Borges, 2021, SãoPaulo, Brazil		Prognosis OS PFS	Training model to calssify patients that needed careful care (low PFS)	GBM	NS	Adults	N.S	Hemispheric		
Verma, 2020, Cleveland, USA		Prognosis OS PFS	PFS with RRS, and Histologic correlation with RRS	GBM	IDH,	Adults	KPS	Hemispheric		
Chaddad, 2018, Montreal, Canada		Prognosis OS PFS	Feature Predictibility: multiscale texture	GBM	N.S	Young adults, adults	KPS	Hemispheric		
Sanghani, 2018, Queenstown, Singapore		Prognosis OS PFS	Classifier model of OS in 2 and 3 classes	GBM	N.S	N.S	KPS	Hemispheric		
Pak, 2018, Seoul, Korea		Prognosis OS PFS	PFS using DCE imaging	GBM	MGMT	Adults	N.S	Hemispheric		
Chang, 2016, Boston, USA		Prognosis OS PFS	Bevacizumab treatment: pre- and posttherapy MRI features—> PFS and OS	GBM	N.S	Adults	N.S	Hemispheric		
Chaddad, 2016, Montreal, Canada		Prognosis OS PFS	Shape features for characterizing GBM and doing prognosis OS	GBM	N.S	Young adults, adults	N.S	Hemispheric		
Farzana, 2023, Norfolk, USA		Prognosis OS PFS, Classifier	Show usefulness of fractal texture for predict REP vs no REP and OS	AnaplasticAstrocytoma&GBM(g.4)	IDH, IDH wild type, MGMT	NS	N.S	Hemispheric		
Su, 2021, Tianjin, China		Prognosis OS PFS, Classifier	Model that classifies GBM molecular subtype and OS	GBM	IDH	NS	N.S	Hemispheric		
Zhou, 2017, Hunan, China		Prognosis OS PFS, Classifier	VASARI features &/or texture and clinical features for predicting OS, PFS, & molecular biomarkers	LGGs	IDH, 1p/19q codeletion,	N.S	KPS	Hemispheric		
Zhang, 2023, Beijing, China		Progression after treatment	Classify injury from treatment vs recurrence	GBM(g.2, 3, 4)	IDH	Adults	KPS	Hemispheric		
SunY, 2021, Shaanxi, China		Progression after treatment	Differentiating pseudoprogression from true progression after standard treatment	GBM	N.S	Adults	KPS	Hemispheric		
Patel, 2021, Birmingham, UK		Progression after treatment	Classify early true progression (rPD) from pseudoprogression (psPD)	GBM	MGMT	Young adults, adults	ECOG	Hemispheric		
Yan, 2020, Cambridge, UK		Progression after treatment	Identify the progression areas outside the contrast-enhancing tumour	GBM	IDH, IDH wild type, MGMT	Adults	KPS	Hemispheric		
Kebir, 2020, Essen, Germany		Progression after treatment	Differentiation from tumor progression and PSP in IDH-wildtype glioblastoma	GBM	IDH-wildtype	Adults	KPS	Hemispheric		
Kim, 2019, Seoul, Korea		Progression after treatment	Multiparametric MRI to differentiate pseudoprogression from ETM	GBM	IDH, MGMT	Adults	KPS	Hemispheric		
Zhuge, 2017, Bethesda, USA		Segmentation	Model for automatic segmentation of HGG (GBM) (map of the brain tumor)	GBM (HGG)	N.S	N.S	N.S	Hemispheric		
Atzri, 2013, TelAviv, Israel		Segmentation	Explore automatic methodology for FLAIR lesion segmentation with available software	BT, AIS	N.S	Adults	N.S	Hemispheric		
Model information										
Model classification	Model's algorithm	Classes	Class cut-off	Model disponibility and link		K-fold cross-validation		Held-out test set form fe. se		
Stats	Individual Features	OS: {short, long} PFS: {short, long}	Median	No AI model, individual features	N.A		N.A			
ML	RF	PFS: {shorter, longer}	3 months	No	4		No			
Stats	LASSO Cox Regression	PFS: {High-risk, low-risk}	N.S	featureExtraction:github.com/ccipd/BrIC_Lab/	10		Yes			
ML & Stats	RF	PFS: {short, long} OS: {short, long}	Median	No	5		Yes			
ML	SVM	OS2: {short, long}. OS3: {short, medium, long}	OS2:400d OS3:10 m < x < 15 m	No	5		No			
Stats	LASSO Cox Regression	PFS: {High-risk, low-risk}	1 year (cutoff risk score .374)	All model's coeffs. stated	10		Yes			
ML	RF	PFS {short, long}. OS {short, long}	N.S	No	-		Yes			
Continued										

Model information						
Model classification	Model's algorithm	Classes	Class cut-off	Model dispoibility and link	K-fold cross-validation	Held-out test set form fe. se
ML	RF	OS: {short, long}	Median	No	5	Yes
ML	Class:CatBoost OS:RF	Class: {Rep, no-Rep} OS: {low-risk, high-risk}	OS: median	No	5	No
ML	SVM	OS:{short, medium, long}.Subtype:{Classical, Neural, Preneural, Mesenchymal}	6 m < x < 18 m	github.com/RanSulab/GBM-subtype-survival	10	No
ML	LR	IDH1 mutation {1, 0}. 1p/19q codeletion: {1, 0}. HistGrade: {2, 3}. LGG: {prog, noProg}	N.S	No	-	No
ML	SVM	Class: {RIB1 , TR} (injury, recurrence)	N.A	No	-	Yes
ML	RF	Class: {Pseudoprogression, progression}	N.A	No	5	No
ML	BN	Class: {tPD, psPD}	N.A	No	5	No
ML	CNN	Class: {Progression, no-progression} (per voxel)	N.A	No	-	Yes
ML	LDA	Class: {Tumor progression, PSP}	N.A	On request	3	Yes
Stats	LASSO Logistic Regression	Class: {Pseudoprogression, early tumor progression (ETM)}	N.A	No	10	Yes
ML	HNN	Class: {1 , 0 } (pixel in manually segmented ROI)	N.A	Public domain	2	Yes
N.A	N.A	Class: {tumor, no tumor} (region)	N.A	N.A	N.A	N.A
Model performance						
AUC (train)	AUC (test)	Accuracy (train)	Accuracy (test)	Sensitivity	Specificity	Others
OS: 68.6% PFS: 78.1%	-	-	-	OS 85.1%tr PFS 52.7%tr	OS 36%tr PFS 100%tr	N.A
-	OS: 81%f	-	-	-	-	Recall: 100% prec: 39%
-	-	-	-	-	-	p < .012te, p < .0001tr
-	PFS 85.4%f OS: 85.5%f	-	-	-	-	-
-	-	-	OS2: 98.7%f OS3: 89.0%f	-	-	-
-	-	-	-	-	-	p < .001
-	-	OS: 77% PFS:65%	OS: 76% PFS: 57%	-	-	-
-	OS: 67.0%f	-	-	-	-	-
-	CI 79.3%f OS 76.2%f	-	CL 78.1%f OS 73.48%f	-	-	FPR CI14.5%f OS 32.2%f
-	OS: 98%f subtt: 95%f	-	OS: 94.8%f subtt: 82%f	OS: 93.3%f PFS: 82.5%f	OS: 96.7%f PFS: 94.0%f	-
86%,96%,86%,80%	-	-	-	75%,90%,74%,76%tr	78%,89%, 79%,72%tr	PFS, OS:feature pValues
96%	94%	-	-	94.4%te	91.7%te	-
-	-	-	72.8%f	78.4%f	61.3%f	-
-	80%f	-	73.7%f	78.2%f	66.7%f	-

Model performance						
AUC (train)	AUC (test)	Accuracy (train)	Accuracy (test)	Sensitivity	Specificity	Others
–	–	92.6%	78.5%	–	–	PPV:16.8%(t) NPV:78.5%(t)
95%	93%	–	–	–	–	–
–	85%,90% ^f	–	not clear	71.4% ^{te}	90% ^{te}	–
–	–	–	–	85% ^{te} 81% ^{tr}	–	DSC 83% ^{te} 78% ^{tr}
–	–	–	–	–	–	Person corr: 95%, $p < .001$

Table 2. Summarized information on the development of the AIRI models of the 19 studies (What?). N.A. not applicable, N.S. not specified, OS overall survival, PFS progression-free-survival, GBM glioblastoma multiform, #%*f* K-fold score, *e.f.* each fold, – value not reported, ML machine learning, Stats statistics, SVM support vector machines, RF random forest, LDA linear discriminant analysis, LASSO least absolute shrinkage and selection operator, LR logistic regression, CNN convolutional neural network, HNN holistic neural network, ET enhancing tumor, CI class, *subs* subtype, *tr* train set, *te* test set, RFE recursive feature elimination, BFE backward feature elimination, P.C. Pearson correlation, AIS acute ischemic stroke, BraTS brain tumor segmentation competition, *fe.* feature, *se.* selection, *clt**mol* clinical and molecular features, BC CatBoost, DMG diffuse midline glioma, ECOG eastern cooperative oncology group), IDH isocitrate dehydrogenase, KPS Karnofsky performance scale, MGMT O6-methylguanine-DNA methyltransferase, MRI magnetic resonance imaging, ASL arterial spin labeling, DSC dynamic susceptibility contrast, DWI diffusion-weighted imaging, FLAIR fluid attenuated inversion recovery), T1WI T1-weighted images, T1-CE contrast-enhanced, T2WI T2-weighted images, AUC area under curve, AUROC area under the receiver operating characteristic, GMM Gaussian mixture models.

features addition to the standard features were used in Sanghani's very good-performing model. Su's innovative model used only deep imaging features (4D map, 2D fisher vectors).

The algorithms behind the models that had the best performance for the prognosis of OS or PFS were SVM, RF, and LASSO. The two models that used the SVM algorithm were among the top-performing models^{13,21} challenging to generalize which is the best algorithm to use. It was not a univariate analysis where only the algorithm changes. The features and the methods used in each article varied additionally to the algorithm used. Despite that being the case across articles, some indeed performed a univariate analysis of the algorithms. These articles had the profitable approach of training several models, each with a different algorithm while maintaining the other variables constant, and then picking the best-performing one. For example, Su trained three models with different algorithms; SVM, RF, and LR, and reported accuracies of 94.8%, 69.6%, and 84.6% respectively¹³. In this case, the SVM algorithm outperformed the others.

Progression after treatment There were six studies whose application topic was progression after treatment, but only four of those diagnosed patients between presenting pseudo-progression from tumor early tumor progression (which excludes Yan's and Zhang's studies). Zhang's study classified patients between presenting injury from treatment from recurrence, and Yan's study predicted future progression areas. The four previously mentioned articles and Zhang's study were included in this analysis because they all diagnose the state of progression after treatment.

The models that have the best performance are Zhang's-2022, Kebir's-2020, and Kim's-2018. For models that diagnosed pseudo-progression from true early progression, the top test scores for the held-out test set are AUC of 93%²² and 85%¹⁹. For models that classified injury against recurrence, Zhang got scores of 94% AUC, 94.4% sensitivity, and 91.7% specificity on the held-out test set³¹. All this information can be found in Tables 2, 3.

The features that proved to be predictive for progression after treatment were more noticeable to identify than those for prognosis. This was because four out of the five articles compared use similar features. These features were: 1st order, texture, shape plus volume, and wavelet transform features of the previously mentioned ones. Despite that, the only features used in Kebir's study were PET features (TBRmax, TBRmean, TTP), and the model performed well²².

There are different ML and statistics algorithms in the studies that performed progression after treatment. The above good-performing models used SVM³¹, Linear Discriminant Analysis (LDA)²², LASSO Logistic Regression¹⁹. Zhang used the strategy of training four models with different algorithms and then highlighted SVM as the best one³¹.

Proposal of a pipeline to develop and implement an AIRI model

The outcome of the systematic review and AIRI meta-analysis led to the proposal of the following two-fold AIRI universal roadmap.

Fold one: AIRI universal roadmap development (1) MRI: For optimum performance, all the MRIs should be taken with the same scanners and with the same parameters such as magnetic field intensity (T), caption time frame, matrix size, number of excitations, slice thickness, gap, and equipment brand^{14,19}. This hold true except if the model will be used to diagnose patients with images that will be taken with scanners or parameters different from the ones used to train the model. In that case, it would be better to use MRI from a variety of different equipment and parameters to train the model, or best, it would be to use MRI from all the equipment and parameters that are going to be used to diagnose patients. This would be the case if the model is intended to be used in several medical centers that have MRI equipment that differs. It is important to keep this point in mind because one could incur in selection bias if one does it improperly.

The MRI imaging types that are suggested to be incorporated into the model are T1W, CE-T1WI, T2W, and FLAIR. Some studies only used CE-T1W, and T2 or FLAIR and also achieved good performances. According to the Brain Tumor Segmentation competition (BraTS) CE-T1W1 compared against T1, and FLAIR are key for the automatic segmentation of the tumor³⁰.

(2) Image pre-processing: It's important to do image pre-processing before segmentation. Some general steps are normalization, interpolation (to achieve the same high resolution on all the MRI), removing outlier voxels, registration, and to skull strip the MRI. Some available software is 3D Slicer, PyRadiomics¹⁵, FSL²⁷, Matlab—SPM plug-in^{25,31}, ITK; N4ITK; ITK-SNAP^{15,24}, R—ANTsR and WhiteStripe packages¹⁹, FMRIB¹⁸, NardiIce¹⁸. It would be ideal to automatize this step with a specific development so that no user needs to interact and do this process.

(3) Segmentation: It is recommended to segment the tumor into three regions, as the studies which used three regions for prognosis had the highest performances (Tables 2, 3). These regions are tumor, edema, and necrosis. On the studies that did treatment progression, most good performing articles only segmented the tumor in one region, tumor from no tumor, so that should be sufficient for that application topic. Once segmented, the segmentation masks should be applied to all the imaging types. For it to be feasible to implement the model and diagnose patients, the segmentation process should ideally be automatized¹⁶. Available software for doing this process automatically are 3D U-Net³², and RA-Unet¹³. 3D U-Net was used to segment into three regions (edema, tumor, necrosis), only using as input CE-T1 MRI sequences³². This makes it very practical, although the accuracy of this method might be lower than the semi-automated or manual approach (refer to other available software's on Table 3). RA-Unet on the other hand, needs all four T1WI, CE-T1WI, T2 and FLAIR, and is not a conventional segmentation method, it must be used with Su's publicly available prognosis model workframe¹³. All the other software's used on the studies were semi-automatic or manual. For it to be viable to implement the model to diagnose patients, it is recommended that the automatic approach is chosen because the others are time consuming¹⁶. The current state of automatic software's is rudimentary, and usually radiologist do it better.

Information		Segmentation			
General information		Level of automatization	Software	#Regions	Name of regions
Chilaca, 2023, MexicoCity, Mexico		Semi-automated	LIFEx	3	Tumoral, edema, EMINT
Borges, 2021, SãoPaulo, Brazil		Manual	3D Slicer	1	Tumor (GTV)
Verma, 2020, Cleveland, USA		Manual	3D Slicer	3	Necrosis, ET, FLAIRhy-perintensity
Chaddad, 2018, Montreal, Canada		Manual	3D Slicer	3	Active tumor, edema, necrosis
Sanghani, 2018, Queenstown, Singapore		Pre-done	pre-done	3	CEtumor, peritumoralEdema, Necrotic&NEtumor
Pak, 2018, Seoul, Korea		Semi-automated	NordicIce	2	ET(T1), non-ET(T2)
Chang, 2016, Boston, USA		Manual	3D Slicer	3	Whole tumor (WT),ET, non-ET
Chaddad, 2016, Montreal, Canada		Manual	3D Slicer	3	Necrosis, active tumor, oedema
Farzana, 2023, Norfolk, USA		Automated	3D U-Net	3	Edema, ET, necrosis
Su, 2021, Tianjin, China		Automated	RA-Unet	10	10 patches, 400 initial segmentation patches
Zhou, 2017, Hunan, China		Manual	N.S	1	Tumor
Zhang, 2023, Beijing, China		Manual	ITK-SNAP	1	Tumor
SunY, 2021, Shaanxi, China		Semi-automated	ITK-SNAP	1	Tumor
Patel, 2021, Birmingham, UK		Semi-automated	ITK-SNAP	2	EnhancingDisease, perilesionalOedema
Yan, 2020, Cambridge, UK		Semi-automated	N.S	2	Progression, no-progression
Kebir, 2020, Essen, Germany		N.A	N.A	N.A	N.A
Kim, 2019, Seoul, Korea		Semi-automated	MITK	1	Tumor
Zhuge, 2017, Bethesda, USA		Manual	Own	1	Tumor from no tumor
Atzri, 2013, TelAviv, Israel		Manual	FMRI(BFSL)	1	Tumor from no tumor

Information		Features					Feature Ext&Software	#Study's total features
General information	Initial features	#Initial features by Patient	Feature selection	#Final features by patient	Final features			
Chilaca, 2023, MexicoCity, Mexico	1st, 2nd:{GLCM, GLZLM, NGLDM};shape&volume	480	Quartiles&QQR, AUROC	OS: 1(4) PFS: 1(13)	1st:{CONV_RIM_sdev, CONV_min, DISCR_std}, 2nd:{GLCM:{contrVar, entropy, jointEntropy}, GLZLM:{GLNU}, NGLDM:{contrast}}	LIFEx	5760	
Borges, 2021, SaoPaulo, Brazil	1st, 2nd:{grayMatrixes};shape&volume, cli&mol{treatmentDoze}	105	RF	1	1st:{Kurtosis}	PyRadiomics and 3D Slicer	4515	
Verma, 2020, Cleveland, USA	1st:{conventional, gradient}, 2nd:{GLCM, LawsE, CoLiAGe}	3024	LASSO cox regression, removed correlated 10-f	17	2nd:{CoLiAGe[ET_perinecroticZoneCT], Laws{Edema_LE, Edema_HyperPlastBloodVesselsLE, ET_InfiltratingTumor, ET_HPBinCT}}	Matlab	471,744	
Chaddad, 2018, Montreal, Canada	1st:{loG Filter Texture (Average, Entropy, St.Dev)}	27	Spearman rank correlation	27	1st:{LoGFilter(edemaStDev(allScales), edemaEntropy(allScales), activeTumorAverage(mediumScale))}	Own	1080	
Sanghani, 2018, Queenstown, Singapore	2nd{grayMatrixes};shape&volume, cli&mol:{age}	2201	RFE with SVM	150	1st:{5}, 2nd:{(GLCM2D:{32}, GLCM3D{12}, GarborText:{86}), shape&volume:{9 (shape)}, cli&mol:{1 (age)}}	PyRadiomics	358,763	
Pak, 2018, Seoul, Korea	1st, 2nd:{GLDM, NGTDM, GLRLM, GLSZM, GLCM};shape&volume{shape} (Krans, Ve, Vp maps of all features)	642	Outliers and LASSO Cox Regression	16	1st:{1}, 2nd{GLDM(3), NGTDM{1}, GLRLM(5), GLSZM(2), GLCM{1}}, shape&volume{3(shape)} (mainly Krans, Ve maps of non-ET T2 high SI)	3D Slicer and PyRadiomics	96,300	
Continued								

Information		Features				#Final features by patient		#Initial features by Patient		Feature selection		#Final features by patient		Final features		Feature ExtrSoftware		#Study's total features	
General information	Initial features	#Initial features by Patient		Feature selection		#Final features by patient		#Initial features by Patient		Feature selection		#Final features by patient		Final features		Feature ExtrSoftware		#Study's total features	
Chang, 2016, Boston, USA	1st, 2nd, shape&volume, changes from pre to post treatment imaging, cli&mol {clinical}	12,104		Univariate AUC (ROC) analysis, Top 128		128		12,104				128		1st:{maxProbability},2nd:{GLRLM{emphasis},GL{nonuniformity}},shape&volume{surfArea/vol}		N.S		1,525,104	
Chaddad, 2016, Montreal, Canada	shape&volume{shape 2D}	33		N.A		33		33				33		shape&volume;Extent (necrosis),Eccentricity (necrosis),Solidity (necrosis)}		Matlab		1320	
Farzana, 2023, Norfolk, USA	2nd:{GTSDM, NGTDM, GLSZM},fractalTexture:{PTPSA, mBm, HE, GmBm},shape&volume, cli&mol:{IDH, MGMT}	600		Copula Modeling and Progress Index		CI: 8 OS: 8		600				CI: 8 OS: 8		2nd:{NGTDM(GmBm, strengthDir37, strength),GTSDM{autocorrelation}},shape&volume{eccentricityE.dema, yAxisLengthNecrosis}		MatLab		42,000	
Su, 2021, Tianjin, China	Deep imaging features (4D maps)	4096		BFE		subt: 370 OS: 174		4096				subt: 370 OS: 174		N.S		Own:Unsupervised Learning		290,816	
Zhou, 2017, Hunan, China	1st, 2nd{GLCM, GLRLM, GLSZM, NGTDM}	3390		4 multivariate models, L.R. imbalanced adj		14		3390				14		2nd{14}		N.S		284,760	
Zhang, 2023, Beijing, China	1st, 2nd{wavelet transform} shape&volume	4255		ICC>0.85, LASSO, RFE, Mann-Whitney Utest		8		4255				8		1st{maxInt, 10thPerc, avgInt},2nd:{GLCM:{diffVariance, JointEnergy, correl2},GLRLM:{highEmphasis},GLSZM:{non-uniformity}}		N.S		536,130	
Sun,Y, 2021, Shaanxi, China	1st, 2nd:{GLCM, GLSZM, RLM, Gabor and Haar wavelet transform}	9675		P.C., >.75, RF: Gini impurity		50		9675				50		1st{6},2nd:{GLSZM:{1},GLCM:{19},RLM:{24}}{92% were wavelet}		Analysis-Kinetics(GE Healthcare)		744,975	
Patel, 2021, Birmingham, UK	1st, 2nd:{GLDM, GLRLM, GLSZM, NGTDM},shape&volume{shape},cli&mol{clinical}	314		P.C., RF RFE tenfold, 6 models,bestFeatures		8		314				8		1st{Kurtosis(ADC)},2nd:{GLCM:{Correl(ADC)},NGTDM{Contrast(ADC)}},shape&volume:{Elongation, Sphericity},cli&mol:{MGMT, Age}		PyRadiomics		23,864	
Yan, 2020, Cambridge, UK	1st, 2nd:{GLCM, GLRLM}	294		N.S		n.s		294				n.s		1st{35},2nd:{77}		Matlab		16,758	
Kebir, 2020, Germany	PET:{TBRmax, TBRmean, TTP}	3		N.A		3		3				3		PET{TBRmax, TBRmean, TTP}		N.S		132	
Kim, 2019, Seoul, Korea	1st, 2nd(original and wavelet-transform: LLL, LLH, LHL),volume&shape, cli&mol{clinical}	6472		LASSO tenfold		12		6472				12		1st{stDev, intentstyRange, sumOfIntensities(originalAll)},2nd:{GLCM{8(waveletTransf)},GLRLM{emphasis},Mean(waveletTransf)}}}		N.S		614,840	
Continued																			

Information		Features			Information		Data					Feature Exts		#Study's total features
General information		Initial features	#Initial features by Patient	Feature selection	#Final features by patient	Final features	#Imaging types	Imaging types	StatusUponImaging (trainingdata)	Data source	#Patients (final#)	#Train	#Test (fold out test)	
Zhuge, 2017, Bethesda, USA	1st: (multiscale and multilevel hierarchical intensity representations)		N.S	N.S	N.S	N.S	2	CE-T1WI, T2WI	Pre-treatment	National Medical Center S.XXI IMSS, Mexico City	12	12	-	N.A
							1	CE-T1	Pre-treatment & radiationPlanning	Ribeirão Preto Medical School Hospital and Clinics	43	32 e.f	11 e.f	
Attrzi, 2013, TelAviv, Israel	no features		N.A	N.A	N.A	N.A	3	CE-T1WI, T2W, FLAIR	Pre-treatment	TCIA, CCE, IVY GAP	156	110 e.f	12 e.f	N.A
							2	T1WI, FLAIR	Pre-treatment	TCGA	40	32 e.f	8 e.f	
Sanghani, 2018, Queenstown, Singapore							4	T2W, FLAIR, T1W, CE-T1WI	N.S	BraTS 2017 challenge	163	130 e.f	33 e.f	
							2	CE-T1WI, T2, DCE	Pre-treatment	Seoul National University Hospital	150	105	-	45
Chang, 2016, Boston, USA							5	T2, T2/FLAIR, DiffusionMRI, T1, T1-CE,	Pre-treatment, post-treatment	Brigham and Women's Hospital	126	84	-	42
							2	PC-T1WI, T2WI FLAIR	Pre-treatment	TCGA(TCIA)	40	32 e.f	8 e.f	-
Farzana, 2023, Norfolk, USA							1	CE-T1	Pre-treatment & radiationPlanning	OhioHealth hospital	70	56 e.f	14 e.f	-
							4	T1, CE-T1, T2, FLAIR	Pre-treatment and post-treatment	TCGA, Brats18 challenge (TCIA), cBioPortal (overlap)	71	64 e.f	7 e.f	-
Zhou, 2017, Hunan, China							4	T1, T2, T2F(FLAIR),CE-T1	Pre-surgical	TCGA + TCIA (overlap)	84, 67, 84, 75	84, 67, 84, 75	-	-
							5	T1WI, T2WI, CE-T1WI, ADC, CBF	N.S	TheFirstMedicalCenter, ChinesePLAGeneralHospital	126	96	-	30

Information		Data								
General information	#Imaging types	Imaging types	StatusUponImaging (trainingdata)	Data source	#Patients (final#)	#Train	#Test (fold out test)	#Test (hold-out test)		
Sun Y, 2021, Shaanxi, China	1	T1-CE	Post-treatment	Tangdu hospital	77	62 e.f	15 e.f	-		
Patel, 2021, Birmingham, UK	3	CE-T1WI and ADC map, T2WI	Post-treatment	Queen Elizabeth Hospital Birmingham	76	61 e.f	15 e.f	-		
Yan, 2020, Cambridge, UK	7	T1C, FLAIR, rCBV, FA, ADC, DTI-p, DTI-q	Pre-treatment	Cambridge University Hospital(tr), other N.S.(te)	57	37	-	20		
Kebir, 2020, Essen, Germany	1	N.S. (N.S.)	endOfRadiotherapy, weeksLatter	the Division of Clinical Neurooncology at the University of Bonn Medical Center and University Hospital of Cologne	44	20 e.f	10 e.f	14		
Kim, 2019, Seoul, Korea	4	FLAIR, ADC, CBV, CE-T1WI	Post-treatment	Department of Radiology at tertiary center University of Ulsan College of Medicine	95	55e.f	6.e.f	34		
Zhuge, 2017, Bethesda, USA	4	T2WI, FLAIR, T1WI, CE-T1WI	N.S	BraTS 2013 (tr) + National Cancer Institute (te)	30	10 e.f	10 e.f	10		
Atrzi, 2013, TelAviv, Israel	1	FLAIR	Pretreatment, duringTherapy, followingAIS	Tel Aviv Sourasky Medical Center	44	-	-	44		

Table 3. Summarized information on the building of the model of the 19 articles (How?).

However, if a neural network is created with good accuracy, it would be a steppingstone for automatizing and diagnosing patients on medical centers. The BraTS challenge is an annual competition where researchers are trying to do just that. Researchers from all over the world compete to develop an automatic model that does segmentation-related tasks.

(4) Features extraction: It is suggested to extract the following standard radiomic features: 1st order (all), 2nd order—Gray-Level Texture Matrixes (GLCM, GLDM, NGTDM, GLRLM, GLSZM) shape and volume features, and clinical and molecular features (age, functional status and molecular status in respect to IDH, MGMT). This can be done in Python with the package PyRadiomics, or in MatLab. Additionally, it is advisable to include innovative features, such as texture LoGFilter first order features, and Garbor textures, which can help to improve the model's performance. Studies whose application topic is progression after treatment, also included wavelet transform features, achieving good results. Therefore, if this type of model is going to be trained, it is recommended to include those types of features. It is recommended to automatize steps two to four on a single software development so that later, it can be used on the implementation phase as well (Fig. 8).

(5) Split of training and testing data: It is important to separate the testing data from the data that will be used for feature selection and model training. If one does not leave apart the data for testing before doing feature selection, there is a risk that the features that will be selected are the ones that best describe the testing set. Separating both sets from the beginning, ensure that the reported testing performance is the one to expect when the model is implemented and avoid bias. In a way, this could be termed confirmation bias because one is only looking at the features that one already knows will correctly describe the testing set. It is as knowing ahead of time which features will work best for the particular testing set.

(6) Feature Selection and Model training: Best performing models combined feature selection with model training^{13,21} using either RFE or BFE. This process can be done by evaluating the training performance alone, or computationally more time-consuming but better, is to do it with k-fold cross-validation²¹. In RFE and BFE, the number and combination of features that achieve the highest performance are selected. To make a more robust model, proof against overfitting, it is best to choose less than that number, which usually has many features. To reduce computational time, it is recommended to do Pearson Correlation coefficients before doing RFE or BFE to eliminate highly correlated features. Many articles did this as a first step for feature selection.

(7) Choosing the best model: For best results, it is a good approach to train several models with different algorithms, for example, SVM, LR, and RF¹³, and then identify the best-performing model¹³. Performing k-fold cross-validation for each model is highly recommended, if not already done (on the feature selection step), to evaluate the model's performance.

(8) Testing the model:

- Input: Use unknown new data (held-out test set from step 5).
- Prediction: Make predictions with the trained model.
- Performance Indicators: Obtain all performance indicators (AUC, accuracy, specificity, and sensitivity).
- Interpretation: Check for overfitting and general outcome.
- Reporting: Report all performance indicators in order to ensure transparency and foster comparability between studies¹⁵.

Testing the model with this dataset ensures the reported performance is trustworthy, making it a reliable expectation for clinical tasks such as prognosis, diagnosing progression after treatment, and other classifications. To check for overfitting, compare the test score with the training score. If the training score is significantly higher, the model is overfitted. One way to address overfitting, which is crucial to report, is to reduce the number of features. Selecting fewer features than the optimum number obtained in step 6 using RFE or BFE can help make the model more robust, universal, and less overfitted. If this adjustment is made and a new model with fewer features is trained, it is essential to note this in the report. If the model is retrained with less features and this is not reported, one is leading to a biased test performance indicator because the new model will be optimized against the testing set, by selecting the features that at the end describe the testing set best. This type of bias has been previously explored on step 5 of this section.

However, further adjusting the model to improve the held-out test set performance may compromise the validity of the held-out test set. This is because optimizing the model against the held-out test set effectively creates a model tailored to that specific test set, thereby nullifying its use as an unbiased test.

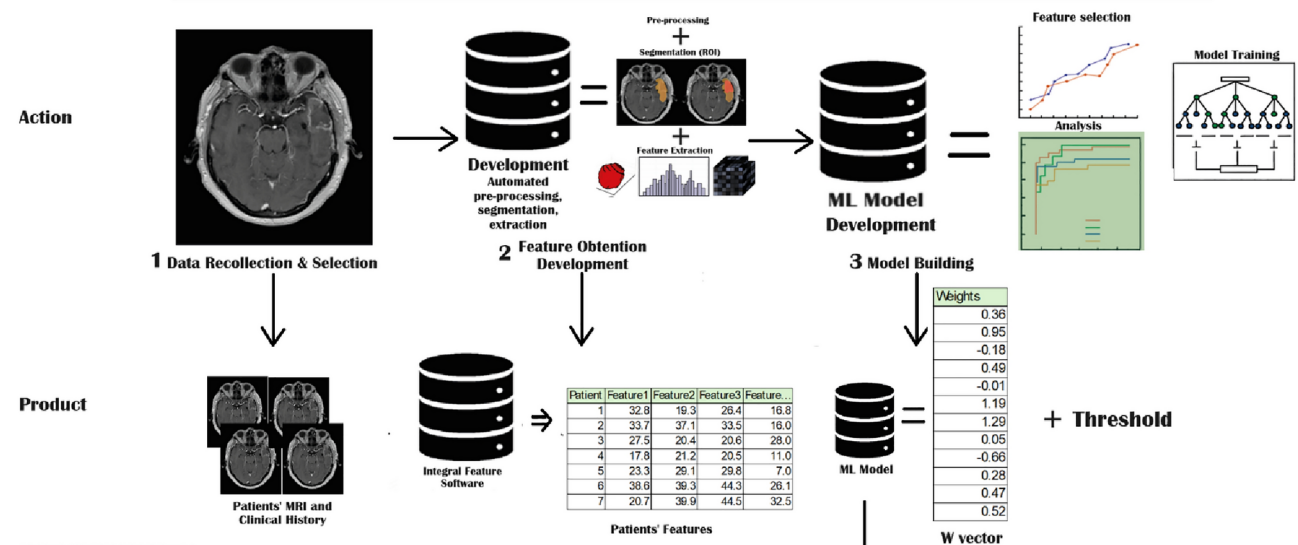
Fold two: AIRI universal roadmap implementation It is suggested that when the model is developed, the development process be divided into two software's, as exemplified in Fig. 8, and leave the performance analysis as another separate step from those software's. Additionally, to this two software's, a third software needs to be developed to get the patient data, apply the first and second previously developed software's, and present both the classification result and the reasons behind it to the oncologist.

(1) A window where the patient's MRI and clinical data is loaded must be developed. Another option is that the MRI and clinical data be automatically fed to this software.

(2) After that, the previously developed software on steps two to four which automates the image preprocessing, ROI segmentation and does feature extraction should be implemented using those MRI and clinical data as input. This process output is the radiomic and clinical features of the patient. To have an efficient data storage, it is advisable that the radiomic features that are not part of the feature selection cohort be eliminated or not stored on the database.

(3) After having obtained the radiomic and clinical features of the patient, the machine learning model trained in the development process should be applied to predict or classify the patient's condition, taking as an input those features. The output of this process is the classification of the patient.

DEVELOPMENT



IMPLEMENTATION

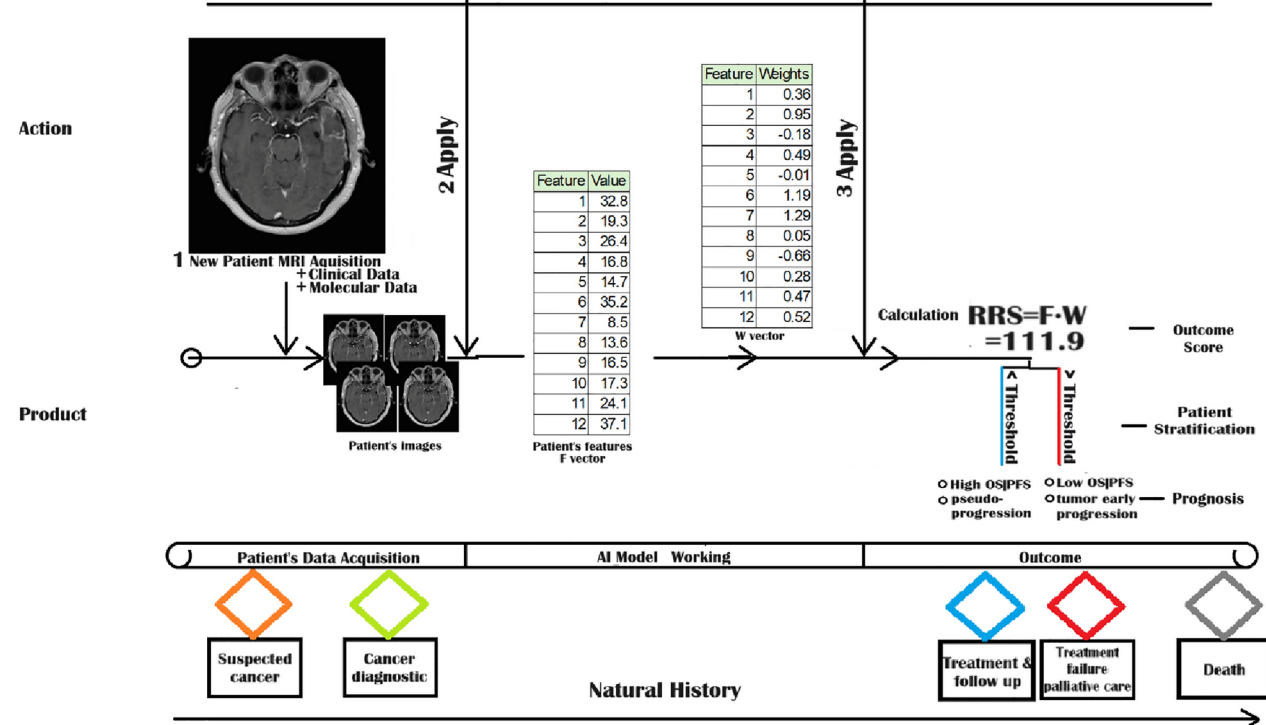


Fig. 8. Model's Development and Implementation Roadmap. Top part, model development. Middle part, implementation process. Bottom part, patient's clinical natural history. This figure is exemplified with the model algorithm behind being LASSO.

(4) The output of the classification should be presented on a visual manner together with the explanation as to why the model predicted that output. It is important to present this explanation, as this will aid the oncologist to understand the process behind it and therefore allow him to take better decisions. The explanation is also important so that the oncologist can tell the patient why he was classified in such a way.

Discussion

Epistemic meta-analysis is an approach used in research to analyze and synthesize existing knowledge, theories, or findings. On this research, the term 'epistemic' refers to the examination of knowledge or understanding by evaluating the underlying assumptions, methodologies, and theoretical frameworks of AIRI studies. It allows to understand the relevance of several types of features and to see how they get integrated into AIRI models. This approach generates the creation of ad hoc mental maps in order to identify the common backbone of the research trajectories that are common in all the investigations in the AIRI and distinguish them from the variables that

are specific of the backbone of each study. This is invaluable as it allows for the creation of a navigable roadmap that can evolve. As it evolves, it can acquire new properties that enable a more generalized interpretation of results and open the epistemic horizon. Beyond academia this aids oncologist for the decision making in the data interpretation, diagnostic, prognostic, treatment, and follow up for the best care of the patients by giving insights on how to develop and implement AIRI models.

This study aims to unify the AIRI field and to expand the epistemic horizon so that in the future, a robust quantitative meta-analysis with meta-analytic techniques can be carried out. This study demonstrates and raises awareness to the scientific community the lack of comparability and heterogeneity of the methods, processes, and metrics used in the AIRI field, showed in the “AIRI meta-analysis Table” section (Table 2—What?, Table 3—How?), which was done meticulously. A general roadmap was presented as a suggested guideline for scientists to follow on future research, so that the field can evolve into a more robust one.

In this research, the expansive potential of the AIRI in high-grade glial tumors, primarily glioblastoma was explored. Swift diagnoses are imperative for conditions such as high-grade gliomas, given their bleak prognosis. The intricate journey from diagnosis to treatment and progression evaluation underscores the importance of specialized oncology reference centers with health care professionals training and becoming familiar with the AIRI. The promotion of integrating automation technologies for capturing, identifying, and discriminating between tumor characteristics, habitat, and surrounding healthy tissue in digital images of magnetic resonance studies is crucial. This integration facilitates precise diagnosis and treatment planning customized to meet each patient's specific needs, thereby minimizing toxicity to healthy tissues in individuals with brain tumors. It encompasses the consideration of personalized-integral and effective management strategies, particularly pertinent in challenging neoplasms such as diffuse midline gliomas.

Towards a standard for AIRI interfiled: good practice

There must be a consensus as to how training and testing data is handled. Many studies have certain bias in their reported performances. Some models^{16,20} are particularly robust compared to others^{13,21}. The robust ones, leave the test set completely separated from the process of developing the model, as shown in Table 2 on the column “held-out test set form fe. se.”. On the other hand, some studies performed feature selection (RFE and BFE) with the data that was used for testing on the k-fold cross-validation, and have no further external validation data. This approach makes the model susceptible to overfitting, giving the possibility for the model's performance to drop when tested with real-world data. This scenario could happen because the features selected to build the model are those that describe the testing set particularly well.

Several clinical characteristics are excluded in different proposed models due to their perceived potential bias and as covariates that may alter the direction of progression prediction results. However, few studies have assessed stratification by molecular identity as a means to mitigate bias.

Table 2 illustrates that only 57.6% of the studies evaluate molecular alterations, indicating an area for improvement in stratifying these molecular alterations alongside clinical data and functional status.

The Radiomics Quality Score (RQS) was not utilized in these studies. RQS serves as a valuable tool aimed at enhancing the quality of radiomics research and necessitates widespread acceptance. It employs a system of rewards and penalties to evaluate the methodology, analysis, and reporting, thereby fostering optimal performance in radiomics analysis and scientific practice. While not yet established as the definitive standard for evaluating radiomics studies, efforts have been made to develop two versions of RQS, with the latest iteration (RQS 2.0) currently under development. Both versions comprise a comprehensive questionnaire consisting of 36 checkpoints, focusing on principles such as Fairness, Universality, Traceability, Usability, and Robustness, elucidating the efficacy of radiomics analysis.

It is also necessary that all the studies calculate and report all the model's performance variables. Most studies reported the AUC score, but not all of them (Tables 2, 3). Some others reported the accuracy, but many did not. The same happened for specificity and sensitivity. It is necessary that all studies report all the machine learning performance indicators: AUC, accuracy, sensitivity, and specificity to ensure transparency of the results. It is straightforward to calculate them. It is advisable to show all the results and not only the best ones.

Excluding the testing set from the feature selection process, reporting all of the model's performance indicators, including clinical and molecular features, and using RQS would enhance the comparability of studies, enabling more conclusive analyses. This approach would expedite progress in the AIRI by facilitating the identification of optimal strategies regarding feature selection, algorithms, MRI sequences, and ROI.

This work paved the way with a roadmap and methodology (Figs. 1, 6, 7, Tables 2, 3) and opened a new epistemic horizon constructed with the effort of all the studies included on this research^{11–27,31,32}. AIRI is now an interdisciplinary field, promoting collaboration and the creation of new projects where information is obtained and integrated universally, allowing research to be strengthened and fostered among different groups worldwide, sharing their knowledge as tools, ends, and practices. Thus, advancing AIRI in a joint direction that encourages rapid evolution in this field.

Conclusion

The integration of Artificial Intelligence and Radiomics Inter-field (AIRI) in neuro-oncology, particularly for predicting high-grade glioma progression, represents a paradigm shift in precision medicine. This systematic review and epistemic meta-analysis highlight the strengths and limitations of existing AIRI models while proposing a structured road map to improve standardization, reproducibility, and clinical implementation (Figs. 1, 2, 3, 4, 5, 6, 7, 8, Tables 2, 3).

Key findings indicate that robust radiomic features, particularly first- and second-order texture metrics, combined with machine learning models such as Support Vector Machines (SVM) and Random Forest (RF), yield high predictive accuracy for overall survival (OS) and progression-free survival (PFS). However, methodological

heterogeneity across studies, including variations in imaging protocols, segmentation techniques, and feature selection strategies, underscores the need for standardized AIRI frameworks to enhance comparability and translational applicability (Fig. 8).

By synthesizing existing methodologies and identifying best practices, this study paves the way for a more concise approach to AI-driven radiomics in glioma prognosis. The proposed universal roadmap outlines critical steps from data acquisition to model development, emphasizing automated workflows for segmentation, feature extraction, and AI model training. Implementing these strategies will improve diagnostic consistency, facilitate early intervention and ultimately enhance patient outcomes (Figs. 4, 5, 6, 7, 8).

Future research should focus on refining AI models with multi-center datasets, integrating molecular and clinical biomarkers, and adopting rigorous validation protocols to bridge the gap between research and real-world application. By fostering interdisciplinary collaboration and adhering to good practices in AIRI, the field can move towards standardized, clinically actionable framework that optimally supports oncologists in patient management and decision-making.

This epistemic meta-analysis gives insights as to how up to today, there is not an established methodology to develop and report AIRI models. On the AIRI table (Tables 2, 3) it becomes evident that some studies report only certain performance indicators and not all of them (AUC, accuracy, sensitivity, and specificity). This helps to raise awareness of the current situation so that it can be amended on future research. We consider this article to be conspicuous and therefore is a vector of change to evolve the AIRI field because it helps to evaluate the field on its' current state and point out its' aspect to improve to become a more robust field (Figs. 4, 5, 6, 7, 8).

Furthermore, despite there being AIRI roadmaps, they do not go into the fine details of what to do and not to do on model development to get a robust model and comparable outcomes. This paper proposes specific guidelines and a novel roadmap, which we believe is more complete than others, on the section *Proposal of a pipeline to develop and implement an AIRI model* for future research to follow. Additionally, the roadmap that we propose at the end, includes how to implement the trained model on a clinical setting, which is a breakthrough.

Data availability

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request (M.M.A.B. myriamab@unam.mx).

Received: 17 June 2024; Accepted: 9 April 2025

Published online: 08 May 2025

References

- Carrete, L. R., Young, J. S. & Cha, S. Advanced imaging techniques for newly diagnosed and recurrent gliomas. *Front. Neurosci.* **16**, 787755 (2022).
- Jacobs, A. H. et al. Imaging in neurooncology. *NeuroRx* **2**, 333–347 (2005).
- Kelly, C., Majewska, P., Ioannidis, S., Raza, M. H. & Williams, M. Estimating progression-free survival in patients with glioblastoma using routinely collected data. *J. Neurooncol.* **135**, 621–627 (2017).
- Leao, D. J., Craig, P. G., Godoy, L. F., Leite, C. C. & Policeni, B. Response assessment in neuro-oncology criteria for gliomas: Practical approach using conventional and advanced techniques. *AJNR Am. J. Neuroradiol.* **41**, 10–20 (2020).
- Overcast, W. B. et al. Advanced imaging techniques for neuro-oncologic tumor diagnosis, with an emphasis on PET-MRI imaging of malignant brain tumors. *Curr. Oncol. Rep.* **23**, 34 (2021).
- Erickson, B. J., Korfiatis, P., Akkus, Z. & Kline, T. L. Machine learning for medical imaging. *Radiographics* **37**, 505–515 (2017).
- Juárez-Villegas, L. E., Altamirano-Bustamante, M. M. & Zapata-Tarrés, M. M. Decision-making at end-of-life for children with cancer: A systematic review and meta-bioethical analysis. *Front. Oncol.* **11**, 1–26 (2021).
- Liberati, A. et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration. *PLoS Med.* <https://doi.org/10.2427/5768> (2009).
- Richardson, W. S. Users' guides to the medical literature. VII. How to use a clinical decision analysis. A. Are the results of the study valid? Evidence-Based Medicine Working Group. *JAMA J. Am. Med. Assoc.* **273**, 1292–1295 (1995).
- Richardson, W. S. et al. Users' guides to the medical literature: VII. How to use a clinical decision analysis B. What are the results and will they help me in caring for my patients?. *JAMA J. Am. Med. Assoc.* **273**, 1610–1613 (1995).
- Borges, P. H. D., Lizar, J. C., Faustino, A. C. C., Arruda, G. V. & Pavoni, J. F. Kurtosis is an MRI radiomics feature predictor of poor prognosis in patients with GBM. *Braz. J. Phys.* **51**, 1035–1042 (2021).
- Chilaca-Rosas, M.-F., Garcia-Lezama, M., Moreno-Jimenez, S. & Roldan-Valadez, E. Diagnostic performance of selected MRI-derived radiomics able to discriminate progression-free and overall survival in patients with midline glioma and the H3F3AK27M mutation. *Diagnostics (Basel, Switzerland)* **13**, 849 (2023).
- Su, R., Liu, X., Jin, Q., Liu, X. & Wei, L. Identification of glioblastoma molecular subtype and prognosis based on deep MRI features. *Knowledge-Based Syst.* **232**, 107490 (2021).
- Sun, Y.-Z. et al. Differentiation of pseudoprogression from true progression in glioblastoma patients after standard treatment: A machine learning strategy combined with radiomics features from T(1)-weighted contrast-enhanced imaging. *BMC Med. Imaging* **21**, 17 (2021).
- Patel, M. et al. Machine learning-based radiomic evaluation of treatment response prediction in glioblastoma. *Clin. Radiol.* **76**(628), e17–628.e27 (2021).
- Pak, E. et al. Prediction of prognosis in glioblastoma using radiomics features of dynamic contrast-enhanced MRI. *Korean J. Radiol.* **22**, 1514–1524 (2021).
- Verma, R. et al. Tumor habitat-derived radiomic features at pretreatment MRI that are prognostic for progression-free survival in glioblastoma are associated with key morphologic attributes at histopathologic examination: A feasibility study. *Radiol. Artif. Intell.* **2**, 1–12 (2020).
- Yan, J.-L. et al. A neural network approach to identify the peritumoral invasive areas in glioblastoma patients by using MR radiomics. *Sci. Rep.* **10**, 9748 (2020).
- Kim, J. Y. J. H. et al. Incorporating diffusion- and perfusion-weighted MRI into a radiomics model improves diagnostic performance for pseudoprogression in glioblastoma patients. *Neuro Oncol.* **21**, 404–414 (2019).
- Chaddad, A., Sabri, S., Niazi, T. & Abdulkarim, B. Prediction of survival with multi-scale radiomic analysis in glioblastoma patients. *Med. Biol. Eng. Comput.* **56**, 2287–2300 (2018).

21. Sanghani, P., Ang, B. T., King, N. K. K. & Ren, H. Overall survival prediction in glioblastoma multiforme patients from volumetric, shape and texture features using machine learning. *Surg. Oncol.* **27**, 709–714 (2018).
22. Kebir, S. et al. A Preliminary study on machine learning-based evaluation of static and dynamic FET-PET for the detection of pseudoprogression in patients with IDH-wildtype glioblastoma. *Cancers (Basel)* **12**, 3080 (2020).
23. Zhou, H. et al. MRI features predict survival and molecular markers in diffuse lower-grade gliomas. *Neuro Oncol.* **19**, 862–870 (2017).
24. Zhuge, Y. et al. Brain tumor segmentation using holistically nested neural networks in MRI images. *Med. Phys.* **44**, 5234–5243 (2017).
25. Chang, K. et al. Multimodal imaging patterns predict survival in recurrent glioblastoma patients treated with bevacizumab. *Neuro. Oncol.* **18**, 1680–1687 (2016).
26. Chaddad, A., Desrosiers, C., Hassan, L. & Tanougast, C. A quantitative study of shape descriptors from glioblastoma multiforme phenotypes for predicting survival outcome. *Br. J. Radiol.* <https://doi.org/10.1259/bjr.20160575> (2016).
27. Artzi, M. et al. FLAIR lesion segmentation: Application in patients with brain tumors and acute ischemic stroke. *Eur. J. Radiol.* **82**, 1512–1518 (2013).
28. Mahajan, A. et al. Glioma radiogenomics and artificial intelligence: Road to precision cancer medicine. *Clin. Radiol.* **78**, 137–149 (2023).
29. Park, Y. W. et al. The 2021 WHO classification for gliomas and implications on imaging diagnosis: Part 2—Summary of imaging findings on pediatric-type diffuse high-grade gliomas, pediatric-type diffuse low-grade gliomas, and circumscribed astrocytic gliomas. *J. Magn. Reson. Imaging* **58**, 690–708 (2023).
30. Marquez, J. BraTS 2018 proceedings. *J. Dev. Econ.* **24**, 1–27 (1986).
31. Zhang, J. et al. Diffusion-weighted imaging and arterial spin labeling radiomics features may improve differentiation between radiation-induced brain injury and glioma recurrence. *Eur. Radiol.* **33**, 3332–3342 (2023).
32. Farzana, W. et al. Prediction of rapid early progression and survival risk with pre-radiation MRI in WHO grade 4 glioma patients. *Cancers (Basel)* **15**, 4936 (2023).

Acknowledgements

The authors would like to thank Dr. Roxana Pelayo and Dr. Laura Bonifaz for their support and Dr. Elsa de la Chesnaye for the fruitful discussion, and Carlos Alberto Flores for some art graphs, and Karen Werner and Perla Sueiras for editing and proof reading the manuscript.

Author contributions

M.M.A.B. and N.F.A.B., contributed to the conception, design, data extraction, investigation, epistemic meta-analysis, figure development, resources, management, quality assessment, data processing, drafting, and revising the manuscript of the study. M.F.C.R. contributed to the conception, resources, investigation, quality assessment, data processing, drafting, and revising the manuscript of the study. M.T.C.A. and D.R.S. contributed to data extraction, investigation, initial manuscript drafting, initial epistemic meta-analysis, and figure development. F.P.L. data extraction, investigation, manuscript drafting and polishing, final epistemic meta-analysis, figure development, and mind map scheme proposal. C.R.M., J.C.H.G., B.C.C., R.M.G. investigation, resources, and final quality assessment. All the authors read and approved the final manuscript.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-98058-0>.

Correspondence and requests for materials should be addressed to M.M.A.-B.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025