Journal of Big Data

RESEARCH Open Access

Automatic magnetic resonance imaging series labelling for large repositories

Armando Gomis-Maya[1,3*], Leonor Cerdá-Alberich[1*], Diana Veiga-Canuto[1], Salvatore Claudio-Fanni[4], Amadeo Ten-Steve[1], Gloria Ribas-Despuig[1], Pedro José Mallol-Roselló[1], Joan Vila-Frances[3] and Luis Marti-Bonmati[1,2]

*Correspondence: armago@alumni.uv.es; leonor_cerda@iislafe.es

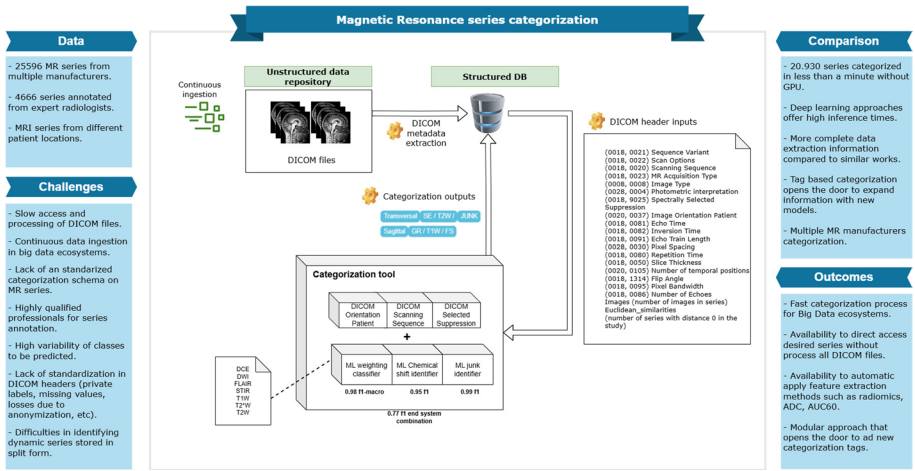
[1] Grupo de Investigación Biomédica en Imagen, Instituto de Investigación Sanitaria La Fe, Avenida Fernando Abril Martorell, Torre 106 A 7th Floor, 46026 Valencia, Spain
[2] Área Clínica de Imagen Médica, Hospital Universitario y Politécnico La Fe, Avenida Fernando Abril Martorell, Torre 106, 46026 Valencia, Spain
[3] Department of Electronics Engineering, Universitat de València, Avenida de Blasco Ibáñez 13, 46010 Valencia, Spain
[4] Department of Translational Research, Academic Radiology, University of Pisa, Lungarno Pacinotti 43, 56126 Pisa, Italy

Abstract

Large medical image repositories present challenges related to unstructured data. A data enrichment process allows the storage of additional information for fast identification of the content and properties of medical imaging studies. The aim of this study is to develop a metadata enrichment pipeline to facilitate the secondary use of medical images in a high-throughput environment. Our aim was to develop a categorization tool for the MR series to generate standardized tags that identify relevant image characteristics such as patient orientation, sequence type, weighting type, or the presence of fat suppression. Three models that make use of machine learning (ML) and DICOM tags are proposed. The dataset for their development consists of 4666 MR series from cancer patients, labeled by expert radiologists and acquired from different manufacturers, clinical centers and anatomical regions, covering as much variability as possible with the aim of making the models generalizable to other databases. Moreover, the inference performance of the end system has been evaluated on 25,596 MR series as well as the final model outputs with an external evaluation set of 1286 MR series. The weighting model achieves very reliable results with a f1 score of 88% in the validation set. Junk and chemical shift models achieved scores of 82% and 83% respectively. These results open the door to the automatic application of image post-processing and deep learning algorithms after accurate labeling, minimizing human intervention. Furthermore, the proposed solution can infer thousands of DICOM series in less than one minute. Thanks to the fast inference times provided by this solution, it fits well in a big data ecosystem, eliminating any performance issues on ingestion in a semi-real-time environment.

Keywords: Magnetic resonance, Categorization, Weighting identification, Medical imaging, Data mining, Radiology, Machine learning, Big data, DICOM, Metadata

Graphical Abstract



Introduction

Machine learning (ML) capabilities are being used to foster research in clinical environments [1–3]. One area that has particularly benefited is radiology. The type of handled data and the increasing improvement of the results in tasks such as segmentation, classification, or object detection, make medical imaging one of the clinical areas that is growing the most in this technological revolution [4]. For the development and operation of AI technologies, it is essential to create data repositories ingested with the highest possible volume of quality annotated data. The creation of large projects in which multicentric imaging data is collected and harmonized has been encouraged in recent years. Relevant examples of research projects involving cancer patients, imaging data and AI-based algorithms are PRIMAGE [5], CHAIMELEON [6], ProCancer-I [7], Incisive [8], and EuCanImage [9]. These projects are challenging due to the partially unstructured data obtained from the real-world environment.

Medical images are stored in a standardized file format (DICOM) where some partially unstructured metadata fields are linked to their corresponding images in terms of their pixel intensity values. When it is desired to collect those images that meet a defined condition (e.g., the presence of a tumor), each DICOM file needs to be inspected to verify if the condition is fulfilled, iterating through all the data and its associated metadata in the repository. This inefficient process could be improved by the creation of a separate table to store such information. In big data repositories, manual extraction of this information is a challenging and tedious task [10, 11].

Therefore, data indexing and data enrichment is a crucial aspect to improve data readiness. In medical imaging, the identification and categorization of MR weighting and sequence is a relevant and challenging mining operation. As an example, the automatic extraction of the Apparent Diffusion Coefficient (ADC) requires the previous

identification of the specific diffusion-weighted sequence [1]. Usually, weighting information is manually entered by technicians with non-standard nomenclature as free text in the DICOM series metadata, which leads to problems of interpretation by automatic processing tools. Therefore, using series names or DICOM series description tags is not an option in large data repositories from different clinical centers using other languages, where anonymization algorithms are used and sometimes destroy such information.

Previous works

Several studies [12–14], have underscored the necessity and utility of developing machine learning (ML) models that automatically annotate sequences in magnetic resonance images (MRI) within large clinical image repositories. These works propose methodologies for training deep learning models using convolutional neural networks (CNNs) directly on images treated as volumetric matrix data. This approach is agnostic to the metadata attached to the images and does not address the lack of standardization in DICOM headers across different manufacturers, nor the issues arising from anonymization algorithms that remove important header information necessary for model creation. These solutions are the only ones capable of reliably identifying contrast in MRI series, as the associated DICOM header (Contrast/Bolus Agent) is not always available [13].

Alternatively, solutions that use DICOM metadata as inputs [15–18] typically offer very short inference times and require less computational power, eliminating the need for GPUs, since metadata processing is simpler and lighter. The main challenge of this approach lies in the extensive and intricate processing required to address the lack of standardization among MRI scanner manufacturers, particularly when different private headers are employed to represent the same information (e.g., b-values in DWI series). This involves handling missing values resulting from optional DICOM headers and interpreting headers written in natural language—such as "Series Name" and "Series Description"—by radiologists from various centers and countries. Furthermore, the extraction of image files from PACS systems can introduce complications during anonymization processes (e.g., deleting DICOM headers), potentially hindering the reconstruction of dynamic series like DCE by storing time volumes individually.

Assembling MRI sequence datasets that encompass extensive heterogeneity in anatomical regions, varying machine manufacturers and comprehensive labeling of sequence types and weightings—including both static and dynamic series—is an arduous task. Among the cited works, only [15] addresses a wide range of weighting labels beyond the cerebral anatomical region; the remaining studies focus on a limited subset of static series pertaining to the most common sequence types. The creation of highly heterogeneous MRI datasets is particularly challenging in Europe due to stringent data confidentiality restrictions.

The datasets used introduce an intrinsic limitation in the development of ML models. Without training on certain sequences and weighting types, these models cannot generalize effectively and may produce erroneous inferences when applied to new, unseen series.

Proposed work

This paper proposes an automatic and modular approach for the categorization of MRI series using machine learning methods applied to an MRI database not previously utilized, derived from the PRIMAGE project [5]. The use of this novel dataset has introduced new challenges, such as confusion arising from storing dynamic DCE series separately from individual static T1 series and the need for expert radiologists to annotate new sequence types and weightings not previously considered—namely DCE, DWI, STIR and CHS. The dataset employed encompasses a wide range of anatomical regions, is multicentre and includes images acquired from different scanner manufacturers.

One of the challenges is managing the anonymization process applied to the images, which removes relevant DICOM headers. Another challenge is avoiding reliance on non-standardized headers like "Series Name" and "Series Description" due to the multicentre and multilingual nature of the data. Additionally, this work proposes a technique to facilitate the identification of dynamic sequences such as DCE and DWI when they are stored in a split manner as multiple static volumes.

This research aims to enhance the application of segmentation techniques and radiomic feature extraction in large image repositories by efficiently identifying relevant information within images. By meeting the input requirements of these techniques and expanding the sample size of data available for dataset creation, we facilitate more robust analyses. We have developed a scheme that identifies critical parameters—including sequence type, weighting, fat suppression and other factors deemed important for such applications—while establishing a new labeling nomenclature that summarizes the pertinent information. This approach allows for rapid and efficient identification of MRI series, benefiting both radiologists and MRI data processing analytics.

Three machine learning models have been developed to perform automatic labeling in accordance with the schema defined by radiologists. The first model is a multi-label classifier for MRI sequence weightings, considering the most common types observed in related studies, such as T1-weighted (T1W), T2-weighted (T2W), T2*-weighted (T2*W) and FLAIR (Fluid-Attenuated Inversion Recovery, specific to the cerebral region). Additionally, it incorporates weightings like STIR (Short Tau Inversion Recovery), which typically represent anatomical regions of the extremities and the spine, along with dynamic series such as DCE (Dynamic Contrast-Enhanced) and DWI (Diffusion-Weighted Imaging), which have significant sample sizes in the PRIMAGE database.

A second binary model is capable of labeling images exhibiting Chemical Shift (CHS), which is not a weighting type per se but a physical phenomenon arising from differences in the resonance frequencies of water and fat protons. This phenomenon is utilized in MRI to generate contrast and provide additional information about tissue characteristics, particularly when differentiating between fat and water areas in the body. Therefore, an independent classifier separate from the weighting classifier has been developed and its labels will incorporate information from DICOM tags regarding fat suppression.

The inclusion of this information, along with fat suppression data, is particularly significant because it can fundamentally alter the interpretation of pixel intensity values within the image matrix. Such alterations can lead to erroneous inferences in machine learning models that have been trained exclusively on T2-weighted images without fat suppression. Introducing a T2-weighted image with fat suppression changes the

intensity values, potentially resulting in unexpected outcomes. Therefore, accounting for fat suppression is essential to prevent inaccuracies in ML models and to enhance their robustness and reliability in clinical application.

The third model identifies "junk" series or those unsuitable for clinical applications or further machine learning analyses. This model is responsible for detecting calibration scans, localizers, screenshots and images with a limited number of slices, rendering them non-informative.

Ultimately, a post-processing algorithm combines the relevant DICOM tags with the predictions from these models to align with the labeling schema defined by radiologists, ensuring comprehensive and accurate tagging for future clinical and analytical use.

Materials and methods

This section outlines the materials and methods used to establish a structured system for MRI data categorization and annotation within the PRIMAGE project, part of the EU's AI4HI cluster. The project utilizes a categorization scheme developed with radiologist input, covering imaging orientation, sequence type, weighting, and suppression methods.

Key stages in this approach include comprehensive data curation, manual annotation and feature engineering, followed by specific transformations and feature selection to refine model inputs. We address data variability challenges, particularly missing DICOM tags, by using the CatBoost algorithm, which is robust in handling missing data and efficient for large-scale datasets. This method provides a reliable, scalable solution for pediatric oncology imaging, supporting data quality essential for predictive modeling and enhancing the precision of MRI annotations.

Data

PRIMAGE (Grant Agreement No. 826494) is a project within the Artificial Intelligence for Health Imaging (AI4HI) cluster—a consortium of EU-funded research initiatives dedicated to developing cancer imaging data repositories and AI solutions based on medical imaging to enhance clinical practice [19]. Other key projects in this cluster include CHAIMELEON [6], ProCancer-I [7], INCISIVE [8] and EuCanImage [9]. PRIMAGE focuses on predictive in-silico multiscale analytics to support personalized diagnosis and prognosis in cancer, empowered by imaging biomarkers. The project collects retrospective data from multiple clinical partners on childhood cancers, specifically Neuroblastoma and diffuse intrinsic pontine glioma (DIPG). While DIPG imaging is confined to the brain, Neuroblastoma can manifest in various regions, including the abdominal, thoracic and pelvic areas. Consequently, the study images encompass a diverse range of patient locations and MRI manufacturers, as illustrated in Fig. 1.

In the PRIMAGE project, all DICOM metadata are extracted during the ingestion phase and stored in a semi-structured MongoDB database [20, 21]. During this phase, a rigorous anonymization process is applied to remove all DICOM headers that could be used to re-identify patients and compromise their privacy. However, this process may result in the loss of important DICOM headers, potentially hindering the performance of subsequent training processes. Furthermore, since each clinical center may apply its

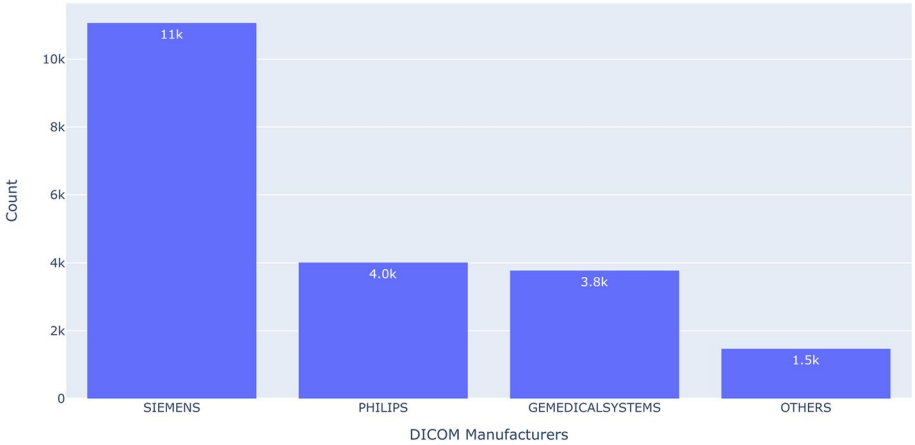


Fig. 1 Number of DICOM series per manufacturer in the PRIMAGE database

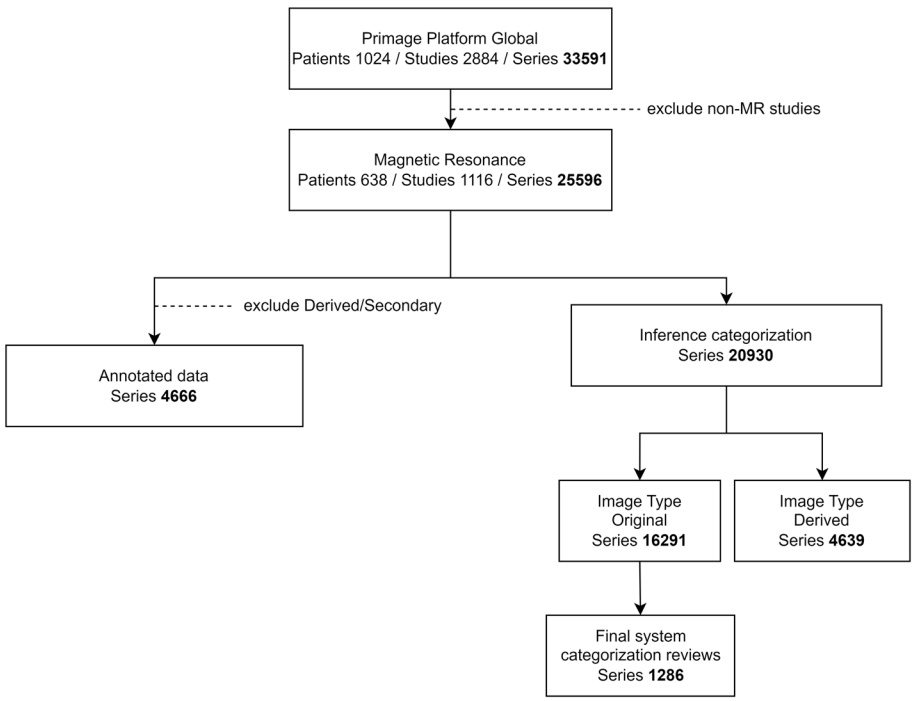


Fig. 2 Diagram of the study design in terms of the MR series distribution

own anonymization process prior to the ingestion phase according to its preferences, this can exacerbate the issue by removing additional essential data and increasing variability in the dataset. At the time of developing the categorization tool, the platform contained 25,596 MR series, of which 4666 had been manually annotated by radiologists for model development and 1286 for final system evaluation. The remaining series were used to evaluate the final solution's performance during inference. Figure 2 presents a comprehensive diagram of the MR series distribution.

Categorization scheme definition

A categorization scheme is defined in combination with expert radiologists. That includes location, type of sequencing, type of weighting and information about the use of different suppressions. The information contained in the categorization is helpful to get an overview of the content of the series and covers all preconditions necessary for the post-processing and feature extraction analyses [27].

To comply with this scheme defined in Fig. 3, three different ML models have been developed (weighting classifier, chemical shift detector and junk detector) together with a final algorithm that merges model outputs with raw values of DICOM headers.

This scheme is founded on two primary criteria. First, it is preferable to extract information from DICOM tags, which offer 100% confidence, rather than relying on models that incorporate probabilistic approaches. Second, machine learning models should possess a singular objective and maintain simplicity. For example, employing two separate models that individually classify weighting, and junk is more effective than using a single model that simultaneously classifies both. The latter approach would double the number of target classes, potentially introducing confusion and diminishing classification performance.

The DICOM headers used in the schema of Fig. 3 to obtain the non-ML tags are "Image Orientation Patient" (0020, 0037), "Scanning Sequence" (0018, 0020), "Spectrally Selected Suppression" (0018, 9025) and "Scan Options" (0018, 0022).

It is relevant to note that one of the possible values that the "Scanning Sequence" DICOM tag can adopt, namely "Research Mode", does not provide useful information for the MR series categorization. Therefore, these cases are treated as empty values in this DICOM tag, forcing the experts to manually label the scanning sequence field when this occurs.

In the "surnames" section of the schema, a combination of two different DICOM tags to detect the possible suppression in a series is applied. This is due to the different casuistry observed in the data around empty values of the two relevant tags

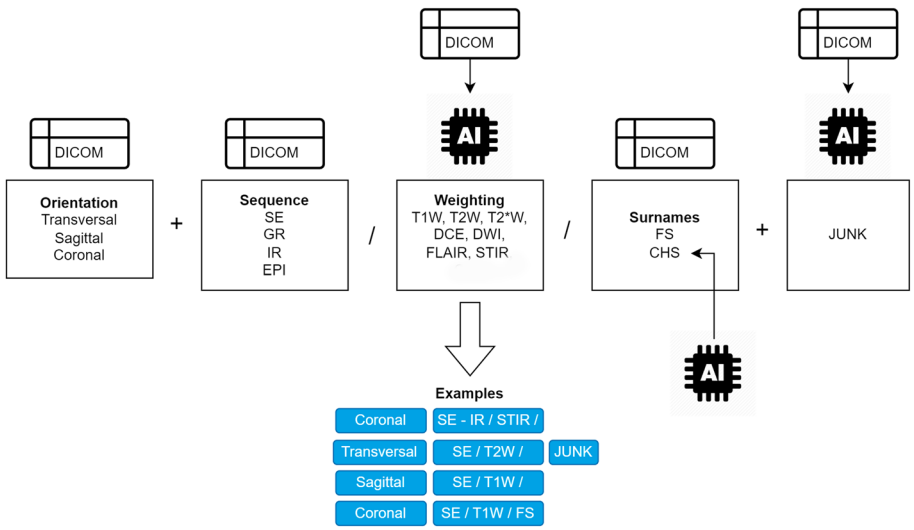


Fig. 3 Categorization schema for the labeling of MR series

("Spectrally Selected Suppression", in all cases and "Scan Options", when adopting the "FS" value). By combining them is it possible to identify the suppression in a larger set of cases.

Annotation

The repository contains not only original or primary images obtained directly from the PACS but also secondary images, such as ADC, T1 and T2 maps, which necessitate filtering to optimize radiologists' annotation efforts.

To ensure efficient manual annotation, specific exclusion criteria were applied to filter out these secondary images. First, the DICOM image type property (0008, 0008) was used to exclude images marked with "Derived" or "Secondary" values, identifying them as post-processed. Additionally, a pattern-matching approach was applied to the series description attribute (0008, 103E) for images containing this DICOM header, filtering for terms such as "ADC", "preprocessed", "registered" and "map." In both scenarios, these series were labeled as "Derived" (see Fig. 2) and subsequently excluded from model training.

Following the exclusion of secondary images, radiologists conducted semi-automatic labeling of the remaining 4666 DICOM MR series, overseeing a traditional pattern-matching algorithm that classifies the weighting type based on the series description header.

Table 1 displays the frequency of the manually labeled samples used to create the ML models. The annotations are used to develop three different ML models (weighting classifier, chemical shift detector and junk detector), but for their development not all the data in Table 1 is used as each particular model requires a specific set of annotations.

To develop a binary junk model capable of identifying irrelevant data within the database, MR series categorized as "JUNK" were assigned as the positive class, with all other data labeled as the negative class. Similarly, for the chemical shift model, data associated with the "CHS" category were designated as the positive class, while the remaining data

Table 1 Sample size by MR series category in dataset

Weighting	Number of manually labeled samples
T2W	1326
T1W	1256
DCE (dynamic contrast enhance)	496
JUNK (localizer, calibration, others)	484
DWI (diffusion weighted)	480
CHS (chemical shift)	236
STIR (Short Tau Inversion Recovery)	236
FLAIR (fluid attenuated inversion recovery)	113
T2*W	36
PDW (proton density weighted)	2
SW (susceptibility weighted)	1

*It's a variant of T2-weighted imaging but incorporates magnetic field inhomogeneities and dephasing effects

were classified as the negative class. This model will further enrich categorization within the classification scheme.

The MRI series used to develop the weighting model comprises a range of manually annotated, classified samples, including T2W, T1W, DCE, DWI, STIR, FLAIR and T2W. Categories "PDW" and "SW" contained insufficient samples for training and were therefore excluded from the weighting dataset. Likewise, the "T2*W" category also had too few samples to develop a robust model. Consequently, this category was consolidated under the "T2W" label, with a subsequent a posteriori classification by checking if the DICOM tag "Scanning Sequence" adopted the gradient echo (GR) value. Where this condition was met, the "T2W" label was reassigned as "T2*W."

Models

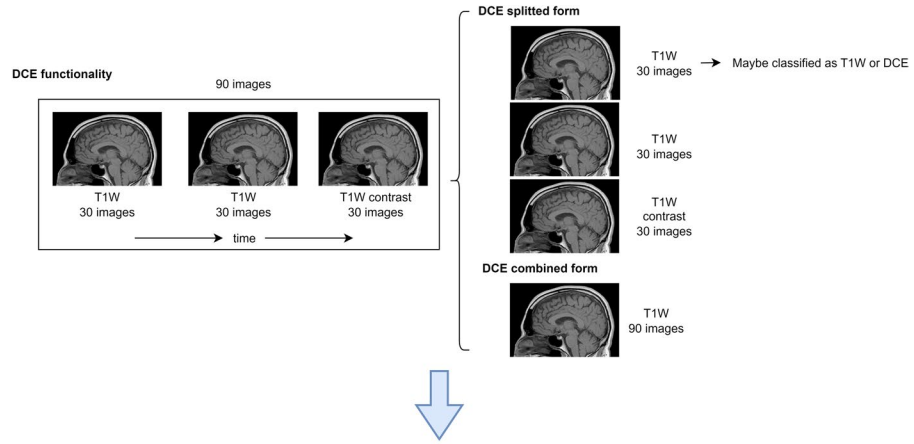
This section presents the implementation of the classification models, specifying the DICOM headers used as input features across the three models, the transformations applied to each input feature, and the algorithm employed to develop the final machine learning models. Furthermore, a technique for creating a new feature, named "Euclidean_similarities" is introduced to improve the models' ability to classify dynamic sequences more effectively.

Input features

The input features utilized in the ML models developed in this study include a combination of features employed in previous works [15, 16] and additional features suggested by PRIMAGE radiologists. Only machine-generated features with a high prevalence across the dataset were selected, specifically those with less than 15% missing values across all series. In addition to the previously noted DICOM metadata, two engineered features were incorporated: the number of images within a series ("Images") and the count of series with a Euclidean distance of zero within the same MR study ("Euclidean_similarities"). A comprehensive list of input features is provided in Table 2.

Table 2 Input features for ML model development

Categorical features	Numerical features
(0018, 0021) sequence variant	(0018, 0081) echo time
(0018, 0022) scan options	(0018, 0082) inversion time
(0018, 0020) scanning sequence	(0018, 0091) echo train length
(0018, 0023) MR acquisition type	(0028, 0030) pixel spacing
(0008, 0008) image type	(0018, 0080) repetition time
(0028, 0004) photometric interpretation	(0018, 0050) slice thickness
(0018, 9025) spectrally selected suppression	(0020, 0105) number of temporal positions
(0020, 0037) image orientation patient	(0018, 1314) flip angle
	(0018, 0095) pixel bandwidth
	(0018, 0086) Number of Echoes, (Optionally studied (0019, 10A9)
	Images (number of images in a series)
	Euclidean_similarities (number of series with Euclidean distance equal to 0 in the same study)



Solution to detect splitted forms: Euclidean distance within study

Name	Sequence	Orientation	EchoTime	FlipAngle	RepetitionTime	...	EuclideanSimilarities
AXIALT2FATSAT384NAVEG	SE	Axial	102.00	150.0	2000.00		0
T2-COR-STIRLUMBAR	['SE', 'IR']	Coronal	71.00	150.0	3300.00		0
T2-STIR-SAGITALLUMBAR	['SE', 'IR']	Sagital	72.00	150.0	3300.00		0
VIBE-IPAT	GR	Axial	3.56	10.0	7.55		3
VIBE-IPAT	GR	Axial	3.56	10.0	7.55		3
VIBE-IPAT	GR	Axial	3.56	10.0	7.55		3
VIBE-IPAT	GR	Axial	3.56	10.0	7.55		3

Fig. 4 Visualization of storage problems encountered in dynamic series. The Euclidean distance is shown as the "Euclidean Similarities" variable: it adopts a value equal to 0 when the MR series is a unique acquisition and a value equal to N (different than 0) when it belongs to a composite series, being N the number of series that fulfill the criteria of having an Euclidean distance equal to 0 with respect to a particular MR series

Dynamic MRI series represent the aggregation of individual series into a unified set with a specific analytical purpose. For instance, the DCE series combines multiple T1W series over time to observe the effect of contrast enhancement as a temporal function. While a model can readily identify a dynamic series based on image count when the data is stored in its combined form, detection becomes challenging when series are stored in split form. As illustrated in Fig. 4, identifying whether a T1W series is part of a DCE combination or a standalone T1W acquisition is complex under these circumstances.

The "Euclidean_similarities" feature addresses this challenge by detecting composite series stored in a split manner within the same study. Typically, composite series retain identical acquisition parameters across internal series, with only minor variations in attributes such as contrast or time. To identify these similarities, a selection of variables is used to calculate the Euclidean distance between series, yielding the count of identical series within the same study. The parameters employed for this Euclidean distance calculation include "Scanning Sequence", "Orientation Patient", "Echo Time", "Flip Angle", "Repetition Time", "Pixel Bandwidth", "Number of Images", "Number of Temporal Positions" and "Slice Thickness".

Additional DICOM tags were evaluated but ultimately excluded due to system anonymization constraints; however, these may be worth revisiting in future studies. One promising tag is the diffusion b-value (0018, 9087) of the first and last images within a series, which could offer an alternative means of classifying DWI series.

Table 3 Transformations of input features

DICOM features	Preprocessing
(0018, 0021) sequence variant	One hot encoding for values SK, SS, SP, OSP
(0018, 0022) scan options	One hot encoding for values PER, RG, CG, PPG, FC, PFF, PFP, SP, FS
(0018, 0020) scanning sequence	One hot encoding for values SE, GR, IR, EP, RM
(0018, 0023) MR acquisition type	Categorical label encoded
(0008, 0008) image type	One hot encoding for values DERIVED, SECONDARY, LOCALIZER, ADC, SCREENSAVE
(0028, 0004) photometric interpretation	Binary value depending if contains RGB
(0018, 9025) spectrally selected suppression	Binary value depending if contains FAT
(0020, 0037) image orientation patient	Transformation from numeric vectors to categories Sagittal, Axial, Coronal One hot encoded transformation
(0018, 0081) echo time	Round the two numeric values to the first decimal Label encoded transformation
(0018, 0082) inversion time	Set empty values simulating infinity with 100,000,000 value
(0018, 0091) echo train length	–
(0028, 0030) pixel spacing	Categorical label encoded
(0018, 0080) repetition time	–
(0018, 0050) slice thickness	–
Number of images in series (custom feature)	–
(0018, 1314) flip angle	Set empty values simulating infinity with 100,000,000 value
(0018, 0095) pixel bandwidth	Set empty values simulating infinity with 100,000,000 value
(0018, 0086) number of echoes	Set empty values simulating infinity with 100,000,000 value
(0020, 0105) number of temporal positions	Set empty values simulating infinity with 100,000,000 value

SK: segmented k-space; SS: steady state; SP: spoiled; OSP: oversampling phase; PER: phase encode reordering; RG: respiratory gating; CG: cardiac gating; PPG: peripheral pulse gating; FC: flow compensation; PFF: partial Fourier—frequency; PFP: partial Fourier—phase; SP: spatial presaturation; FS: fat suppression; SE: spin echo; GR: gradient echo; IR: inversion recovery; EP: echo planar; RM: research mode; RGB: red; green; blue; FAT: fat suppression

Transformations

As commonly done when training ML algorithms, several feature transformations have been applied to maximize the learning in trained models. Table 3 details the transformations applied to each feature.

The transformations applied to the target variable vary according to the model being developed. For the weighting model, categorical data were processed using label encoding to convert each category into a unique numerical label. In contrast, for the junk classifier, a binary encoding was implemented, assigning a value of "1" to the "JUNK" category and "0" to all other categories. This binary encoding approach was similarly applied to the chemical shift model, where data in the "CHS" category were assigned a "1," with the remaining classes assigned a "0."

After applying the transformations, a total of forty-three transformed features were generated. To identify the most relevant features and minimize the number of inputs for model training, the Boruta feature selection algorithm [22] was utilized. Given that Boruta is a supervised algorithm, feature selection was conducted independently for each model (weighting, junk and chemical shift) to optimize relevance and accuracy within each specific model context.

The Boruta algorithm has been chosen over other alternatives because it is a feature detection algorithm that does not require input parameters and thus avoids the user having to make decisions regarding the cut-off threshold for separating the features to keep from the ones to remove. Boruta's only requirement for its application is the choice of the ML algorithm to be used internally for selection. Typically, a *RandomForest* is used, due to its robustness to data with anomalies or noise and its ability to detect non-linear relationships between features [23].

Boruta outputs a classification into three groups of features according to the level of certainty in keeping the variables in a predictor. Only the features of the group with certainty to be eliminated have been filtered out, thus keeping the features recommended to be kept and the features of which there is no certainty to be kept or eliminated. This has resulted in a total of 26 features selected for each model.

Algorithms

A significant challenge in the PRIMAGE project is managing the anonymization process applied to images, which removes critical DICOM headers, as well as handling additional anonymization procedures performed by the PACS systems of individual hospitals. Furthermore, since the dataset contains images from various manufacturers, essential information is often stored within private DICOM metadata fields. These fields are particularly targeted by anonymization algorithms, as they frequently contain sensitive data. It is therefore especially important to employ an approach that is able to deal with missing values.

Traditional classification algorithms are typically unable to handle datasets with missing values, often requiring the addition of imputation models capable of identifying and filling data gaps. However, integrating an imputation model increases the system's overall complexity and leads to longer inference times per sample. In large-scale data environments, especially when processing computationally demanding medical images, minimizing inference latency is essential. To address this, the CatBoost algorithm [24] was selected, given its native capacity to manage missing data, rapid inference capabilities on both CPU and GPU and strong performance in detecting non-linear relationships.

This choice followed iterative optimizations comparing a Random Forest model with a K-Nearest Neighbors (KNN) imputer against a CatBoost classifier [25]. The low variability in the scores and the high performance in the obtained results were the main reasons in choosing this feature selection and model combination in this study, resulting in a simple, stable, robust and explainable tool.

Results

The proposed approach is designed as a modular system comprising three distinct models that are ultimately integrated into a unified framework. The evaluation methodology for this system includes two key components: first, assessing the performance of each individual model using classical train-test partitions to obtain baseline scores and second, testing the performance of the combined model on an independent dataset as external evaluation not utilized in the initial phase. Notably, the weighting model operates as a multiclass classifier, while the other two models are binary classifiers. Due to class imbalances, f1 scores are employed as the primary evaluation metric, providing

Table 4 Weighting classifier evaluation results by category

Class	Precision	Recall	F1	Support
DCE	0.98	0.98	0.98	99
DWI	0.99	0.96	0.97	96
FLAIR	0.96	1.00	0.98	23
STIR	0.98	0.96	0.97	47
T1W	0.99	1.00	1.00	126
T2W	0.99	1.00	0.99	133

Fig. 5 Weighting model evaluation with test dataset confusion matrix

a balanced measure that accounts for both precision and recall across uneven class distributions.

Weighting model results

A 90–10 data partition was conducted to train the Weighting model, with stratification based on the frequency of each class. Specifically, 90% of the data was allocated

for training and 10% for testing. This methodology addresses a multi-label classification problem involving predictions across six different types of potentiation. The test results yielded a f1 score (macro) of 98%. Table 4 presents the individual results for each class within the test partition. The predictions demonstrated remarkable homogeneity and the performance metrics were quite similar across classes including dynamic weightings, despite the existing class imbalance.

An examination of the confusion matrix in Fig. 5 reveals that the FLAIR, T1W and T2W series are perfectly classified, while the dynamic series (DCE and DWI) exhibit a few misclassifications, confirming the increased difficulty in classifying these more complex categories.

Regarding feature importance within the weighting model, Fig. 6 illustrates that "echo time", "repetition time" and "flip angle" are critical parameters in defining image weighting during the acquisition process [25, 26]. Therefore, it is expected that these features are among the most predictive for this model. Notably, "pixel bandwidth" emerges as the second most important feature. This may be attributed to its direct relationship with voxel size and image resolution. A well-known trade-off exists between spatial and temporal resolution: in dynamic series, shorter acquisition times result in lower achievable resolution, whereas non-composite series can be acquired at higher resolutions due to the absence of temporal information requirements.

Furthermore, the categorical binary variable "scanning sequence" when adopting the inversion recovery (IR) value is identified as one of the most significant features in the model. The IR "scanning sequence type" is exclusively used in STIR and FLAIR series, which justifies the high importance of this variable, as it assists in discriminating these types of weightings. This suggests that the model's behavior is coherent and aligns with expected theoretical considerations.

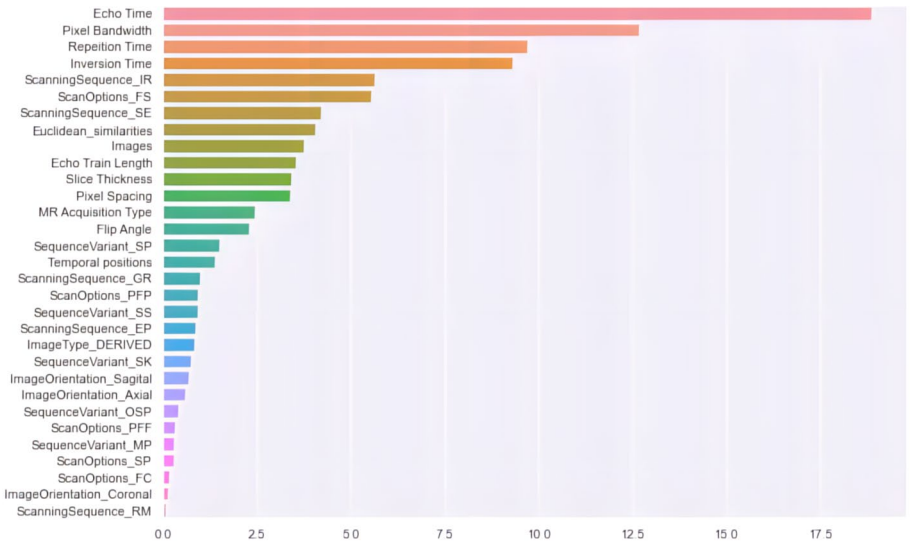


Fig. 6 Feature importance of weighting classifier

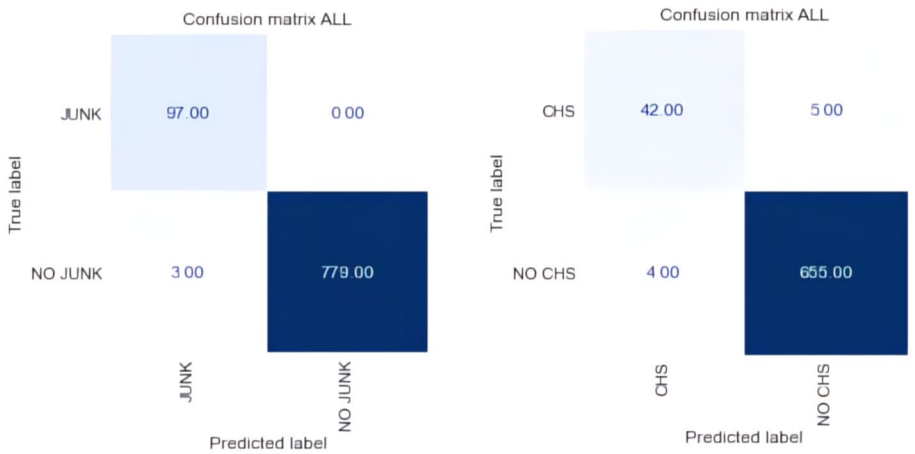


Fig. 7 Evaluation results of the (left) junk and (right) chemical shift models in terms of the confusion matrices per category (model class—binary) with the test data set

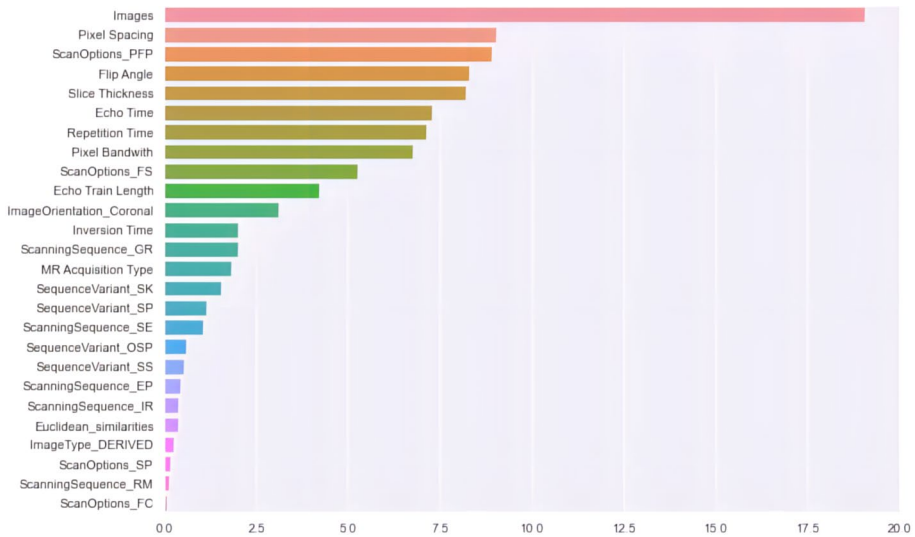


Fig. 8 Feature importance of Junk classifier

Junk and chemical shift model results

The junk and chemical shift models are binary classifiers. For these models, all annotations of the non-target class are set to zero, while the "JUNK" or "CHS" class annotations are set to one, respectively. The same 90–10 train-test partition proportion used in the weighting classifier is applied here. The evaluation of the junk classification model on the test dataset yields strong performance, achieving an accuracy of 100% and an f1 score of 99%. An examination of the confusion matrix (Fig. 7, left) reveals that only three misclassifications occurred out of eight hundred and seventy-nine samples evaluated, corresponding to an error rate of 34%. These results demonstrate the model's effectiveness and its potential for integration into a combinatory model for the final solution developed in this study.

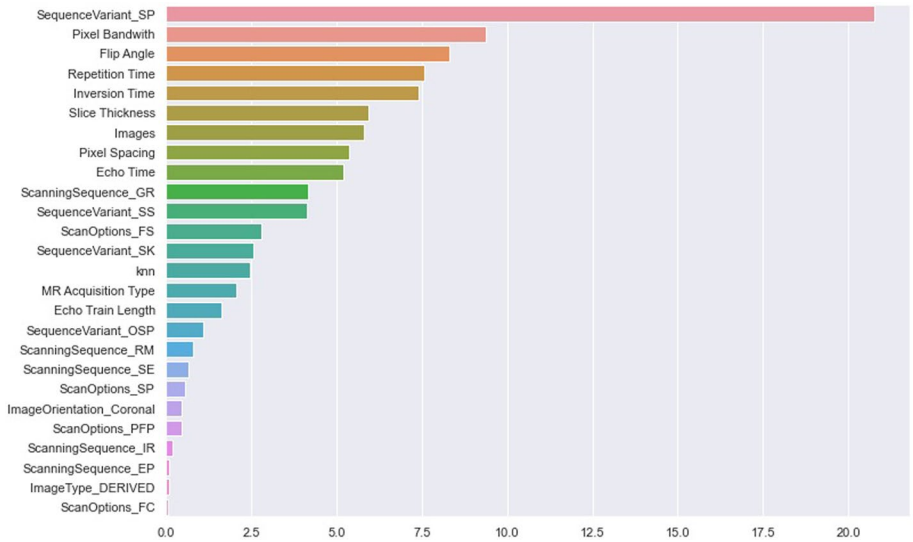


Fig. 9 Feature importance of CHS classifier

In terms of the importance of the features of the junk model (Fig. 8), it can be seen that the number of images present in the series is of great relevance. This observation is coherent with the fact that localizers and calibrations, which constitute a large proportion of the junk data in the study, tend to have a small number of images. The feature "*ScanOptions_PFP"* is used when technicians aim to shorten acquisition times, so it is also expected that localizers and calibrations may contain this particular DICOM tag and therefore, to be discriminative. The same applies to the *Pixel Spacing* feature, which includes valuable information regarding image resolution and voxel size.

In parallel, the feature importance of the CHS classifier, shown in Fig. 9, highlights "SequenceVariant" with the value "SP" (spoiled gradient echo) as the most predictive feature. This strong predictive capacity can be attributed to several factors. Firstly, "SP" sequences are specifically designed to reduce chemical shift artifacts, a crucial factor in distinguishing between fat and water tissues. The spoiled gradient echo sequences are particularly effective in minimizing these artifacts, allowing for clearer differentiation between fatty and aqueous tissues, a distinction vital for accurate chemical shift classification.

Moreover, SP sequences uniquely modify signal phase and echo timing, thereby enhancing the contrast between tissues with varying chemical compositions, such as fat and water. This feature manipulation aids in highlighting tissue characteristics that are essential for the model's classification tasks, making "SP" a highly predictive value within the CHS classifier. Finally, SP-marked sequences often employ specific repetition time and echo time parameters that further optimize the visualization of chemical shifts. These tailored imaging parameters are instrumental in capturing subtle variations between fat- and water-based sequences, thereby reinforcing the importance of "SP" in the classification of chemical shift variations.

Table 5 Evaluation results of the weighting classifier in terms of metrics per category (model class) with the external validation data set and performed in a production environment

Class	Precision	Recall	F1	Support
DCE	0.89	0.85	0.87	158
DWI	0.79	1.00	0.88	76
FLAIR	1.00	0.85	0.90	11
STIR	1.00	0.34	0.51	71
T1W	0.94	0.96	0.95	460
T2W	0.83	0.92	0.87	226

Table 6 Results of the end system evaluation in terms of the f1 score and the accuracy metrics per model (weighting, chemical shift and junk) and for the final model combination with the external validation data set and performed in a production environment

Model	F1	Accuracy
Weighting	0.88 (weighted)	0.89
Chemical shift	0.83 (standard)	0.98
Junk	0.82 (standard)	0.88
Combined	0.77 (macro)	0.88

The results obtained with the chemical shift model are marginally worse than the previous models, resulting in a f1 score of 95% and an accuracy of 99%. Looking at the evaluation results of the model in terms of the confusion matrix (Fig. 7), nine failures out of the six hundred and ninety-seven samples evaluated are observed (corresponding to a total of 1.29% of errors). The main reason for this slight decrease in performance is the unbalanced nature of the data employed to develop this model, as the number of cases labeled as chemical shift constitute only 9.32% of the total data set. Nevertheless, this model will also be used as part of the final combination model as it can help with additional information into the complete categorization process.

External validation in production

Following the individual training and evaluation of the models, after the individual training and evaluation of the models, an additional evaluation phase has been carried out to assess their integrated performance according to the scheme defined in Fig. 3. To facilitate this evaluation, we developed and implemented a prediction tool for MR series categorization within the PRIMAGE platform. Radiologists employed this tool to validate one thousand two hundred and eighty-six MRI series, of which two hundred and eighty-four were identified as unsuitable for model evaluation and were consequently excluded from further analysis.

Table 5 presents the evaluation results of the weighting model across various metrics for each individual class. Most series types achieve f1 scores close to 90%. However, STIR series are often misclassified as FLAIR or T2W series due to the subtle differences between these weightings in the DICOM headers.

Table 6 provides a comprehensive summary of the external evaluation results for each individual model and the combined algorithm, following the defined scheme. The weighting model achieved an accuracy of 89% and an f1 score (weighted by frequency) of 88%. The junk model attained an accuracy of 88% and an f1 score of 0.82, while the chemical shift model reached an accuracy of 98% and an f1 score of 83%. All models exhibited a decrease in performance metrics compared to the training and testing phase results, as anticipated, due to the use of less curated data. In this final evaluation, images with missing header information, from different manufacturers and originating from various centers were not filtered or adapted to replicate performance as closely as possible to real clinical practice. New images added to the repository are those that have received annotation predictions and have been evaluated by radiologists.

In the combined solution, a prediction is considered successful only when all three individual model predictions are simultaneously correct. The cumulative errors from the three models led to a decrease in the overall performance metrics, yielding an accuracy of 88% and an f1 score of 77%.

Discussion

Artificial intelligence (AI) and big data technologies are significantly enhancing the field of medical image repositories, facilitating the initiation of new research projects [28]. Numerous studies have demonstrated how feature extraction from medical images can be leveraged to develop clinical predictive models and to better understand the behavior and heterogeneity of various pathologies [1, 29, 30]. However, automating the extraction of magnetic resonance (MR) imaging features requires prior knowledge of specific information, such as the MR sequence type and dynamic acquisition parameters.

One of the major challenges in medical imaging research is the scarcity of large and diverse datasets. It is particularly difficult to find extensive collections of medical images and most similar works employ neuroimaging datasets, which are limited to the modalities used in that area [15]. These datasets often focus on a single anatomical region—the brain—owing to the availability of publicly accessible data predominantly originating from neuroimaging studies. This reliance restricts the types of sequences employed and limits the variability of labels. In contrast, our study utilizes a unique and comprehensive data repository that includes not only images from the brain region of patients with diffuse intrinsic pontine glioma (DIPG) but also images from different anatomical regions obtained from patients with neuroblastoma, who may present tumours in various parts of the body. This diversity enhances the robustness of our models and broadens their applicability across different clinical scenarios. For example, T1-weighted and T2-weighted images are commonly utilized to extract radiomic features, while diffusion-weighted imaging (DWI) and dynamic contrast-enhanced (DCE) sequences are well-established for obtaining apparent diffusion coefficient (ADC) and AUC60 parametric maps, respectively. AUC60 is defined as the area under the time-activity curve of gadolinium contrast over sixty seconds from the start of contrast enhancement. Traditionally, identifying which MRI series correspond to each relevant sequence necessitates a manual selection process performed by highly qualified radiologists, making data selection tedious, challenging and time-consuming.

The lack of diversity in publicly available datasets makes direct comparisons with other works challenging. Most existing research does not employ annotations for Short Tau Inversion Recovery (STIR) weightings and only study [18] utilizes dynamic series such us perfusion (DCE) and diffusion (DWI) weightings. Moreover, that study relies on a rule-based annotation system using DICOM tags containing natural language descriptions such as “Series Name” or “Series Description.” This approach is impractical in multi-centre and multinational clinical practice settings due to the lack of standardization among practitioners in composing these headers. For instance, a German radiologist may write the "Series Description" differently from a Spanish radiologist, as they often do not adhere to the same nomenclature standards. Additionally, this method depends on the presence of the DICOM header "Series Description"; without it, the annotation process cannot be performed. In this work, aggressive anonymization processes often remove these headers, which would result in a significant number of unlabelled series if this method were applied. Furthermore, the issue of identifying dynamic series that store their individual series separately is not addressed.

Regarding model performance, studies that train classifiers using DICOM metadata [15–18] achieve f1 scores above 90%, comparable to the individual models in the current work. These scores suggest that with effective data cleaning and ML model training methodologies, high accuracy rates can be achieved when performing automatic annotations in MRI series enhancements. The primary limitation is the heterogeneity of data and annotations, which affects the ability to extend the types of enhancements as extensively as possible.

The proposed machine learning (ML) approach offers the capability to automate the identification of both static and dynamic MR series without relying on DICOM headers containing natural language descriptions and with the understanding that some DICOM header values may be absent. This work enables the application of analysis algorithms across the entire database. The extracted information must not only satisfy the requirements of automatic analysis but also meet the needs of potential users who intend to perform searches within the image database.

Moreover, in studies like [15], where the detection of “junk” data is unified with the weighting model, this approach can introduce noise into the main model. A series categorized as Calibration or Localizer could simultaneously be a T1 or T2-weighted sequence, causing confusion at the metadata level. By separating the identification of junk data into two phases—first labeling the weighting and then determining the utility of the series—we have achieved better results, reducing confusion and improving model accuracy.

In the context of image selection for radiomic feature extraction and tumor segmentation, T2-weighted images with fat suppression are preferable to basic T2-weighted images. Recognizing the presence of fat suppression in an image can positively impact the segmentation phase by facilitating easier delineation of tumor boundaries by the radiologist. Moreover, it helps prevent potential biases in the extraction of radiomic features, which are expected to differ significantly depending on whether fat suppression is applied. This is particularly important when the dataset includes both types of images (with and without fat suppression).

Employing deep learning (DL) models on heavy data such as volumetric pixel matrices require powerful GPUs for both training and inference, capable of fully storing the volume in memory [12–14]. Even then, inference times remain higher than those of lighter ML models that utilize traditional algorithms like Random Forests on tabular data [23]. This makes the current work, which leverages DICOM metadata and employs CatBoost algorithm, more efficient and accessible, especially in environments with limited computational resources [24].

This work acknowledges that it does not cover all desired enhancements, such as proton density, susceptibility-weighted imaging, or magnetic resonance angiography (MRA), due to the lack of images annotated with these classes in the employed database. However, this limitation could be addressed when the PRIMAGE repository will be merge with the rest of AI4HI databases in the EUCAIM project [31].

Rather than harmonizing the entire database to unify dynamic series in their combined form, this work adopts a more generalist approach that is applicable in a wider range of scenarios. A custom feature, based on Euclidean similarities, has been created for each MR series to assist the models in identifying individual series that belong to a dynamic series and can thus be combined. This solution is essential for achieving satisfactory results in DCE and DWI classifications.

Regarding the junk classification model, it has proven especially useful as an initial step in cleaning the database of unwanted series. It has supervisedly reduced the number of stored series by approximately 2%. Not only are localizers and calibrations removed, but also other unhelpful series, such as synchronized respiratory series, which should not be stored in the repository and might otherwise have gone unidentified.

The CatBoost algorithm was chosen due to its ability to handle empty data inputs in both training and testing phases, reducing system complexity by eliminating the need for a missing value imputer. Moreover, it can natively process categorical data, making it well-suited for DICOM metadata. Given these advantages and its demonstrated performance in big data environments, CatBoost is considered an optimal choice for this type of study [32].

Performance testing of the final algorithm, which comprises three CatBoost models running in parallel, revealed that the entire image repository (approximately twenty thousand MR series) can be categorized in just forty to fifty seconds with a high level of accuracy using a standard CPU with twelve threads. Alternative models that utilize matrices as input for deep learning algorithms [12–14] require significantly longer categorization times compared to models based on DICOM metadata. As discussed in ref. [13], a DL approach that solely detects contrast in MR studies take between four and six seconds to classify a single MR study using a GPU. In addition to this performance disparity, the proposed work offers a modular approach that allows for the addition of a "contrast" classification model if needed. While this would necessitate a GPU and increase inference times, it provides flexibility for future enhancements.

Conclusion

Working with large medical image databases presents great challenges and problems. In a warehouse with thousands of image files, it is necessary to extract meta-information in a tabular, structured and standardized format that allows for quick identification and

search of the desired content. This is especially relevant in environments where time and computational requirements are critical. At the same time, the lack of a standard nomenclature is a problem that makes the consumption of data a big challenge.

In this work, we propose a standardized categorization schema aimed at unifying (1) the nomenclature requirements for the automated execution of image processing tasks and analyses and (2) the additional relevant information required by expert radiologists for accurate and robust definition of MR sequences. Alongside this schema, we designed a machine learning (ML) data mining process capable of handling both static and dynamic series. This process extracts information within a big data ecosystem, enabling automatic image analysis without human intervention.

The proposed ML data mining process delivers inference outputs with low computational demands and features a modular design. This allows for the expansion of functionality by integrating new models into the data mining process, such as the identification of contrast using deep learning algorithms. By addressing the challenges of large-scale medical image management and standardization, our approach facilitates more efficient data handling and paves the way for advanced automated analyses in medical imaging.

Abbreviations

MR	Magnetic resonance
ML	Machine learning
DL	Deep learning
PDW	Proton density weighting
DCE	Dynamic contrast enhance
DWI	Diffusion weighted imaging
SW	Susceptibility weighting
STIR	Short Tau Inversion Recovery
FLAIR	Fluid attenuated inversion recovery
CHS	Chemical shift
GR	Gradient echo
SE	Spin echo
IR	Inversion recovery
RM	Research mode
NB	Neuroblastoma
DIPG	Diffuse intrinsic pontine glioma
KNN	K nearest neighbors
ADC	Apparent diffusion coefficient

Acknowledgements
This study was funded by PRIMAGE (PRedictive In-silico Multiscale Analytics to support cancer personalized diaGnosis and prognosis, empowered by imaging biomarkers), a Horizon 2020|RIA project (Topic SC1-DTH-07-2018), Grant Agreement No: 826494.

Author contributions
L.M.-B conceived the idea for this study and supervised the work. A.G.-M and L.C.-A created the datasets, developed the algorithms, and took the lead in writing the manuscript in consultation with L.M.-B. From the radiological point of view, D.V.-C, L.M.-B S.C.-F, and A.T.-S. performed the series annotation and provided expert information. All authors discussed the results and contributed to the preparation of the final manuscript submitted. All authors have read and agreed to the published version of the manuscript.

Funding
This study was funded by PRIMAGE (PRedictive In-silico Multiscale Analytics to support cancer personalized diaGnosis and prognosis, empowered by imaging biomarkers), a Horizon 2020|RIA project (Topic SC1-DTH-07-2018), Grant Agreement No: 826494. This research was funded by MCIN/AEI/10.13039/501100011033 'ERDF A way of making Europe' through Grant Number PID2021-127946OB-I00.

Availability of data and materials
The datasets generated and/or analysed during the current study are available in the Primage Platform Web repository, [https://primage.quibim.com] Anyone interested can request the creation of an account for research on the platform by email to anajimenez@quibim.com.

Declarations

Ethics approval and consent to participate
This study has been approved by the Hospital's Ethics Committee (The Ethics Committee for Investigation with medicinal products of the University and Polytechnic La Fe Hospital, ethic code: 2018/0228).

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 4 May 2024 Accepted: 17 January 2025
Published online: 19 February 2025

References

1. Cerdá Alberich L, et al. A confidence habitats methodology in MR quantitative diffusion for the classification of neuroblastic tumors. Cancers. 2020;12(12): 12. https://doi.org/10.3390/cancers12123858.
2. Rodríguez-Ortega A, et al. Machine learning-based integration of prognostic magnetic resonance imaging biomarkers for myometrial invasion stratification in endometrial cancer. J Magn Reson Imaging. 2021;54(3):987–95. https://doi.org/10.1002/jmri.27625.
3. Suter Y. Radiomics for glioblastoma survival analysis in pre-operative MRI: exploring feature robustness, class boundaries, and machine learning techniques. Cancer Imaging. 2020;20:13.
4. Scapicchio C, Gabelloni M, Barucci A, Cioni D, Saba L, Neri E. A deep look into radiomics. Radiol Med. 2021;126(10):1296–311. https://doi.org/10.1007/s11547-021-01389-x.
5. Martí-Bonmatí L, et al. PRIMAGE project: predictive in silico multiscale analytics to support childhood cancer personalised evaluation empowered by imaging biomarkers. Eur Radiol Exp. 2020;4(1):22. https://doi.org/10.1186/s41747-020-00150-9.
6. Mart L. CHAIMELEON project: creation of a pan-European Repository of health imaging data for the development of AI-powered cancer management tools. Front Oncol. 2022;12:11.
7. An AI Platform integrating imaging data and models, supporting precision care through prostate cancer's continuum | ProCAncer-I Project | Fact Sheet | H2020', CORDIS | European Commission. https://cordis.europa.eu/project/id/952159 Accessed 6 Sep 2022.
8. A multimodal AI-based toolbox and an interoperable health imaging repository for the empowerment of imaging analysis related to the diagnosis, prediction and follow-up of cancer | INCISIVE Project | Fact Sheet | H2020 | CORDIS | European Commission. https://cordis.europa.eu/project/id/952179 Accessed 6 Sep 2022.
9. A European Cancer Image Platform Linked to Biological and Health Data for Next-Generation Artificial Intelligence and Precision Medicine in Oncology | EuCanImage Project | Fact Sheet | H2020 | CORDIS | European Commission. https://cordis.europa.eu/project/id/952103/es. Accessed 11 Oct 2022.
10. Tanwar M, Duggal R, Khatri SK. Unravelling unstructured data: A wealth of information in big data. In: 2015 4th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO) (Trends and Future Directions). 2015. pp. 1–6. https://doi.org/10.1109/ICRITO.2015.7359270.
11. Eberendu A. Unstructured data: an overview of the data of big data. Int J Comput Trends Technol. 2016;38:46–50. https://doi.org/10.14445/22312803/IJCTT-V38P109.
12. Ranjbar S, et al. A deep convolutional neural network for annotation of magnetic resonance imaging sequence type. J Digit Imaging. 2020;33(2):439–46. https://doi.org/10.1007/s10278-019-00282-4.
13. Pizarro R, et al. using deep learning algorithms to automatically identify the brain MRI contrast: implications for managing large databases. Neuroinformatics. 2019;17(1):115–30. https://doi.org/10.1007/s12021-018-9387-8.
14. de Mello JPV, et al. Deep learning-based type identification of volumetric MRI sequences. In: 2020 25th International Conference on Pattern Recognition (ICPR). 2021. pp. 1–8. https://doi.org/10.1109/ICPR48806.2021.9413120.
15. Liang S, et al. Magnetic resonance imaging sequence identification using a metadata learning approach. Front Neuroinform. 2021. https://doi.org/10.3389/fninf.2021.622951.
16. Gauriau R, et al. Using DICOM metadata for radiological image series categorization: a feasibility study on large clinical brain MRI datasets. J Digit Imaging. 2020;33(3):747–62. https://doi.org/10.1007/s10278-019-00308-x.
17. Florea F, Rogozan A, Bensrhair A, Dacher J-N, Darmoni S. Modality categorization by textual annotations interpretation in medical imaging. 2022.
18. Na S, et al. Sequence-Type classification of brain MRI for acute stroke using a self-supervised machine learning algorithm. Diagnostics. 2024;14(1): 1. https://doi.org/10.3390/diagnostics14010070.
19. PRedictive In-silico Multiscale Analytics to support cancer personalized diaGnosis and prognosis, Empowered by imaging biomarkers | PRIMAGE Project | Fact Sheet | H2020 | CORDIS | European Commission. https://cordis.europa.eu/project/id/826494. Accessed 6 Sep 2022.
20. MongoDB Atlas: Cloud Document Database. MongoDB. https://www.mongodb.com/cloud/atlas/lp/try4. Accessed 7 Sep 2022.
21. Budd S, Robinson EC, Kainz B. A survey on active learning and human-in-the-loop deep learning for medical image analysis. Med Image Anal. 2021;71:102062. https://doi.org/10.1016/j.media.2021.102062.
22. Kursa M, Jankowski A, Rudnicki W. Boruta—a system for feature selection. Fundam Inf. 2010;101:271–85. https://doi.org/10.3233/FI-2010-288.
23. Breiman L. Random forests. Mach Learn. 2001;45(1):5–32. https://doi.org/10.1023/A:1010933404324.

24. Prokhorenkova L, Gusev G, Vorobev A, Dorogush AV, Gulin A. CatBoost: unbiased boosting with categorical features. 2019. arXiv: arXiv:1706.09516. https://doi.org/10.48550/arXiv.1706.09516.
25. Batista G, Monard M-C. A study of K-nearest neighbour as an imputation method. 2002;30:260.
26. Erturk SM, Alberich-Bayarri A, Herrmann KA, Marti-Bonmati L, Ros PR. Use of 3.0-T MR imaging for evaluation of the abdomen. Radiographics. 2009. https://doi.org/10.1148/rg.296095516.
27. Skalski M. MRI sequence parameters | Radiology Reference Article | Radiopaedia.org'. Radiopaedia. https://radiopaedia.org/articles/mri-sequence-parameters. Accessed 11 Oct 2022.
28. Dash S, Shakyawar SK, Sharma M, Kaushik S. Big data in healthcare: management, analysis and future prospects. J Big Data. 2019;6(1):54. https://doi.org/10.1186/s40537-019-0217-0.
29. Carles M, et al. 18F-FMISO-PET hypoxia monitoring for head-and-neck cancer patients: radiomics analyses predict the outcome of chemo-radiotherapy. Cancers. 2021;13(14):3449. https://doi.org/10.3390/cancers13143449.
30. Marti-Bonmati L, et al. Pancreatic cancer, radiomics and artificial intelligence. Br J Radiol. 2022;95(1137):20220072. https://doi.org/10.1259/bjr.20220072.
31. EUCAIM. Building a Data-Driven Future for Cancer Care. Valencia, Spain. Available from https://cancerimage.eu/. Accessed 28 June 2023.
32. Hancock JT, Khoshgoftaar TM. CatBoost for big data: an interdisciplinary review. J Big Data. 2020;7(1):94. https://doi.org/10.1186/s40537-020-00369-8.

Publisher's Note