On MRI-Based Central Nervous System Tumor Differentiation using Deep Neural Networks and Conformal Predictors

Luis Balderas[1,2,4][0000−0002−3845−8848]⋆, Francisco Javier Pérez García[4,5], María Moreno de Castro[1,2][0000−0003−0440−0864], Miguel Lastra[2,3][0000−0002−7278−2668], José Pablo Martínez Barbero[4,5][0000−0003−4217−0633], Antonio Jesús Láinez-Ramos Bossini[4,5][0000−0002−0663−2856], Antonio Arauzo-Azofra[2,7][0000−0002−2486−5792], and José M. Benítez[1,2,4][0000−0002−2346−0793]

[1] Department of Computer Science and Artificial Intelligence, University of Granada, Spain
[2] DiCITS Lab, iMUDS, DaSCI University of Granada
[3] Department of Software Engineering, University of Granada
[4] Advanced Medical Imaging Group (TeCe-22). Instituto de Investigación Biosanitaria de Granada
[5] Department of Radiology, Hospital Universitario "Virgen de las Nieves", Granada
[6] Rural Engineering Department, University of Cordoba

Abstract. Central nervous system tumors (CNST) rank among the top 10 causes of cancer-related deaths worldwide. Gliomas compose a group of CNST demanding precise classification for effective treatment. While several previous studies have considered the application of Artificial Intelligence to differentiate between high and low-grade gliomas (HGG and LGG, respectively), all of them lack a reliable confidence measurement. We propose a deep learning-based technique for differentiating HGG and LGG using MRI imaging, which incorporates two novel approaches. First, perfusion sequences are represented as time series, which serve as inputs for simplified convolutional and dense models trained to distinguish between LGG and HGG according to WHO grades. Second, we employ Conformal Prediction to calibrate the results, ensuring they include the true category with a high probability. The experimental results demonstrate that our methodology not only provides accurate outputs but also ensures trustworthy predictions.

Keywords: Glioma classification, time series, conformal prediction, CNN, neural network simplification

1 Introduction

Gliomas are among the most common and aggressive central nervous system (CNS) malignancies, posing a severe and growing global healthcare challenge

⋆ Corresponding author: luisbalru@ugr.es

aligned with an aging population [19]. Precise differentiation of glioma subtypes is critical, as timely, tailored therapeutic strategies significantly improve patient survival outcomes [29]. While the definitive gold standard for classification relies on histological and molecular analysis [22], obtaining these samples requires invasive stereotactic brain biopsies that carry inherent clinical risks, including hemorrhage, infection, and neurological deficits [26]. Consequently, establishing robust, non-invasive diagnostic alternatives is a critical priority.

Perfusion Magnetic Resonance Imaging (MRI) provides a powerful non-invasive window into tissue hemodynamics by tracking contrast agent transit over time to generate dynamic perfusion curves [20]. While deep learning (DL) has successfully accelerated perfusion processing by bypassing time-intensive, physically constrained traditional tracer-kinetic modeling [8], current state-of-the-art methods typically collapse this rich temporal data into static, aggregated parametric maps. This discards critical fine-grained sequential information inherent to raw perfusion curves.

Furthermore, deploying black-box AI models in high-stakes clinical workflows necessitates rigorous uncertainty quantification. Conformal Prediction (CP) addresses this imperative by serving as a distribution-free statistical framework that transforms point predictions into prediction sets guaranteed to contain the true label at a user-specified confidence level [28]. Despite its mathematical rigor, CP remains entirely unexploited in the domain of non-invasive glioma grading.

To bridge these gaps, this paper presents a novel framework that leverages raw perfusion curves for comprehensive, trustworthy glioma differentiation. The principal highlights of this work are two-fold:

- We treat raw MRI perfusion curves directly as sequential time series, constructing a hierarchical deep-learning pipeline tailored to classify tumors across all four WHO glioma grades.
- We incorporate inductive conformal calibration into the pipeline, providing clinicians with mathematically guaranteed confidence levels alongside every diagnostic prediction.

2 Related Work and Background

This section reviews the recent literature across three pillars underpinning the proposed methodology: deep learning for brain tumor diagnosis, conformal prediction for uncertainty quantification, and the clinical paradigm of glioma differentiation.

2.1 Deep Learning-Based Brain Tumor Diagnosis

Deep learning (DL) combined with magnetic resonance imaging (MRI) has become standard for non-invasive tumor evaluation [11], [33]. In glioma management, DL architectures (e.g., U-Net) and radiomics are widely applied to augment image resolution [35], [18], segment tissues, and predict genetic profiles such as IDH mutation or 1p19q codeletion [27], [6].

Perfusion MRI parameters, such as relative cerebral blood volume (rCBV) and flow (rCBF), have also been integrated with machine learning models (e.g., random forests, multilayer perceptrons, autoencoders) to forecast glioma grades, MGMT methylation, and distinguish progression from pseudoprogression [9], [14], [23]. However, current state-of-the-art approaches rely heavily on aggregated parameter maps rather than raw perfusion curves. Furthermore, existing works rarely attempt a comprehensive differentiation across the entire spectrum of WHO grades (I–IV) as addressed in this study.

2.2 Conformal Prediction in Clinical AI

Conformal Prediction (CP) is a distribution-free framework that provides rigorous, well-calibrated uncertainty bounds for machine learning outputs [28]. For binary classification, the Inductive Venn-ABERS Predictor (IVAP) variant yields valid probabilistic outputs independently of the underlying base model [31], [24]. While CP is gaining traction for risk prediction and segmentation in breast [1], prostate [12], and colorectal cancers [17], its application to quantify prediction uncertainty in glioma characterization remains entirely unexplored.

2.3 Glioma Differentiation and Perfusion Curves

Gliomas are the most frequent primary brain malignancies. The World Health Organization (WHO) classifies them into grades I–IV based on histological and molecular biomarkers (e.g., IDH status), frequently grouped into Low-Grade (LGG, I–II) and High-Grade Gliomas (HGG, III–IV) for clinical management. Because definitive grading traditionally requires invasive and risky stereotactic biopsies, identifying reliable non-invasive alternatives via multi-sequence MRI (T1, T2, FLAIR, DWI) is paramount.

Perfusion MRI is uniquely suited for this task as it captures temporal hemodynamic changes reflecting blood-brain barrier disruption. In practice, clinical metrics like rCBV, rCBF, mean transit time, and time-to-peak are derived by consolidating time-intensity perfusion curves across successive frames [7]. However, aggregating these curves into static scalar metrics discards fine-grained temporal patterns. Leveraging the full sequential structure of raw perfusion curves via machine learning offers an untapped data source to optimize non-invasive glioma characterization.

3 Our proposal

In this paper, we present a comprehensive pipeline for glioma differentiation. We address the problem of glioma differentiation at two level of precision. The first one is concerned with differentiation between low-grade gliomas (LGG) and high-grade gliomas (HGG), which is an important heuristic dichotomization with relevant clinical implications in patient management. The second one goes on step ahead by trying to categorize gliomas according to WHO grades, a

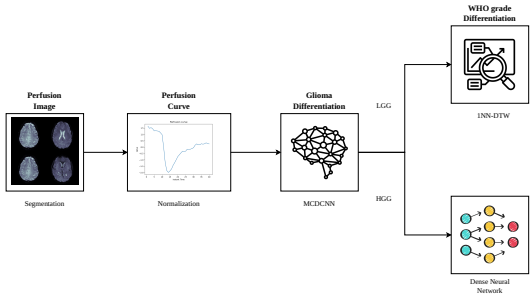


Fig. 1. Hierarchical architecture of the proposed glioma differentiation system. Extracted, normalized MRI perfusion curves serve as time-series inputs to the pipeline. In the first phase, an MCDCNN classifier distinguishes between LGG and HGG. In the second phase, the case is routed downstream based on the initial prediction: LGG cases are transferred to a 1NN-DTW model for WHO grade I vs. II differentiation, while HGG cases are processed by a dense neural network to differentiate between WHO grades III and IV.

widely used standard in the global medical community that aids physicians in formulating an effective diagnosis and treatment for the disease.

To this end, we start from a certain sequence of the patient's 3D MRI. After segmentation, specific information in the shape of a time series is extracted. This information is referred to as the perfusion curve of the MRI, which will be the fundamental information for the machine learning models. Given the particular relationship between the different types of gliomas and WHO grades, we approach the learning task as hierarchical classification, combining highly complex deep learning models with state-of-the-art methods for time series classification. Finally, we utilize conformal prediction to quantify the uncertainty in the models' outputs. A graphical representation of the overall procedure can be found in Figure 1. A detailed explanation of each phase is provided below.

Sequence Selection and Image Segmentation We utilize T2*-dynamic susceptibility contrast (DSC) MR perfusion imaging, where gadolinium administration induces a characteristic negative signal intensity slope over time [4], [5]. Data were acquired on a Philips Ingenia CX 3T scanner using a 2D gradient-echo/echo-planar sequence with the following parameters: TR/TE = 2059/40 ms, slice thickness = 4 mm, matrix = 128×128, flip angle = 75°, and 40 temporal sequences over an 88.5 s acquisition window.

Due to the lack of robust automated methods for perfusion MRI segmentation [21], [13], [32], [16], manual segmentation was performed to guarantee optimal data quality. Using 3D Slicer (v. 5.8.0), a neuroradiologist with over 10 years of experience delineated the glioma and a control region of normal-appearing white matter (NAWM), utilizing contrast-enhanced T1-weighted images for precise lesion localization.

Perfusion curves: from images to time series As previously mentioned, the perfusion sequence contains a number of snapshots that allow the study of contrast flow over time. This is expressed in terms of image luminance. Therefore, by calculating the mean luminance value of voxels at each time instant, a time series, called perfusion curve, is generated. It encapsulates the information of the perfusion sequence and is consequently very useful for tumor differentiation. Finally, perfusion curves are normalized to a common scale before being fed into machine learning models. As a result, the glioma differentiation problem, originally conceived as an image classification problem, can be casted as a time series classification task, and, as a result, it can be addressed with state-of-art techniques known in Artificial Intelligence.

Addressing a hierarchical classification challenge: Distinguishing between HGG and LGG and assigning WHO grades The normalized perfusion sequence serves as input to machine learning models to address two distinct classification problems. The first aims to differentiate gliomas into low-grade gliomas (LGG) and high-grade gliomas (HGG). The second objective is to assign a WHO grade (I-IV) to each case. These problems, however, are not independent; rather, they exhibit a close relationship: a low-grade glioma can be categorized as WHO grade I or II, while a high-grade glioma can be classified as WHO grade III or IV. Consequently, despite being binary and multiclass classification problems, respectively, we establish a hierarchical classification system. In the first stage, utilizing a MCDCNN model, we distinguish between LGG and HGG. Subsequently, to determine the WHO grade, we employ a specialized 1NN-DTW model to differentiate between grades I and II for cases previously labeled as LGG. Conversely, if the preceding model classifies the case as HGG, we utilize a dense neural network to discriminate between WHO grades III and IV.

Uncertainty Quantification: Conformal Prediction A point of the utmost relevance is assessing the confidence on the output of classifiers. Many classifiers can provide a probability, that, however, is not usually well calibrated. An innovative approach can be considered instead, conformal predictors. Specifically, since the aforementioned scheme involves three distinct binary classifiers, we employ the Inductive Venn-Abers conformal predictor [31].

This approach ensures that the models, namely MCDCNN, 1NN-DTW, and a dense deep neural network, generate calibrated probabilities. This tool is crucial as it enhances model predictive accuracy while simultaneously enabling clinicians to assess the confidence of each model at various predictive stages. In conclusion, the proposed pipeline for glioma differentiation is both accurate, trustworthy and reliable, combining state-of-the-art models with robust uncertainty quantification tools.

4 Results

A thorough experimental study has been designed to assess the efficacy of our novel methodology for differentiating gliomas into low-grade, high-grade, and specific WHO grade categories (I-IV). This section outlines the experimental goals, the dataset, evaluation metrics, and the employed classifiers.

4.1 Experimental Goals

The primary objective of this study is to develop a robust, hierarchical computer-aided diagnosis (CAD) system for non-invasive glioma characterization. Specifically, the pipeline addresses two sequential time series classification tasks: distinguishing between LGG and HGG, and subsequently classifying tumors across all four WHO grades. To bridge the gap between high-predictive performance and clinical deployability, we integrate Conformal Prediction into deep learning and time-series architectures. This provides radiologists with mathematically sound uncertainty quantification alongside each diagnostic prediction. As metrics, accuracy metrics such as accuracy and F1-score are combined with calibration metrics like Brier Score (BS) and Log Loss (LL).

4.2 Dataset and Preprocessing

We utilized a homogeneous cohort of 106 patients from the Hospital Universitario Virgen de las Nieves (Granada, Spain). Mirroring real-world clinical prevalence [22], the dataset exhibits a substantial class imbalance. The cohort comprises 70% HGG and 30% LGG cases; broken down by WHO grades, the distribution is 62% grade IV, 18% grade III, 15% grade II, and 5% grade I.

To accommodate inductive conformal calibration, the data were partitioned into distinct training, calibration, and test sets. To mitigate the small patient cohort size and class imbalance during training, a data augmentation strategy was implemented: for each patient, distinct localized perfusion curves were extracted by traversing the z-axis throughout the manually segmented tumor volumes.

4.3 Classifiers

We evaluated a diverse suite of state-of-the-art time series classifiers: Random Forest [10], 1NN-DTW [25], Shapelet [15], Support Vector Classifier (SVC) [30], Multi-Channel Convolutional Neural Network (MCDCNN) [34], and a Dense Neural Network (DNN). To optimize efficiency, MCDCNN and DNN models were integrated via their streamlined and optimized variants, MCDCNN* and DNN*, derived using OCNNA [2] and ODF2NNA [3], respectively. All models were evaluated in both unconformalized and conformalized configurations using Inductive Venn-ABERS.

Empirical analysis identifies a hierarchical configuration as the optimal pipeline architecture (Fig. 1). The optimized MCDCNN* model first isolates HGG from

LGG tumors. Subsequently, a 1NN-DTW classifier differentiates between LGG subgroups (WHO grades I vs. II), while the DNN* model differentiates between HGG subgroups (WHO grades III vs. IV). Candidate classifier performance across a stratified 5-fold cross-validation is reported in Tables 1, 2, and 3, where table cells follow the notation (unconformalized / conformalized).

Model	**Acc**	**F1**	**BL**	**LL**
RF	0.762 / 0.760	0.863 / 0.86	0.557 / 0.556	0.176 / 0.182
1NN-DTW	0.737 / 0.742	0.837 / 0.847	9.486 / 0.546	0.263 / 0.181
MCDCNN*	0.698 / **0.778**	0.803 / **0.876**	3.265 / **0.528**	0.288 / **0.173**
Shapelet	0.703 / 0.777	0.815 / 0.874	5.0373 / 0.531	0.235 / 0.173
SVC	0.779 / 0.777	0.875 / 0.875	0.522 / 0.531	0.170 / 0.174
DNN*	0.661 / 0.772	0.766 / 0.870	1.585 / 0.564	0.300 / 0.182

Table 1. Average test performance for HGG vs. LGG differentiation across 5-fold stratified cross-validation. Metrics are reported as (unconformalized / conformalized). Asterisks (*) denote models simplified using OCNNA [2] (MCDCNN*) and ODF2NNA [3] (DNN*). Conformalized MCDCNN achieves the highest accuracy and calibration performance.

5 Results and Discussion

The experimental results validate our hierarchical classification architecture and underscore the value of uncertainty quantification via conformalization. For the primary stage (HGG vs. LGG, Table 1), the optimized MCDCNN* model consistently outperformed all alternative approaches across all four evaluation metrics. Crucially, except for Random Forest, all candidate models benefited from inductive conformalization. In fact, relying solely on baseline, unconformalized outputs would have shifted the optimal model selection entirely, highlighting how calibration alters diagnostic decision-making.

For the subsequent sub-classifications, different architectures emerged as optimal. In the LGG branch (WHO grade I vs. II, Table 2), 1NN-DTW provided the strongest performance, where conformalization expanded its performance margin over competing models. Conversely, for the HGG branch (WHO grade III vs. IV, Table 3), the DNN* model was selected; it yielded the top results for F1-score, Brier Loss, and Log Loss, while maintaining competitive parity in accuracy with Shapelet and SVC methods.

A granular inspection of the Venn-ABERS framework (Fig. 2) reveals how conformalization directly enhances clinical utility. By computing a calibrated probability range $[p_0, p_1]$, the pipeline successfully recovers and corrects baseline misclassifications in instances 6–9. For remaining false predictions, the framework provides an essential safety layer. For instance, while cases 1 and 2 remain misclassified, their wide probability intervals (> 0.6) alert clinicians to

high model uncertainty. Admittedly, edge cases like instances 12 and 18 remain misclassified with deceptively high confidence. This limitation is primarily attributed to data scarcity, as conformal calibration relies heavily on empirical coverage over sample size. Expanding the calibration dataset in future cohorts will further refine these bounds, reinforcing the pipeline's reliability as a clinical decision-support tool.

Model	**Acc**	**F1**	**BL**	**LL**
RF	0.584 / 0.555	0.547 / 0.704	1.289 / 0.744	0.319 / 0.262
1NN-DTW	0.453 / **0.748**	0.472 / **0.853**	19.715 / **0.651**	0.547 / **0.215**
MCDCNN*	0.454 / 0.627	0.603 / 0.746	2.687 / 0.723	0.489 / 0.249
Shapelet	0.587 / 0.668	0.561 / 0.780	8.437 / 0.676	0.352 / 0.230
SVC	0.748 / 0.684	0.644 / 0.808	0.654 / 0.748	0.225 / 0.245
DNN*	0.574 / 0.635	0.724 / 0.763	6.265 / 0.694	0.420 / 0.239

Table 2. Average test performance for WHO grade I vs. II differentiation across 5-fold stratified cross-validation. Metrics are reported as (unconformalized / conformalized). Asterisks (*) denote models simplified using OCNNA [2] (MCDCNN*) and ODF2NNA [3] (DNN*). Conformalized 1NN-DTW achieves the highest accuracy and calibration performance.

Model	**Acc**	**F1**	**BL**	**LL**
RF	0.699 / 0.755	0.640 / 0.860	0.952 / 0.559	0.243 / 0.185
1NN-DTW	0.619 / 0.752	0.609 / 0.858	13.720 / 0.552	0.381 / 0.183
MCDCNN*	0.585 / 0.760	0.725 / 0.863	2.003 / 0.593	0.357 / 0.195
Shapelet	0.662 / **0.767**	0.634 / 0.868	4.810 / 0.562	0.268 / 0.187
SVC	0.774 / 0.767	0.676 / 0.867	0.597 / 0.607	0.191 / 0.203
DNN*	0.735 / 0.766	0.846 / **0.869**	3.481 / **0.550**	0.257 / **0.182**

Table 3. Average test performance for WHO grade III vs. IV differentiation across 5-fold stratified cross-validation. Metrics are reported as (unconformalized / conformalized). Asterisks (*) denote models simplified using OCNNA [2] (MCDCNN*) and ODF2NNA [3] (DNN*). The conformalized dense neural network achieves the highest accuracy and calibration performance.

6 Conclusions

In this paper, the brain glioma differentiation problem has been addressed by using brain T2*-DSC MRI perfusion sequences of patients as inputs. The goal is to precisely characterize the glioma subtype at two precision levels, HGG vs. LGG and WHO I-IV. The innovations of our approach are focusing on the perfusion

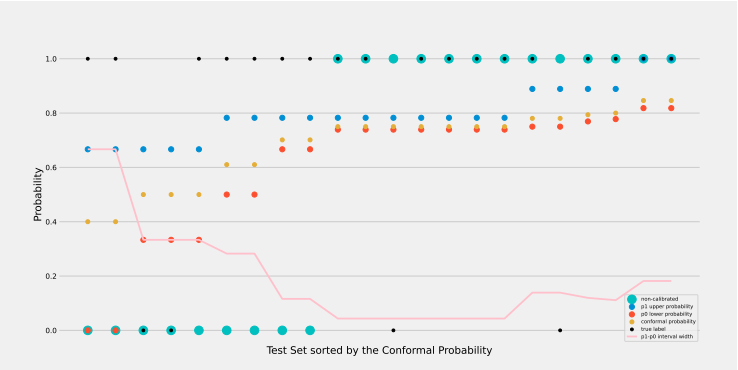


Fig. 2. Prediction analysis and uncertainty quantification for HGG vs. LGG differentiation across test instances. Individual points represent true labels (black), unconformalized predictions (thick turquoise), lower/upper Venn-ABERS probabilities p_0 (red) and p_1 (light blue), and final conformalized probabilities (yellow). Vertical intervals (pink lines) measure model uncertainty. Conformalization successfully corrects the misclassifications in cases 6–9. Furthermore, for the first two cases, the framework highlights high prediction uncertainty, effectively flags incorrect baseline outputs, and improves diagnostic reliability.

curves, each of which is actually a time series, and formulate the differentiation problem as a time series classification problem.

Taking the perfusion curve as input, our proposal prescribes the construction of a classifier to predict the diagnosis. Different state-of-the-art classifiers have been considered. After a thorough empirical evaluation, the highest performance has been obtained with convolutional neural networks for HGG and LGG differentiation. In addition, conformal predictors have been developed to bound the uncertainty on the prediction. The use of conformal predictions showed a significant improvement in the predictive capacity of the applied deep learning models, as well as a high degree of certainty.

Acknowledgments. This research has been partially supported by Proyecto PID2023-151336OB-I00 financiado por MICIU/AEI/10.13039-501100011033 y por FEDER, UE.

Disclosure of Interests. The study was conducted in accordance with the Declaration of Helsinki and approved by our Institutional Review Board. (approval code: IANeuro24; approval date: 29 June 2021). Due to the retrospective nature of the study, written informed consent was waived.

References

1. Almaslukh, B.: A reliable breast cancer diagnosis approach using an optimized deep learning and conformal prediction. Biomedical Signal Processing and Control **98** (DEC 2024). https://doi.org/10.1016/j.bspc.2024.106743

2. Balderas, L., Lastra, M., Benítez, J.M.: Optimizing convolutional neural network architectures. Mathematics **12**(19) (2024). https://doi.org/10.3390/math12193032, https://www.mdpi.com/2227-7390/12/19/3032
3. Balderas, L., Lastra, M., Benítez, J.M.: Optimizing dense feed-forward neural networks. Neural Networks **171**, 229–241 (2024). https://doi.org/https://doi.org/10.1016/j.neunet.2023.12.015, https://www.sciencedirect.com/science/article/pii/S0893608023007219
4. Bammer, R., Amukotuwa, S.A.: Dynamic susceptibility contrast perfusion, Part 1: The fundamentals. Magn Reson Imaging Clin N Am **32**(1), 1–23 (Feb 2024). https://doi.org/10.1016/j.mric.2023.09.010
5. Bammer, R., Amukotuwa, S.A.: Dynamic susceptibility contrast perfusion, Part 2: Deployment with and without contrast leakage present. Magn Reson Imaging Clin N Am **32**(1), 25–45 (Feb 2024). https://doi.org/10.1016/j.mric.2023.09.011
6. Chang, P., Grinband, J., Weinberg, B.D., Bardis, M., Khy, M., Cadena, G., Su, M.Y., Cha, S., Filippi, C.G., Bota, D., Baldi, P., Poisson, L.M., Jain, R., Chow, D.: Deep-learning convolutional neural networks accurately classify genetic mutations in gliomas. AJNR Am. J. Neuroradiol. **39**(7), 1201–1207 (Jul 2018)
7. Cho, S., Na, D., Ryoo, J., Roh, H., Moon, C., Byun, H., Kim, J.: Perfusion MR imaging: clinical utility for the differential diagnosis of various brain tumors. Korean J. Radiol. **3**(3), 171–179 (2002). https://doi.org/10.3348/kjr.2002.3.3.171
8. Choi, K.S.: Deep learning applications in perfusion MRI: recent advances and current challenges. Investigative Magnetic Resonance Imaging **26**(4), 246–255 (2022)
9. Crisi, G., Filice, S.: Predicting mgmt promoter methylation of glioblastoma from dynamic susceptibility contrast perfusion: A radiomic approach. Journal of Neuroimaging **30**(4), 458–462 (May 2020). https://doi.org/10.1111/jon.12724, http://dx.doi.org/10.1111/jon.12724
10. Deng, H., Runger, G., Tuv, E., Vladimir, M.: A time series forest for classification and feature extraction. Information Sciences **239**, 142–153 (2013). https://doi.org/https://doi.org/10.1016/j.ins.2013.02.030, https://www.sciencedirect.com/science/article/pii/S0020025513001473
11. Egger, J., Gsaxner, C., Pepe, A., Pomykala, K.L., Jonske, F., Kurz, M., Li, J., Kleesiek, J.: Medical deep learning—a systematic meta-review. Computer Methods and Programs in Biomedicine **221**, 106874 (2022). https://doi.org/https://doi.org/10.1016/j.cmpb.2022.106874, https://www.sciencedirect.com/science/article/pii/S0169260722002565
12. Gade, M., Nguyen, K.M., Gedde, S., Fernandez-Quilez, A.: Impact of uncertainty quantification through conformal prediction on volume assessment from deep learning-based MRI prostate segmentation. Insights into Imaging **15**(1) (NOV 29 2024). https://doi.org/10.1186/s13244-024-01863-w
13. Gagnon, L., Gupta, D., Mastorakos, G., White, N., Goodwill, V., McDonald, C.R., Beaumont, T., Conlin, C., Seibert, T.M., Nguyen, U., Hattangadi-Gluth, J., Kesari, S., Schulte, J.D., Piccioni, D., Schmainda, K.M., Farid, N., Dale, A.M., Rudie, J.D.: Deep learning segmentation of infiltrative and enhancing cellular tumor at pre- and posttreatment multishell diffusion MRI of glioblastoma. Radiology: Artificial Intelligence **6**(5), e230489 (2024). https://doi.org/10.1148/ryai.230489, https://doi.org/10.1148/ryai.230489, pMID: 39166970
14. Hashido, T., Saito, S., Ishida, T.: A radiomics-based comparative study on arterial spin labeling and dynamic susceptibility contrast perfusion-weighted imaging in gliomas. Scientific Reports **10**(1) (Apr 2020). https://doi.org/10.1038/s41598-020-62658-9, http://dx.doi.org/10.1038/s41598-020-62658-9

15. Hills, J., Lines, J., Baranauskas, E., Mapp, J., Bagnall, A.: Classification of time series by shapelet transformation. Data Mining and Knowledge Discovery **28**(4), 851–881 (Jul 2014). https://doi.org/10.1007/s10618-013-0322-1, https://doi.org/10.1007/s10618-013-0322-1
16. Hoebel, K.V., Bridge, C.P., Ahmed, S., Akintola, O., Chung, C., Huang, R.Y., Johnson, J.M., Kim, A., Ly, K.I., Chang, K., Patel, J., Pinho, M., Batchelor, T.T., Rosen, B.R., Gerstner, E.R., Kalpathy-Cramer, J.: Expert-centered evaluation of deep learning algorithms for brain tumor segmentation. Radiology: Artificial Intelligence **6**(1), e220231 (2024). https://doi.org/10.1148/ryai.220231, https://doi.org/10.1148/ryai.220231
17. Kapuria, S., Minot, P., Kapusta, A., Ikoma, N., Alambeigi, F.: A novel dual layer cascade reliability framework for an informed and intuitive clinician-AI interaction in diagnosis of colorectal cancer polyps. IEEE Journal of Biomedical and Health Informatics **28**(4), 2326–2337 (APR 2024). https://doi.org/10.1109/JBHI.2024.3350082
18. Lee, J., Jung, W., Yang, S., Park, J.H., Hwang, I., Chung, J.W., Choi, S.H., Choi, K.S.: Deep learning-based super-resolution and denoising algorithm improves reliability of dynamic contrast-enhanced MRI in diffuse glioma. Scientific Reports **14**(1), 25349 (Oct 2024). https://doi.org/10.1038/s41598-024-76592-7, https://doi.org/10.1038/s41598-024-76592-7
19. Lin, D., Wang, M., Chen, Y., Gong, J., Chen, L., Shi, X., Lan, F., Chen, Z., Xiong, T., Sun, H., et al.: Trends in intracranial glioma incidence and mortality in the united states, 1975-2018. Frontiers in Oncology **11**, 748061 (2021)
20. Luypaert, R., Boujraf, S., Sourbron, S., Osteaux, M.: Diffusion and perfusion MRI: basic physics. European journal of radiology **38**(1), 19–27 (2001)
21. Minaee, S., Boykov, Y., Porikli, F., Plaza, A., Kehtarnavaz, N., Terzopoulos, D.: Image segmentation using deep learning: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence **44**(7), 3523–3542 (2022). https://doi.org/10.1109/TPAMI.2021.3059968
22. Nabors, L., Portnow, J., Ahluwalia, M., Baehring, J., Brem, H., Brem, S., Butowski, N., Campian, J., Clark, S., Fabiano, A., Forsyth, P., Hattangadi-Gluth, J., Holdhoff, M., Horbinski, C., Junck, L., Kaley, T., Kumthekar, P., Loeffler, J., Mrugala, M., Nagpal, S., Pandey, M., Parney, I., Peters, K., Puduvalli, V., Robins, I., Rockhill, J., Rusthoven, C., Shonka, N., Shrieve, D., Swinnen, L., Weiss, S., Wen, P., Willmarth, N., Bergman, M., Darlow, S.: Central nervous system cancers, version 3-2020, nccn clinical practice guidelines in oncology. J Natl Compr Canc Netw. **18**(11), 1537–1570 (2022). https://doi.org/10.6004/jnccn.2020.0052. PMID: 33152694
23. Park, J.E., Kim, H.S., Lee, J., Cheong, E.N., Shin, I., Ahn, S.S., Shim, W.H.: Deep-learned time-signal intensity pattern analysis using an autoencoder captures magnetic resonance perfusion heterogeneity for brain tumor differentiation. Scientific Reports **10**(1), 21485 (2020). https://doi.org/10.1038/s41598-020-78485-x, https://doi.org/10.1038/s41598-020-78485-x
24. Pereira, T., Cardoso, S., Guerreiro, M., Mendonça, A., Madeira, S.C.: Targeting the uncertainty of predictions at patient-level using an ensemble of classifiers coupled with calibration methods, Venn-ABERS, and conformal predictors: A case study in AD. Journal of Biomedical Informatics **101**, 103350 (2020). https://doi.org/https://doi.org/10.1016/j.jbi.2019.103350, https://www.sciencedirect.com/science/article/pii/S1532046419302692

25. Rakthanmanon, T., Campana, B., Mueen, A., Batista, G., Westover, B., Zhu, Q., Zakaria, J., Keogh, E.: Addressing big data time series: Mining trillions of time series subsequences under dynamic time warping. ACM Trans. Knowl. Discov. Data **7**(3) (Sep 2013). https://doi.org/10.1145/2500489, https://doi.org/10.1145/2500489
26. Riche, M., Marijon, P., Amelot, A., Bielle, F., Mokhtari, K., Chambrun, M., Joncour, A., Idbaih, A., Touat, M., Do, C., Deme, M., Pasqualotto, R., Jacquens, A., Degos, V., Shotar, E., Chougar, L., Carpentier, A., Mathon, B.: Severity, timeline, and management of complications after stereotactic brain biopsy. J Neurosurg **136**(3), 867–876 (2021). https://doi.org/10.3171/2021.3.JNS21134. PMID: 34507289
27. Sabeghi, P., Zarand, P., Zargham, S., Golestany, B., Shariat, A., Chang, M., Yang, E., Rajagopalan, P., Phung, D.C., Gholamrezanezhad, A.: Advances in neuro-oncological imaging: An update on diagnostic approach to brain tumors. Cancers **16**(3) (2024). https://doi.org/10.3390/cancers16030576, https://www.mdpi.com/2072-6694/16/3/576
28. Shafer, G., Vovk, V.: A tutorial on conformal prediction. J. Mach. Learn. Res. **9**, 371–421 (Jun 2008)
29. Stupp, R., Brada, M., Van Den Bent, M., Tonn, J.C., Pentheroudakis, G.: High-grade glioma: Esmo clinical practice guidelines for diagnosis, treatment and follow-up. Annals of Oncology **25**, 93–101 (2014)
30. Suthaharan, S.: Support vector machine (2016). https://doi.org/10.1007/978-1-4899-7641-3_9, https://doi.org/10.1007/978-1-4899-7641-39
31. Vovk, V., Petej, I.: Venn-Abers predictors (2014), https://arxiv.org/abs/1211.0025
32. Wu, S., Li, H., Quang, D., Guan, Y.: Three-plane–assembled deep learning segmentation of gliomas. Radiology: Artificial Intelligence **2**(2), e190011 (2020). https://doi.org/10.1148/ryai.2020190011, https://doi.org/10.1148/ryai.2020190011, pMID: 32280947
33. Yang, P., Chen, W., Qiu, H.: Mmgcn: Multi-modal multi-view graph convolutional networks for cancer prognosis prediction. Computer Methods and Programs in Biomedicine **257**, 108400 (2024). https://doi.org/https://doi.org/10.1016/j.cmpb.2024.108400, https://www.sciencedirect.com/science/article/pii/S0169260724003936
34. Zheng, Y., Liu, Q., Chen, E., Ge, Y., Zhao, J.L.: Time series classification using multi-channels deep convolutional neural networks. In: Li, F., Li, G., Hwang, S.w., Yao, B., Zhang, Z. (eds.) Web-Age Information Management. pp. 298–310. Springer International Publishing, Cham (2014)
35. Özbay, E., Altunbey Özbay, F.: Interpretable features fusion with precision mri images deep hashing for brain tumor detection. Computer Methods and Programs in Biomedicine **231**, 107387 (2023). https://doi.org/https://doi.org/10.1016/j.cmpb.2023.107387, https://www.sciencedirect.com/science/article/pii/S0169260723000548