



An Explainable MRI-Radiomic Quantum Neural Network to Differentiate Between Large Brain Metastases and High-Grade Glioma Using Quantum Annealing for Feature Selection

Tony Felefly^{1,2,3} · Camille Roukoz¹ · Georges Fares^{1,4} · Samir Achkar⁵ · Sandrine Yazbeck⁶ · Philippe Meyer^{7,8} · Manal Kordahi⁹ · Fares Azoury¹ · Dolly Nehme Nasr¹ · Elie Nasr¹ · Georges Noël^{10,11,12} · Ziad Francis⁴

Received: 19 April 2023 / Revised: 11 June 2023 / Accepted: 17 July 2023 / Published online: 28 July 2023

© The Author(s) 2023

Abstract

Solitary large brain metastases (LBM) and high-grade gliomas (HGG) are sometimes hard to differentiate on MRI. The management differs significantly between these two entities, and non-invasive methods that help differentiate between them are eagerly needed to avoid potentially morbid biopsies and surgical procedures. We explore herein the performance and interpretability of an MRI-radiomics variational quantum neural network (QNN) using a quantum-annealing mutual-information (MI) feature selection approach. We retrospectively included 423 patients with HGG and LBM (> 2 cm) who had a contrast-enhanced T1-weighted (CE-T1) MRI between 2012 and 2019. After exclusion, 72 HGG and 129 LBM were kept. Tumors were manually segmented, and a 5-mm peri-tumoral ring was created. MRI images were pre-processed, and 1813 radiomic features were extracted. A set of best features based on MI was selected. MI and conditional-MI were embedded into a quadratic unconstrained binary optimization (QUBO) formulation that was mapped to an Ising-model and submitted to D'Wave's quantum annealer to solve for the best combination of 10 features. The 10 selected features were embedded into a 2-qubits QNN using PennyLane library. The model was evaluated for balanced-accuracy (bACC) and area under the receiver operating characteristic curve (ROC-AUC) on the test set. The model performance was benchmarked against two classical models: dense neural networks (DNN) and extreme gradient boosting (XGB). Shapley values were calculated to interpret sample-wise predictions on the test set. The best 10-feature combination included 6 tumor and 4 ring features. For QNN, DNN, and XGB, respectively, training ROC-AUC was 0.86, 0.95, and 0.94; test ROC-AUC was 0.76, 0.75, and 0.79; and test bACC was 0.74, 0.73, and 0.72. The two most influential features were tumor Laplacian-of-Gaussian-GLRLM-Entropy and sphericity. We developed an accurate interpretable QNN model with quantum-informed feature selection to differentiate between LBM and HGG on CE-T1 brain MRI. The model performance is comparable to state-of-the-art classical models.

Keywords Brain metastasis · Glioma · Machine Learning · Quantum annealing · Variational classifier

✉ Tony Felefly
tony.felefly@hotmail.com

¹ Radiation Oncology Department, Hôtel-Dieu de France Hospital, Saint Joseph University, Beirut, Lebanon

² ICube Laboratory, University of Strasbourg, Strasbourg, France

³ Radiation Oncology Department, Hôtel-Dieu de Lévis, Lévis, QC, Canada

⁴ Physics Department, Saint Joseph University, Beirut, Lebanon

⁵ Radiation Oncology Department, Gustave Roussy Cancer Campus, 94805 Villejuif, France

⁶ Department of Radiology, University of Maryland School of Medicine, 655 W Baltimore St S, Baltimore, MD 21201, USA

⁷ Medical Physics Department, Institut de Cancérologie de Strasbourg (ICANS), 67200 Strasbourg, France

⁸ IMAGeS Unit, IRIS Platform, ICube, University of Strasbourg, 67085 Strasbourg Cedex, France

⁹ Institut National de Pathologie, Beirut, Lebanon

¹⁰ Radiotherapy Department, Institut de Cancérologie de Strasbourg (ICANS), 67200 Strasbourg, France

¹¹ Radiobiology Department, IMIS Unit, IRIS Platform, ICube, University of Strasbourg, 67085 Strasbourg Cedex, France

¹² Faculty of Medicine, University of Strasbourg, 67000 Strasbourg, France

Introduction

The most frequent malignant brain tumors in adults are metastases (BM) and high-grade gliomas (HGG) with incidence rates of 7 to 14 and 1.95 per 100,000 population per year, respectively [1, 2]. MRI is the gold standard imaging modality for diagnosing and characterizing BM and HGG. However, differentiating these two entities on MRI is often challenging, owing to shared imaging characteristics such as central necrosis, ring enhancement, and peritumoral edema. Furthermore, both tumor types might present as a solitary mass or multifocal disease, albeit with different frequencies.

The management strategy differs significantly between these two disease categories [3–6]. Although biopsy and/or surgical resection with histopathological and molecular analysis [7] is generally recommended to establish a definite diagnosis and guide personalized treatment, there remains an undeniable need to improve the diagnostic ability of imaging to differentiate between these two disease entities. In rare cases where an invasive procedure could not be undertaken due to medical contraindications or eloquent tumor location, imaging accuracy becomes paramount. Moreover, for patients with a known history of malignancy, avoiding a confirmatory biopsy for brain metastases could prevent unnecessary toxicity, and thus, leveraging confidence in the diagnostic capabilities of MRI imaging is eagerly needed.

MRI techniques that help differentiate between BM and HGG represent an active area of research, and significant improvements over conventional MRI techniques have been made in the last few years using spectroscopy, perfusion, and diffusion-weighted imaging techniques [8–14]. A major drawback of such techniques with respect to conventional MRI is the need for additional special imaging acquisitions that would have a negative impact on the procedural cost, patient convenience, and the time to get a diagnosis. Further, many of these techniques are not widely available.

Radiomics extraction is a promising quantitative method that could glean useful information from imaging data that are not discernable to the human eye [15, 16]. Two types of radiomics exist: “handcrafted radiomics” that are distinct mathematically-defined features generated through a stand-alone process and “deep radiomics” that represent task-oriented features extracted by an auto-encoder or as a part of the deep layers’ data of a convolutional neural network [17]. In many tasks, deep radiomics outperformed engineered features [18–20]. In one BM-glioblastoma classification study by Bae et al. comparing both techniques, deep radiomics yielded slightly better results than their counterparts [21].

Quantum computing is gaining much momentum as advances in quantum device engineering and

computational algorithms are surging, and healthcare is a promising arena for these technologies [22]. Quantum machine learning (QML) is an emerging bridging discipline that combines principles from both quantum computing and classical machine learning [23–26]. Few studies have pointed to a potential superiority of QML over classical algorithms under specific circumstances [27–29], and classification tasks with small training datasets might derive some benefit in terms of performance and generalization [30, 31]. However, a clear quantum advantage is yet to be demonstrated. Variational quantum classifiers (VQC), also quoted as quantum neural networks (QNN) due to their analogy with classical neural networks, are hybrid quantum–classical models consisting of three sequential blocks: data encoding or state preparation, trainable parametrized unitary evolution, and measurement of the quantum state. Such algorithms are considered kernel classifiers in that they map classical data to a high-dimensional feature space where a linear separation could be defined by measurement [32]. Although the hope of QML is to achieve an advantage over classical methods through computational speed-ups, such results might not be possible on near-term quantum devices.

In order to enhance model interpretability, dimensionality reduction through feature selection is a pivotal step before training. Mutual information feature selection (MIFS) is a model-agnostic filtering method to select the features that share the maximal amount of information with the output variable and is able to capture non-linear relationships [33]. Conditional mutual information scores have been used to account for feature interaction [34], and such computation is classically intractable. As inspired by the work of Nguyen and colleagues [35], the MIFS problem could be mapped to a quadratic unconstrained binary optimization (QUBO) and solved heuristically using an Ising machine. Quantum annealing is a metaheuristic that consists of evolving an initial multi-qubit system prepared as the ground state of a simple Ising-spin Hamiltonian to a final Hamiltonian whose ground state represents the solution of a desired optimization problem. Quantum annealing has been suggested to match and sometimes outperform classical solvers in combinatorial optimization [36], hypothetically due to quantum tunneling. While quantum annealing is pushing its way towards commercial utilization, its superiority over classical heuristics remains questionable.

Radiomics and deep learning techniques have been widely used for brain tumor classification [21, 37–41]. Deep learning algorithms suffer from the lack of interpretability and are regarded as “black-box” methods, drawing much skepticism around their use in clinical practice. QNNs represent a new family of machine learning algorithms that are interesting to explore on real-world datasets. In this work, we evaluate the performance of a QNN to differentiate between BM

and HGG on a real-world conventional MRI dataset, with a particular focus on model interpretability. We use a state-of-the-art method for feature selection, MIFS, taking into consideration feature interaction, and we implement it on a quantum annealer. Moreover, since BM are often smaller than HGG at diagnosis, there exists a significant confounding effect of tumor size that is seldom considered in many published studies [41]. We herein limit our analysis to large BM. Finally, we compare the QNN performance to prove classical machine learning algorithms.

Materials and Methods

Patients and Data Collection

This study was conducted at the Hôtel-Dieu de France University Medical Center after obtaining approval from the institutional Ethics board. Medical records from the radiation oncology and radiology departments were retrospectively screened for patients diagnosed with brain metastases and high-grade gliomas (WHO grades 3–4) between 2012 and 2019. A total of 423 patients aged more than 18 years who had a confirmatory pathology report and a pre-treatment gadolinium-enhanced T1-weighted (CE-T1W) brain MRI were included. Brain MRI images were thoroughly reviewed, and patients who had significant motion artifacts, lesions without ring enhancement, or tumors with largest diameter of less than 2 cm were excluded. The final dataset consisted of 129 BM and 72 HGG patients. Imaging performed at our institution as well as imported images from other radiology centers were allowed. Different imaging protocols and 2D or 3D acquisition were permitted as this is more representative of the real-world setting. Since this was a tumor-wise analysis, patients who had multiple brain metastases or multifocal gliomas were accepted for inclusion, yet only one tumor per patient was selected. A 2D largest diameter cutoff of 2 cm was chosen to exclude very small lesions that are more likely to represent metastases rather than gliomas and thus might lead to classification bias.

Segmentation

Tumors on CE-T1W brain MRI images were manually segmented by a senior radiation oncology resident and a radiation oncologist using Eclipse™ treatment planning system (Varian Medical Systems, Palo Alto, CA, USA). A 5-mm isotropic peritumoral ring extension was created for each tumor volume and was manually edited to carve out parts extending beyond anatomical barriers such as bone and cerebral falx. All segmentations were verified by both physicians for consistency.

Image Pre-Processing

First, the advanced normalization tools (ANTs) library [42] was used to apply the bias field correction on the MRI images using the N4ITK algorithm [43]. Spatial resampling to a 1 mm × 1 mm × 1 mm voxel size was then performed using ANTs. Automatic brain segmentation was done using two publicly available brain extraction tools, BET from the functional MRI of the brain library (FSL) [44] and the HD-BET [45]. The brain masks were then checked visually, and the best fitting mask was retained for each MRI. The masks were used for Z-score intensity normalization [46] of the resampled MRI images. We used the Z-score method as it resulted in the best performing radiomics-based models when used with a fixed bin number intensity discretization method [47].

Feature Extraction and Data Pre-Processing

Radiomic extraction was performed using PyRadiomics library version 3.0.1 [48] on Python 3.7.9. Most of the features used in PyRadiomics conform to the imaging biomarker normalization initiative (IBSI) [49, 50]. A fixed bin number of 32 was used for gray-level intensity discretization as recommended by Carré et al. [47]. A total number of 913 tumor features and 900 ring features were extracted for each patient. Seven feature classes were considered: shape features (13 features), first-order statistics (17 features), grey-level co-occurrence matrix (GLCM) (22 features), grey-level run length matrix (GLRLM) (16 features), grey-level size zone matrix (GLSZM) (16 features), grey-level dependence matrix (GLDM) (14 features), and neighborhood grey tone difference matrix (NGTDM) (5 features). Aside from original images, 9 filters were used for first- and second-order features prior to extraction: Laplacian of Gaussian (LoG) (1 filter, sigma value: 1) and wavelet filters (8 filters, with the different combinations of high – H and low – L pass filters in the three dimensions: LLH, LHL, LHH, HLL, HLH, HHL, HHH, and LLL). Shape features were omitted for the ring volume due to irrelevance.

The extracted features were rescaled using Scikit-learn StandardScaler function by subtracting the mean value and scaling to unit variance. The dataset was split into a training set (70%) and a validation set (30%) using a stratified approach. The number of patients in the BM cohort is significantly higher than the HGG dataset. To minimize the effect of class imbalance, a combined over-sampling (synthetic minority oversampling technique (SMOTE)) and under-sampling (RandomUnderSampler) strategy was used. Highly correlated features with a Spearman's rank correlation coefficient absolute value $|\rho|$ of more than 0.8 were eliminated.

Feature Selection

Primary feature selection was performed using linear support vector classification (LinearSVC), and then, the best combination of 10 features was determined using MIFS.

Mutual information (MI) between two variables X and Y quantifies the amount of information about one variable derived from observing the other variable. It can be expressed as a function of Shannon entropy (SE) and conditional Shannon entropy (CSE) of these variables as follows:

$$I(X;Y) = H(Y) - H(Y|X)$$

where $H(Y) = -\int p(Y)\log p(Y)dY$ is the SE of Y , a function of its probability distribution $p(Y)$, and $H(Y|X) = H(X, Y) - H(X)$ is the CSE of Y conditional of X , a function of SE of X and joint SE of both variables.

CMI between two variables X and Y given the previous selection of a variable Z takes into account the interaction between these variables and is given by:

$$I(X;Y|Z) = H(X|Z) - H(X|Y, Z)$$

where $H(X|Z)$ is the CSE of X conditional of Z and $H(X|Y, Z)$ is the CSE of X conditional of Y and Z .

Given a set of n features $X = \{X_1, X_2, X_3, \dots, X_n\}$, selecting the subset of variables $S \subset X$ that shares the maximum information with a variable Y requires maximizing $I(S;Y)$, which is an NP-hard problem. The method used herein is inspired by the work of Nguyen et al. [35], and a relevant example can be found at [51].

Under the assumption of variable conditional independence, this feature selection problem could be approximated [35] by:

$$\arg \max_{S \subset X} \left\{ \sum_{X_i \in S} I(X_i;Y) + \sum_{X_i, X_j \in S} I(X_i;Y|X_j) \right\}$$

and could be written in a matrix representation as follows:

$$\arg \max \{x^T Qx\}$$

$$x \in \{0, 1\}^n$$

where Q is a $n \times n$ matrix with $Q_{ii} = I(X_i;Y)$ and $Q_{ij} = I(X_i;Y|X_j)$ for $i \neq j$ and x is a $n \times 1$ vector with $x_i = \begin{cases} 1, & i^{\text{th}} \text{ feature is selected} \\ 0, & \text{otherwise} \end{cases}$

Evidently, this problem fits to the following QUBO formulation:

$$\arg \min \left\{ -\sum_i^n Q_{ii}x_i - \sum_{i < j}^n Q_{ij}x_i x_j \right\}$$

This optimization could be naturally mapped to an Ising model Hamiltonian and thus could be solved by

approximating its ground state on an Ising machine. We use for this purpose a quantum annealing approach on D-Wave systems advantage quantum processing unit (QPU).

In order to solve for a specific number of features k , a penalty factor is added to the QUBO to penalize solutions with $|S| \neq k$:

$$P = -\alpha \left(\sum_{i=1}^n x_i - k \right)^2$$

where α is a tunable penalty amplitude.

We selected a set of most important features based on MI scores, and then, clique embedding was performed using D-Wave's minorminer heuristic tool. The MIFS heuristic was solved for all k values, each with 5000 reads. The set of 10 features was selected to build the classification model.

Quantum Neural Network

We used herein the PennyLane quantum machine learning library [52] to build a QNN, which consists of a parametrized variational circuit that acts as a binary classifier. For this purpose, we used PennyLane's default_qubit quantum simulator, with a circuit of two qubits using an architecture similar to that described in Farhi and Neven [53] and Schuld et al. [54]. A Python code developed by the PennyLane team to implement a variational classifier could be accessed at [55].

An amplitude encoding method was employed to map the feature vector to the Hilbert space of the 2-qubit system, as shown in Möttönen et al. [56] and the book by Schuld and Petruccione [26]. For an n -qubit circuit, a feature vector of 2^n dimensions is encoded, i.e., four features in the current work. Amplitude encoding entails a sample-wise feature vector normalization to unit norm to satisfy the Born rule, which leads to different normalization factors and significant distortion of the feature dataset. To tackle this issue, three informative features were used and were subsequently padded with a non-zero constant term that would ultimately bear information on the normalization factor.

Principle component analysis (PCA) was used to reduce the dimensionality of the feature dataset from 10 to 3 while retaining the maximal amount of information. Rescaling was performed using MinMaxScaler to ensure positivity of the features, since this would lead to positive amplitudes and thus avoidance of a cascade of Z-axis rotations for the state preparation subroutine [26].

To achieve the amplitude encoding of the normalized feature vectors, a set of controlled Y-axis rotations is performed according to a reversed scheme of the algorithm used by Möttönen et al. [56]. The corresponding circuit representation in our case is shown in Fig. 1, where β denotes the

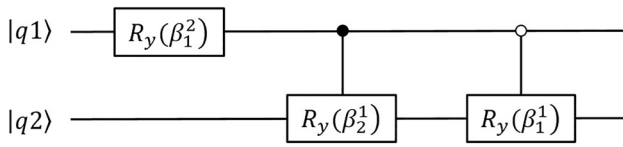


Fig. 1 Circuit representation for amplitude encoding in the case of a 2-qubit system. $R_y(\beta)$ denotes the Y-axis rotation angle; the white circle indicates a control on qubit basis state 1; the black circle indicates a control on qubit basis state 0; $|q1\rangle$ and $|q2\rangle$ are qubits 1 and 2 states, respectively

Y-axis rotation angle, and the white and black circles indicate a control on qubit basis state 1 and 0 respectively.

The angles β are given by [26]:

$$\beta_j^s = 2 \sin^{-1} \left(\frac{\sqrt{\sum_{l=1}^{2^{s-1}} |\alpha_{(2j-1)2^{s-1}+l}|^2}}{\sqrt{\sum_{l=1}^{2^s} |\alpha_{(j-1)2^s+l}|^2}} \right)$$

where α_i is the i^{th} element of the amplitude vector (i.e., the normalized feature vector).

These controlled Y-axis rotations are further decomposed into basic circuit components including controlled-NOT (CNOT) and rotation around Y-axis (Ry) gates.

We define a parametrized entangling circuit model composed of 6 layers of a unitary consisting of an arbitrary single qubit rotation (Rot gate) on each qubit followed by a CNOT gate. A Rot gate is parametrized by 3 angle values φ , θ , and ω and could be written as a function of Y- and Z-axis rotations as $RZ(\omega)RY(\theta)RZ(\varphi)$. This circuit acts on the encoded “ket” vector, and the expectation value of the Pauli Z operator acting on the first qubit is calculated. The quantum circuit (or ansatz) is shown in Fig. 2. If the value is negative, the prediction $p = -1$ is a brain metastasis; otherwise, $p = +1$ for glioma.

The parameters are classically optimized using Adam optimizer, a gradient-descent optimizer with an adaptive learning rate.

Benchmarking and Model Evaluation

In order to evaluate the performance of the QNN relative to classical machine learning models on our 10-features dataset, we benchmarked our results against two

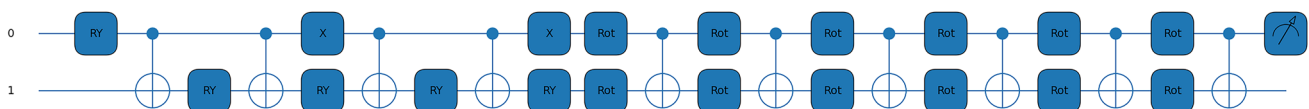


Fig. 2 The 2-qubit circuit acting as a machine learning classifier. RY: Y-axis rotation; X: Pauli X gate; Rot: arbitrary qubit rotation. Qubit numbers 1 and 2 are denoted “0” and “1”, respectively

well-established classical models: extreme gradient boosting (XGBoost) and dense neural network (DNN).

For XGBoost, we used Bayesian optimization to tune 9 hyper-parameter tuning to get the best area under receiver operating characteristic curve (ROC-AUC). The search range for the hyper-parameters was 1×10^{-5} to 1×10^{-1} for the learning rate, 300 to 1000 for the number of trees, 1 to 4 for the minimum sum of instance weight needed in a child, 3 to 5 for maximum tree depth, 0.2 to 0.5 for the subsample ratio of the training instance, 0.2 to 0.5 for the subsample ratio of columns when constructing each tree, 0 to 0.1 for the minimum loss reduction, and 0 to 75 for L1 and L2 regularization terms on weights. Five-fold cross-validation was done on the training set to reduce overfitting.

The DNN used Keras’ Adam optimizer, and included three hidden layers, with the output layer using a sigmoid activation function. We used a dropout layer with a rate of 30% after the input and each hidden layer to minimize overfitting. Early stopping with a patience of 10 was utilized during training.

Balanced accuracy (bACC) and ROC-AUC were used for model comparison. Precision, recall, and F1-scores were also reported.

QNN Model Interpretation

Shapley value [57, 58], a coalitional game-theory solution concept to determine a player’s contribution to an overall cooperative gain, was used herein to compute sample-wise feature attribution to an overall prediction, as implemented by Lundberg and colleagues [59]. We used the model-agnostic KernelExplainer to estimate Shapley values for the purpose of the current study. The mean absolute Shapley values for all features were calculated for impact ranking. Class-wise mean Shapley values and mean feature values were also calculated, and the same was done for true and false predictions.

Results

Patient Characteristics

The BM cohort included 56 patients (43.4%) with primary lung cancer, 40 patients (31%) with breast cancer, and 33

patients (25.6%) with other histologies. Median tumor 3D-largest diameter was 34 mm (IQR: 26–42.6 mm) for BM and 47.9 mm (IQR: 38.4–61.1 mm) for HGG. 3D MRI acquisition was performed in 91 patients with BM (70.5%) and 43 patients with HGG (59.7%).

Feature Selection

After eliminating highly correlated features, a set of 209 features was kept. Using LinearSVC, 54 important features were classically selected. The MI scores for these remaining features were calculated, and 17 most important features were retained. The MIFS heuristic was solved on D-Wave's quantum annealer for k between 1 and 17 (Fig. 3). The best combination of 10 features was determined and was used to build the machine learning classifiers.

QNN Model Performance and Benchmarking

The variational quantum classifier training and validation accuracy learning curves, as well as the training loss curve, are shown in Fig. 4. Model training accuracy was 0.8, and

validation bACC was 0.74. ROC-AUC was 0.86 on the training set and 0.76 on the validation set.

Respective bACC on the test set for XGBoost and DNN were 0.72 and 0.73. Training and test ROC-AUC were 0.94 and 0.79 and 0.95 and 0.75 for XGBoost and DNN, respectively. These metrics along with the precision, recall, and F1-scores for all models are summarized in Table 1. The difference in accuracy between training and validation sets are 6%, 17%, and 15% for QNN, XGBoost, and DNN, respectively, and the respective differences in ROC-AUC are 10%, 15%, and 20%.

QNN Model Interpretability

Instance-wise Shapley values were calculated, and each feature's contribution to the overall model prediction was determined. Figure 5 shows the mean absolute values for each feature, ranked by order of importance. Tumor LoG-3D-GLRLM-Run-Entropy and sphericity were the most important features. Ring GLCM-Contrast ranked third, and was the most important ring feature.

Figure 6a shows the mean Shapley value for each feature according to the predicted class, and colors represent the mean feature value according to the color scale bar. A positive Shapley value shifts the prediction output towards

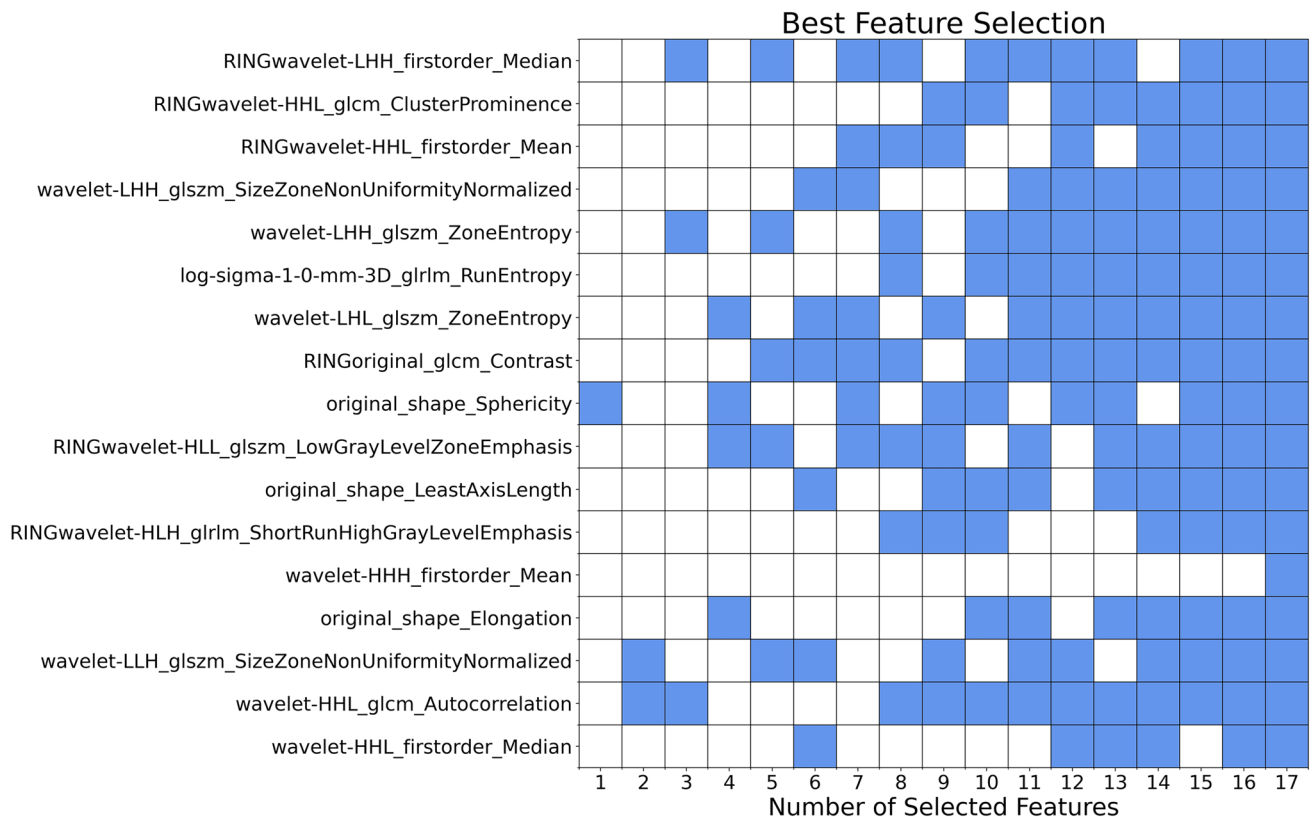
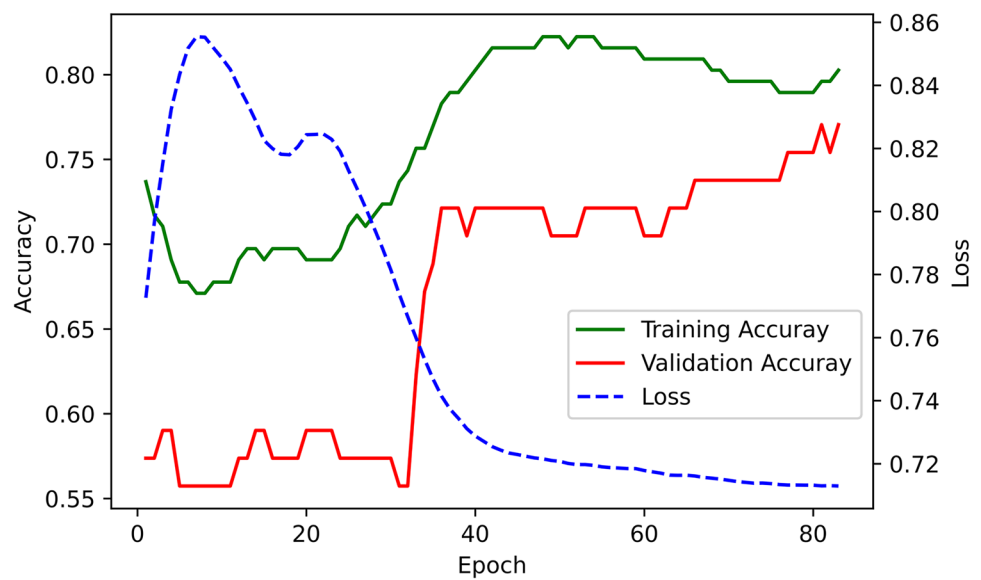


Fig. 3 The results of the MIFS heuristic for all k values between 1 and 17, showing the best combination of k features. Blue squares correspond to selected features

Fig. 4 Learning curves for training and test accuracy and training loss for the quantum neural network model



HGG from the baseline value, i.e., the mean Shapley value for all predictions. For instance, low tumor sphericity is associated with a positive Shapley value for HGG predictions, whereas for BM, a high sphericity value leads to an average negative Shapley value. Further, a high tumor LoG-3D-GLRLM-Run-Entropy that indicates increased texture heterogeneity contributes to negative Shapley values and shifts the prediction towards BM. A lower average GLRLM-Run-Entropy value is associated with a high positive mean Shapley value shifting the outcome significantly towards HGG. On another hand, high values of the two most important ring features, GLCM-Contrast (correlating with

disparity in intensity values among neighboring voxels) and Wavelet-HHL-GLCM-Cluster-Prominence (implying more asymmetry about the mean voxel value), are associated with negative Shapley values pushing the prediction towards BM, and the opposite is true for HGG.

Figure 6b highlights differences in Shapley value and feature values between true and false predictions on the test set. The largest mean Shapley value magnitudes for the false predictions were observed for low tumor sphericity combined with a negative Shapley value, high ring Wavelet-HHL-GLCM-Cluster-Prominence combined with a positive Shapley value, and low tumor Wavelet-LHH-GLSZM-Zone-Entropy combined with a negative Shapley value.

Table 1 Models' performance metrics

		QNN	XGBoost	DNN
Training ACC		0.8	0.89	0.88
Test bACC		0.74	0.72	0.73
Training ROC-AUC		0.86	0.94	0.95
Test ROC-AUC		0.76	0.79	0.75
Test precision		0.77	0.75	0.77
	BM	0.80	0.79	0.79
	HGG	0.7	0.68	0.72
Test recall		0.77	0.75	0.77
	BM	0.85	0.85	0.87
	HGG	0.64	0.59	0.59
Test F1-score		0.77	0.75	0.76
	BM	0.83	0.81	0.83
	HGG	0.67	0.63	0.65

QNN quantum neural network, *XGBoost* extreme gradient boosting, *DNN* dense neural network, *ACC* accuracy, *bACC* balanced accuracy, *ROC-AUC* area under receiver operating characteristic curve, *BM* brain metastases, *HGG* high-grade glioma

Discussion

The integration of machine learning models in conjunction with other clinical instruments holds the potential to significantly enhance the accuracy of diagnosing and planning treatment for brain tumors, specifically LBM and HGG, using MRI scans. By leveraging the power of quantum computing, a QNN model can uncover subtle patterns and relationships within radiomic data, leading to more precise and reliable classification of brain tumors. This enhanced accuracy can aid healthcare professionals in making more informed decisions regarding treatment strategies, such as surgical intervention, radiation therapy, or targeted drug therapies. Developing an accurate QNN model could serve as a valuable tool for non-invasive and automated analysis of medical imaging data, reducing subjectivity and interobserver variability in tumor classification.

In this work, we developed an interpretable radiomic classifier to differentiate between large BM and HGG on

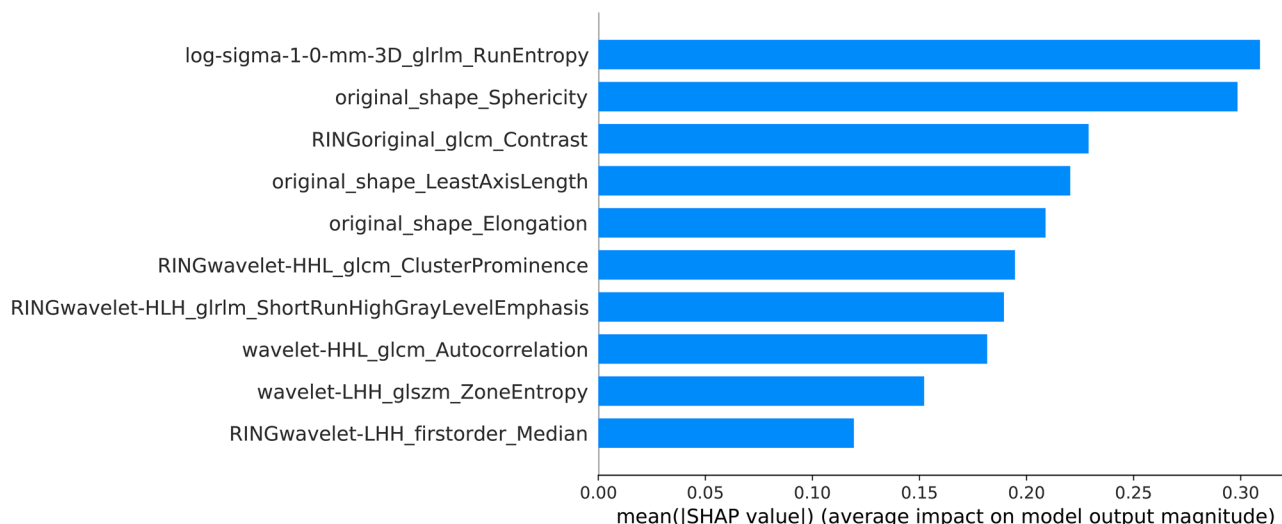


Fig. 5 Mean absolute Shapley value for each feature, ranked by order of importance

CE-T1W brain MRI. We used to this end a QNN model, and we showed that its performance on our dataset is comparable to two state-of-the-art classical algorithms: XGBoost and DNN. We employed a MIFS method and solved the resulting combinatorial optimization on D-Wave’s quantum annealer. To the extent of our knowledge, this is the first study to use quantum algorithms for brain tumor classification on a real-world dataset, and our results shed light on the potential value of these algorithms and the need for further exploration and development.

Since the role of feature selection is cardinal for reducing model complexity and leveraging interpretability, we applied a pipeline of two feature selection methods as this strategy was shown to improve the selection results [60, 61]. We used a MIFS method to get the best 10 features, since this method captures non-linear relationships and accounts for interaction

effects between features. The resulting QUBO, a computationally expensive optimization, was solved using a quantum annealing metaheuristic as it was shown to match or outperform other classical methods [36]. Nonetheless, the limited number of qubits and connectivity in quantum annealers can restrict the scale and complexity of implementable circuits, and this remains a considerable limitation in the current era.

While the QNN model developed herein showed similar performance on the test set as for XGBoost and DNN, the difference between training and validation metrics was smaller for QNN than for the classical algorithms despite using overfitting techniques for the latter, hinting to a better generalization and lower overfitting for the QNN. This is in line with the work of Caro et al. that showed a favorable generalization error for quantum machine learning algorithms [31], and it would be worthwhile testing this further in future studies.

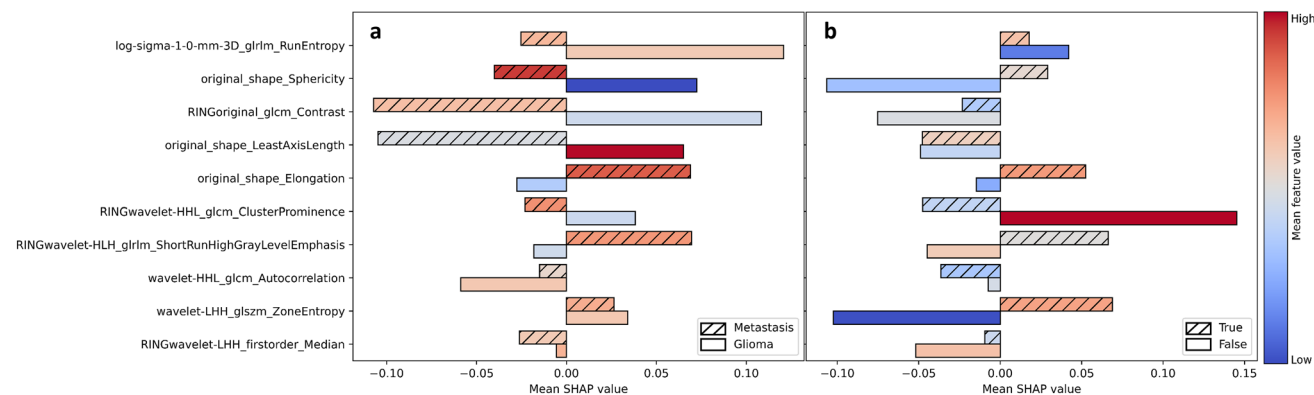


Fig. 6 Mean Shapley values and feature values for the different features according to the predicted class (metastasis or glioma) (a) and the prediction accuracy (true or false) (b). Colors represent the mean feature values according to the color scale bar

Differentiating between BM and HGG has been the subject of many previously published studies [41], including handcrafted radiomics-based machine learning techniques as well as deep learning models, with reported accuracy results ranging from 64 to 98%. Comparing results indirectly from separate studies is trivial, and such heterogeneity in classification performance could stem from many factors including the nature and quality of the data, the methods used, and the quality of reporting. In a systematic review by Jekel et al. [41], less than half of the TRIPOD items reflecting critical points in model development were reported on average. Moreover, most of the published papers do not report the tumor size, which could introduce significant classification bias due to the fact that BM are more likely to be diagnosed at an earlier growth stage than HGG owing to the close surveillance pattern of cancer patients.

de Causans and colleagues [62] developed a tumor-radiomics machine learning algorithm based on CE-T1W MRI to classify glioblastoma and solitary BM larger than 2cm, yielding a balanced accuracy of 0.8 and a ROC-AUC of 0.85 on the test set. Their results were slightly better than ours; however, it should be noted that imaging was done on the same 3-Tesla MRI machine using the same protocol, by contrast to our heterogeneous dataset including images from different institutions using different machines and acquisition parameters. Although heterogeneity of training data might affect the performance of a model, its ability to generalize to unseen data with different acquisition protocols might be better. Nonetheless, the ultimate test of generalization is through external validation. Heterogeneity has been shown to compromise the robustness of texture features, yet most of the selected texture features in this work were shown to have good repeatability for T1-weighted MRI with various MRI scanners and scanning parameters, particularly after pre-processing [63]. In this study, we adhered to the image pre-processing pipeline suggested by Carré et al. [47] to maximize radiomic feature stability. Moreover, 3 out of 10 features were shape-based and thus are very robust with respect to imaging techniques.

In this work, we focused on the model interpretability. Sphericity ranked second in terms of feature attribution to an overall prediction. This is consistent with the results of de Causans et al. [62] and Priya et al. [64] where tumor sphericity was the most important discriminating feature. From a clinical perspective, being able to see how each feature is influencing the decision process for a certain prediction would give more confidence in using the algorithm in practice. For instance, knowing that BM are more spherical on average than HGG, if an algorithm uses low sphericity of a tumor to significantly bias a prediction towards a BM prediction, the probability that this prediction is false increases. Our results support this rationality, and as we can see in Fig. 6b for sphericity, false predictions had high mean

absolute Shapley values attributing low sphericity to BM prediction. We acknowledge that some features are more complex to understand, and significant uncertainty remains as to extrapolating average values to individual instances.

Our analysis had some limitations. Despite being one of the largest studies to report on this subject, the sample size remains relatively small. Additionally, in this study, external validation was not conducted to provide a better evaluation of generalizability, although the heterogeneous nature of our dataset could potentially be advantageous in this aspect. This highlights the difficulty in gathering sufficient data for imaging-based machine learning studies in neuro-oncology and stresses on the need for data-sharing initiatives to improve model development and validation and push this field further towards clinical integration. Furthermore, we limited our model to CE-T1W MRI sequence. Adding more sequences might lead to improved prediction performance, although it might increase model complexity.

We used PennyLane's quantum simulator for the purpose of this study since our scope is to test the algorithm rather than the quantum device. Nevertheless, this algorithm could be easily implemented on a near-term quantum computer with reasonable error. The model architecture is scalable and could include a larger number of features; however, as the number of required qubits increases, the circuit becomes exponentially expensive to simulate classically. The rapid development in quantum device engineering and quantum error-correction algorithms is likely to allow practical implementation of such algorithms in the near future. However, we acknowledge that quantum technology is still in its early stages, and further developments are needed to fully exploit the advantages of quantum computing over its classical counterpart. Efforts are eagerly needed to explore and establish the utility of quantum machine learning algorithms in the field of medicine, more particularly in oncology.

Conclusion

In conclusion, this work led to the creation of a cutting-edge CE-T1W MRI-based radiomic quantum neural network classifier, employing quantum-informed mutual information feature selection. We were able to demonstrate that our developed model effectively discerns between large brain metastases and high-grade gliomas with remarkable accuracy. Notably, our model's performance was found to be on par with two prevailing state-of-the-art algorithms, namely, XGBoost and dense neural network, while seemingly exhibiting less susceptibility to overfitting issues. This algorithm warrants further external validation. We have also provided a game-theory approach for model interpretability using Shapley value, and our results were in line with published data. Further development of such algorithms using

quantum technology has significant implications for the field of medical imaging, paving the way for enhanced diagnostic capabilities and improved patient care in the realm of brain tumor classification.

Author Contribution All authors contributed to the study conception and design. Material preparation and data collection were performed by Tony Felefly, Camille Roukoz, and Sandrine Yazbek. Image pre-processing, code development, and analysis were performed by Tony Felefly. The first draft of the manuscript was written by Tony Felefly and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding This work was partially funded by the Saint Joseph University Research Council, Grant number: FS157.

Data Availability The data that support the findings of this study are available upon reasonable request from the corresponding author. The data are not publicly available due to privacy and ethical restrictions.

Declarations

Ethics Approval This study was performed in line with the principles of the Declaration of Helsinki. Approval was granted by the Ethics Committee of Saint Joseph University – Hôtel Dieu de France Hospital: Date: June 6, 2019; Number: CEHDF 1419.

Competing Interests The authors declare no competing interests.

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

References

- Ostrom QT, Price M, Neff C, Cioffi G, Waite KA, Kruchko C, et al. CBTRUS Statistical Report: Primary Brain and Other Central Nervous System Tumors Diagnosed in the United States in 2015–2019. *Neuro-Oncology*. 2022 Oct;24(Supplement_5):v1–95.
- Achrol AS, Rennert RC, Anders C, Soffietti R, Ahluwalia MS, Nayak L, et al. Brain metastases. *Nature Reviews Disease Primers*. 2019;5(1):5.
- Le Rhun E, Guckenberger M, Smits M, Dummer R, Bachelot T, Sahn F, et al. EANO-ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up of patients with brain metastasis from solid tumours. *Ann Oncol*. 2021;32(11):1332–47.
- Mohile NA, Messersmith H, Gatson NT, Hottinger AF, Lassman A, Morton J, et al. Therapy for Diffuse Astrocytic and Oligodendroglial Tumors in Adults: ASCO-SNO Guideline. *J Clin Oncol*. 2022;40(4):403–26.
- Weller M, van den Bent M, Preusser M, Le Rhun E, Tonn JC, Minniti G, et al. EANO guidelines on the diagnosis and treatment of diffuse gliomas of adulthood. *Nature Reviews Clinical Oncology*. 2021;18(3):170–86.
- Gondi V, Bauman G, Bradford L, Burri SH, Cabrera AR, Cunningham DA, et al. Radiation Therapy for Brain Metastases: An ASTRO Clinical Practice Guideline. *Pract Radiat Oncol*. 2022;12(4):265–82.
- WHO Classification of Tumours Editorial Board. *Central Nervous System Tumours: WHO Classification of Tumours*. 2021.
- Suh CH, Kim HS, Jung SC, Choi CG, Kim SJ. Perfusion MRI as a diagnostic biomarker for differentiating glioma from brain metastasis: a systematic review and meta-analysis. *European Radiology*. 2018;28(9):3819–31.
- Tsolaki E, Svolos P, Kousi E, Kapsalaki E, Fountas K, Theodorou K, et al. Automated differentiation of glioblastomas from intracranial metastases using 3T MR spectroscopic and perfusion data. *International Journal of Computer Assisted Radiology and Surgery*. 2013;8(5):751–61.
- Blasel S, Jurcoane A, Franz K, Morawe G, Pellikan S, Hattingen E. Elevated peritumoural rCBV values as a mean to differentiate metastases from high-grade gliomas. *Acta Neurochirurgica*. 2010;152(11):1893–9.
- Lin L, Xue Y, Duan Q, Sun B, Lin H, Huang X, et al. The role of cerebral blood flow gradient in peritumoral edema for differentiation of glioblastomas from solitary metastatic lesions. *Oncotarget*. 2016;7(42):69051–9.
- Bauer AH, Erly W, Moser FG, Maya M, Nael K. Differentiation of solitary brain metastasis from glioblastoma multiforme: a predictive multiparametric approach using combined MR diffusion and perfusion. *Neuroradiology*. 2015;57(7):697–703.
- Kadota Y, Hirai T, Azuma M, Hattori Y, Khant ZA, Hori M, et al. Differentiation between glioblastoma and solitary brain metastasis using neurite orientation dispersion and density imaging. *J Neuroradiol*. 2020;47(3):197–202.
- Suh CH, Kim HS, Jung SC, Kim SJ. Diffusion-Weighted Imaging and Diffusion Tensor Imaging for Differentiating High-Grade Glioma from Solitary Brain Metastasis: A Systematic Review and Meta-Analysis. *AJNR Am J Neuroradiol*. 2018;39(7):1208–14.
- Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More than Pictures, They Are Data. *Radiology*. 2016;278(2):563–77.
- Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RGPM, Granton P, et al. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer*. 2012;48(4):441–6.
- Afshar P, Mohammadi A, Plataniotis KN, Oikonomou A, Benali H. From Handcrafted to Deep-Learning-Based Cancer Radiomics: Challenges and Opportunities. *IEEE Signal Processing Magazine*. 2019;36(4):132–60.
- Hosny A, Parmar C, Coroller TP, Grossmann P, Zeleznik R, Kumar A, et al. Deep learning for lung cancer prognostication: A retrospective multi-cohort radiomics study. *PLoS Med*. 2018;15(11):e1002711.
- Carneiro G, Oakden-Rayner L, Bradley AP, Nascimento J, Palmer L. Automated 5-year mortality prediction using deep learning and radiomics features from chest computed tomography. In: 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017). 2017. p. 130–4.
- Kooi T, Litjens G, van Ginneken B, Gubern-Mérida A, Sánchez CI, Mann R, et al. Large scale deep learning for computer aided detection of mammographic lesions. *Med Image Anal*. 2017;35:303–12.
- Bae S, An C, Ahn SS, Kim H, Han K, Kim SW, et al. Robust performance of deep learning for distinguishing glioblastoma from single brain metastasis using radiomic features: model development and validation. *Scientific Reports*. 2020;10(1):12110.
- Valdez F, Melin P. A review on quantum computing and deep learning algorithms and their applications. *Soft comput*. 2022 7;1–20.

23. Biamonte J, Wittek P, Pancotti N, Rebentrost P, Wiebe N, Lloyd S. Quantum machine learning. *Nature*. 2017;549(7671):195–202.
24. Schuld M, Sinayskiy I, Petruccione F. The quest for a Quantum Neural Network. *Quantum Information Processing*. 2014;13(11):2567–86.
25. Dunjko V, Briegel HJ. Machine learning & artificial intelligence in the quantum domain: a review of recent progress. *Rep Prog Phys*. 2018;81(7):074001.
26. Schuld M, Petruccione F. *Supervised Learning with Quantum Computers*. 1st ed. Springer Cham; XIII, 287. (Quantum Science and Technology).
27. Liu Y, Arunachalam S, Temme K. A rigorous and robust quantum speed-up in supervised machine learning. *Nature Physics*. 2021;17(9):1013–7.
28. Abbas A, Sutter D, Zoufal C, Lucchi A, Figalli A, Woerner S. The power of quantum neural networks. *Nature Computational Science*. 2021;1(6):403–9.
29. Huang HY, Broughton M, Mohseni M, Babbush R, Boixo S, Neven H, et al. Power of data in quantum machine learning. *Nature Communications*. 2021;12(1):2631.
30. Mott A, Job J, Vlimant JR, Lidar D, Spiropulu M. Solving a Higgs optimization problem with quantum annealing for machine learning. *Nature*. 2017;550(7676):375–9.
31. Caro MC, Huang HY, Cerezo M, Sharma K, Sornborger A, Cincio L, et al. Generalization in quantum machine learning from few training data. *Nature Communications*. 2022;13(1):4919.
32. Schuld M. Supervised quantum machine learning models are kernel methods [Internet]. arXiv; 2021. Available from: <https://arxiv.org/abs/2101.11020>
33. Vergara JR, Estévez PA. A review of feature selection methods based on mutual information. *Neural Computing and Applications*. 2014;24(1):175–86.
34. Brown G, Pocock A, Zhao MJ, Luján M. Conditional Likelihood Maximisation: A Unifying Framework for Information Theoretic Feature Selection. *J Mach Learn Res*. 2012;13(1):27–66.
35. Nguyen XV, Chan J, Romano S, Bailey J. 2014. Effective global approaches for mutual information based feature selection. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD '14)*. Association for Computing Machinery, New York, NY, USA, 512–521. <https://doi.org/10.1145/2623330.2623611>
36. McGeoch CC, Wang C. Experimental Evaluation of an Adiabatic Quantum System for Combinatorial Optimization. In: *Proceedings of the ACM International Conference on Computing Frontiers [Internet]*. New York, NY, USA: Association for Computing Machinery; 2013. (CF '13). Available from: <https://doi.org/10.1145/2482767.2482797>
37. Rastogi R, Chaturvedi DK, Sagar S, Tandon N, Rastogi AR. Deep Learning Application in Classification of Brain Metastases: Sensor Usage in Medical Diagnosis for Next Gen Healthcare. In: *Nandan Mohanty S, Chatterjee JM, Satpathy S, editors. Internet of Things and Its Applications [Internet]*. Cham: Springer International Publishing; 2022. p. 117–35. Available from: https://doi.org/10.1007/978-3-030-77528-5_6
38. Dikaos N. Deep learning magnetic resonance spectroscopy fingerprints of brain tumours using quantum mechanically synthesised data. *NMR Biomed*. 2021;34(4):e4479.
39. Naeem A, Anees T, Naqvi RA, Loh WK. A Comprehensive Analysis of Recent Deep and Federated-Learning-Based Methodologies for Brain Tumor Diagnosis. *Journal of Personalized Medicine [Internet]*. 2022;12(2). Available from: <https://www.mdpi.com/2075-4426/12/2/275>
40. Stadlbauer A, Heinz G, Marhold F, Meyer-Bäse A, Ganslandt O, Buchfelder M, et al. Differentiation of Glioblastoma and Brain Metastases by MRI-Based Oxygen Metabolomic Radiomics and Deep Learning. *Metabolites [Internet]*. 2022;12(12). Available from: <https://www.mdpi.com/2218-1989/12/12/1264>
41. Jekel L, Brim WR, von Reppert M, Staib L, Cassinelli Petersen G, Merkaç S, et al. Machine Learning Applications for Differentiation of Glioma from Brain Metastasis-A Systematic Review. *Cancers (Basel)*. 2022;14(6):1369. Published 2022 Mar 8. <https://doi.org/10.3390/cancers14061369>
42. Avants BB, Tustison NJ, Johnson HJ. *Advanced Normalization Tools [Internet]*. github. Available from: <http://stnava.github.io/ANTs/>
43. Tustison NJ, Avants BB, Cook PA, Zheng Y, Egan A, Yushkevich PA, et al. N4ITK: improved N3 bias correction. *IEEE Trans Med Imaging*. 2010;29(6):1310–20.
44. Smith SM. Fast robust automated brain extraction. *Hum Brain Mapp*. 2002;17(3):143–55.
45. Isensee F, Schell M, Pflueger I, Brugnara G, Bonekamp D, Neuberger U, et al. Automated brain extraction of multisequence MRI using artificial neural networks. *Human Brain Mapping*. 2019;40(17):4952–64.
46. Reinhold JC, Dewey BE, Carass A, Prince JL. Evaluating the Impact of Intensity Normalization on MR Image Synthesis. *Proc SPIE Int Soc Opt Eng*. 2019;10949:109493H.
47. Carré A, Klausner G, Edjlali M, Lerousseau M, Briend-Diop J, Sun R, et al. Standardization of brain MR images across machines and protocols: bridging the gap for MRI-based radiomics. *Sci Rep*. 2020;10(1):12340.
48. van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Res*. 2017;77(21):e104–7.
49. Zwanenburg A, Vallières M, Abdalah MA, et al. The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping. *Radiology*. 2020;295(2):328–338. <https://doi.org/10.1148/radiol.2020191145>
50. Zwanenburg A, Vallières M, Abdalah MA, Aerts HJWL, Andrearczyk V, Apte A, et al. The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping. *Radiology*. 2020;295(2):328–38.
51. Pasvolsky J, Stevanovic R, Condello A, Tong H. *Feature Selection Notebook [Internet]*. D-Wave; 2021. Available from: <https://github.com/dwave-examples/feature-selection-notebook>
52. Bergholm V, Izaac J, Schuld M, Gogolin C, Ahmed S, Ajith V, et al. PennyLane: Automatic differentiation of hybrid quantum-classical computations [Internet]. arXiv; 2018. Available from: <https://arxiv.org/abs/1811.04968>
53. Farhi E, Neven H. Classification with Quantum Neural Networks on Near Term Processors [Internet]. arXiv; 2018. Available from: <https://arxiv.org/abs/1802.06002>
54. Schuld M, Bocharov A, Svore KM, Wiebe N. Circuit-centric quantum classifiers. *Physical Review A [Internet]*. 2020 Mar;101(3). Available from: <https://doi.org/10.1103/PhysRevA.101.032308>
55. Schuld M. Variational Classifier [Internet]. PennyLane; 2021. Available from: https://github.com/PennyLaneAI/qml/blob/master/demonstrations/tutorial_variational_classifier.py
56. Möttönen M, Vartiainen JJ, Bergholm V, Salomaa MM. Transformation of Quantum States Using Uniformly Controlled Rotations. *Quantum Info Comput*. 2005;5(6):467–73.
57. Shapley LS. *Notes on the n-Person Game -- II: The Value of an n-Person Game [Internet]*. Santa Monica, Calif.: RAND Corporation; 1951. Available from: https://www.rand.org/content/dam/rand/pubs/research_memoranda/2008/RM670.pdf
58. *The Shapley Value: Essays in Honor of Lloyd S. Shapley*. Cambridge University Press; 1988.
59. Lundberg SM, Lee SI. A Unified Approach to Interpreting Model Predictions. In: *Guyon I, Luxburg UV, Bengio S, Wallach H, Fergus R, Vishwanathan S, et al., editors. Advances in Neural Information Processing Systems 30 [Internet]*. Curran Associates, Inc.; 2017. p. 4765–74. Available from: <http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf>

60. Tsai CF, Hsiao YC. Combining multiple feature selection methods for stock prediction: Union, intersection, and multi-intersection approaches. *Decision Support Systems*. 2010;50(1):258–69.
61. Rokach L, Chizi B, Maimon O. Feature Selection by Combining Multiple Methods. In: Last M, Szczepaniak PS, Volkovich Z, Kandel A, editors. *Advances in Web Intelligence and Data Mining* [Internet]. Berlin, Heidelberg: Springer Berlin Heidelberg; 2006. p. 295–304. Available from: https://doi.org/10.1007/3-540-33880-2_30
62. de Causans A, Carré A, Roux A, Tauziède-Espariat A, Ammari S, Dezamis E, et al. Development of a Machine Learning Classifier Based on Radiomic Features Extracted From Post-Contrast 3D T1-Weighted MR Images to Distinguish Glioblastoma From Solitary Brain Metastasis. *Front Oncol*. 2021;11:638262.
63. Lee J, Steinmann A, Ding Y, Lee H, Owens C, Wang J, et al. Radiomics feature robustness as measured using an MRI phantom. *Scientific Reports*. 2021;11(1):3973.
64. Priya S, Liu Y, Ward C, Le NH, Soni N, Pillenahalli Maheshwarappa R, et al. Machine learning based differentiation of glioblastoma from brain metastasis using MRI derived radiomics. *Scientific Reports*. 2021;11(1):10478.

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