FULL PAPER

Predicting chemo-brain in breast cancer survivors using multiple MRI features and machine-learning

Vincent Chin-Hung Chen[1,2] | Tung-Yeh Lin[3] | Dah-Cherng Yeh[4] |
Jyh-Wen Chai[5,6] | Jun-Cheng Weng[2,3]

[1]School of Medicine, Chang Gung University, Taoyuan, Taiwan

[2]Department of Psychiatry, Chang Gung Memorial Hospital, Chiayi, Taiwan

[3]Department of Medical Imaging and Radiological Sciences, Chang Gung University, Taoyuan, Taiwan

[4]Breast Medical Center, Cheng Ching Hospital Chung Kang Branch, Taichung, Taiwan

[5]Department of Radiology, Taichung Veterans General Hospital, Taichung, Taiwan

[6]College of Medicine, China Medical University, Taichung, Taiwan

Correspondence:
Jun-Cheng Weng, Department of Medical Imaging and Radiological Sciences, Chang Gung University, No. 259, Wenhua 1st Road, Guishan District, Taoyuan City 33302, Taiwan.
Email: jcweng@mail.cgu.edu.tw

Funding information
Research programs MOST107-2221-E-182-054-MY3, MOST106-2221-E-182-079, MOST104-2314-B-040-001, and NSC103-2420-H-040-002 that were sponsored by the Ministry of Science and Technology, Taipei, Taiwan; the grant (BMRPD1H0101, BMRPD1G1321) of Chang Gung University, Taoyuan, Taiwan; and the grant (CORPG6G0101, CORPG6G0121) of Chang Gung Memorial Hospital, Chiayi, Taiwan.

Purpose: Breast cancer (BC) is the most common cancer in women worldwide. There exist various advanced chemotherapy drugs for BC; however, chemotherapy drugs may result in brain damage during treatment. When a patient's brain is changed in response to chemo drugs, it is termed chemo-brain. In this study, we aimed to construct machine-learning models to detect the subtle alternations of the brain in postchemotherapy BC patients.

Methods: Nineteen BC patients undergoing chemotherapy and 20 healthy controls (HCs) were recruited for this study. Both groups underwent resting-state functional MRI and generalized q-sampling imaging (GQI).

Results: Logistic regression (LR) with GQI indices in standardized voxel-wise analysis, LR with mean regional homogeneity in regional summation analysis, decision tree classifier (CART) with generalized fractional anisotropy in voxel-wise analysis, and XGBoost (XGB) with normalized quantitative anisotropy had formidable performances in classifying subjects into a chemo-brain group or an HC group. Classifying the brain MRIs of HC and postchemotherapy patients by conducting leave-one-out cross-validation resulted in the highest accuracy of 84%, which was attained by LR, CART, and XGB with multiple feature sets.

Conclusions: In our study, we constructed the machine-learning models that were able to identify chemo-brains from normal brains. We are hopeful that these results will be helpful in clinically tracking chemo-brains in the future.

KEYWORDS
breast cancer (BC), chemo-brain, generalized q-sampling imaging (GQI), machine learning, resting-state functional magnetic resonance imaging (rs-fMRI)

1 | INTRODUCTION

Breast cancer (BC), which takes more than 40,000 lives per year globally, is 1 of the most prevalent and deadly cancers. Countless amounts of money have been spent to develop therapies for BC, including chemotherapy, radiation therapy, and hormone therapy. Because of these advanced therapies, the mortality of BC has decreased in recent decades. As a result of higher survival rates, relieving the side effects of treatments has been emphasized to improve the quality of life of BC survivors.

Among these treatments, chemotherapy is a common, crucial treatment that utilizes chemotherapy drugs for inhibiting or killing cancer cells, and chemotherapy can also be combined with surgery or radiation therapy. According to several previous studies,[1,2] chemotherapy drugs proved harmful to human brains. Chemotherapy drugs can impair various cognitive functions, including memory, word finding, processing speed, and other diverse, basic cognitive functions; this damage is called chemo-brain. In a recent published study,[3] 42 BC subjects who had been treated with surgery and adjuvant chemotherapy were asked to complete questionnaires to investigate the influence of chemotherapy on their daily life. In their findings, there was no direct correlation between side effects and problems with attention and memory. For this reason, they suggested independent diagnostics for cognitive impairment. Therefore, an objective examination is now required to identify chemo-brains to track functional impairment postchemotherapeutically. An MRI, a common, formidable, and nonradiation brain imaging tool, is able to perform various imaging methods and was our chosen imaging tool for these studies.

The change of default mode network (DMN) in postchemotherapy BC survivors was previously proposed.[4] A recent study,[5] where researchers constructed a machine-learning model which could discriminate between chemotherapy-treated BC survivors and non-chemotherapy-treated BC survivors using the DMN resting state functional connectivity patterns of subjects, proved the utility of resting-state functional MRI (rs-fMRI) in detecting alternation in gray matter, and diffusion MRI, on the other hand, provides the information about white matter. Therefore, in this study, we aimed to construct machine-learning models by choosing rs-fMRI and diffusion MRI as our features. An fMRI can be categorized into either task-based fMRI or rs-fMRI. Task-based fMRI studies are conducted with unique stimulus protocols during the MRI scan. During an rs-fMRI, subjects are required to remain awake, close their eyes, keep their head still, and not think about anything particular. Because of basic differences between task-based fMRI and rs-fMRI, there are some advantages and disadvantages to these 2 methods. The 4 main advantages of rs-fMRI are as follows: (1) simplifying and facilitating the examination; (2) standardized and reproducible;

(3) no task requirement; and (4) wider use in clinical applications.[6] For diffusion MRI, we adopted generalized q-sampling imaging (GQI) as our feature set. GQI,[7] which is commonly used in diffusion MRI, is a nontensor tractographic method[8] based on the Fourier transform relation between diffusion signals and diffusion displacement.

Scikit-learn,[9] a well-known, publicly available library in Python, is a perfect platform for machine-learning research. Numerous preprocessing methods, feature selection methods, machine-learning algorithms, model evaluation metrics, etc., for machine-learning are included in scikit-learn, which allows operators to simply import and use them. For example, we imported StandardScaler, train_test_split, accuracy_score, classification_report, and all machine-learning models we used in this research. We tried various machine-learning models through the scikit-learn library to find the most suitable model for our data set because other machine-learning studies[10] also implemented comparing multiple models in a row. A total of 9 classification machine-learning models were used in this study: (1) logistic regression (LR); (2) decision tree classifier (CART); (3) XGBoost (XGB); (4) linear discriminant analysis (LDA); (5) Gaussian naïve Bayes (NB); (6) k-nearest neighbors classifier (KNN); (7) support vector machine (SVM); (8) multilayer perceptron (MLP); and (9) random forest (RF).

LR is a typical linear model based on the logistic function for classification rather than a regression model, despite "regression" in its name. Generally, 3 advantages of linear models (e.g., LR, LDA, NB, and linear SVM) could be proposed: (1) higher training and prediction speed; (2) appropriate for sparse data sets; and (3) appropriate for high-dimensional data. The utility of LR for classifying MRI images has been previously studied[11] when researchers aimed to construct LR models for classifying prostate cancer within the transition zone of an MRI. Radiologists participated in their study to validate the performance of these models, and their model reached or surpassed the radiologist. LDA, a rapid algorithm with closed-form solutions, was used in a previous study[12] where researchers developed models based on linear discriminant analysis to classify epileptic electroencephalogram data from epilepsy patients into normal, spike, and seizure. CART is a promising model that uses an if/else question-based machine learning algorithm and has gradually been emphasized and used widely more recently. Two main advantages of CART are: (1) easy to visualize the model and (2) feature scaling is not necessary. However, CART often tends to get overfitting and low performance on induction. Neuroscientific research[13] has developed a high-accuracy CART model that can differentiate multiple system atrophy from Parkinson's disease on a standard MRI and diffusion tensor imaging. NB is 1 of the linear models, but it sacrifices its performance on induction for even higher training speeds. Thus, NB is a perfect model for extremely massive

data set. A voxel-based NB model was developed to identify ischemic stroke lesions in T1-weighted images to provide an alternative to manual lesion delineation by experts.[14] An accuracy of 93% with sensitivity-specificity of 99% was attained by a study where researchers applied optimized NB and CART approaches on rs-fMRI of nicotine-dependent patients for classifying the subjects into N-acetylcysteine and placebo groups.[15] Furthermore, on the same topic, the XGB algorithm with 0.86 precision and an area under the curve (AUC) of 0.92 was used for classifying the subjects into relapse and nonrelapse classes using leave-one-out cross-validation (LOOCV).[16] XGB, eXtreme Gradient Boosting, is highly efficient, flexible, and accurate, which was based on the gradient boosting framework and has shown its impressive performance on many machine-learning competitions.

The KNN algorithm is a simple, nonparametric method that can be used for classification or regression. A KNN model could have reasonable performance without much tuning. However, there are some situations where researchers should not use KNN: (1) high-dimensional data sets, especially sparse data sets (low performance), and (2) massive data sets (low speed). Previous studies have[17] utilized KNN for automated segmentation and volumetric quantification of white matter lesions. The SVM, 1 of the most widely used machine-learning algorithms, is compatible with small amounts of data and has good performance on either low-dimensional or high-dimensional data. Nonetheless, SVM requires careful, proper tuning and preprocessing. In the field of medicine, many studies have used SVM for either classification or regression.[18,19] In a previous study, where researchers used 7 robust classifiers for discriminating between malignant and benign non-mass-enhancing lesions, SVM attained the highest accuracy of 94%.[20] MLP, a model based on the theory of deep-learning, is another strong, potential machine-learning algorithm adopted by many studies.[21] The flexibility of deep learning allows extremely complicated models which could predict at high accuracy. However, deep learning requires long training time, massive data, tuning, and data preprocessing, especially for those powerful, complicated models. The ensemble method family has recently become more popular and includes RF and bagging classifiers, etc., and implies a combination of multiple models that address several problems single models encounter. RF, the combination of several decision trees, addresses the overfitting often occurring in CART and shares many advantages with CART. Nonetheless, RF has a harder visualization, and it is inappropriate for high-dimensional and sparse data sets. With these machine-learning algorithms, we aimed to develop models that can identify chemo-brains from normal brains.

The goal of this study was to establish effective, accurate machine-learning models for identifying chemo-brains with rs-fMRI and GQI. To do this, we calculated the mean fractional amplitude of low-frequency fluctuations (mfALFF),

mean regional homogeneity (mReHo) map from rs-fMRI, generalized fractional anisotropy (GFA), quantitative anisotropy (QA), normalized quantitative anisotropy (NQA), as well as the isotropic value of the orientation distribution function (ISO) map from GQI for each subject. Consequently, a total of 6 different maps were collected as our feature set in this study. We then adopted 8 machine-learning models to use for classifications. In the future, we expect to use machine-learning as a clinical chemo-brain tracking model.

2 | METHODS

2.1 | Participants and clinical characteristics

All 39 participants, including 19 postchemotherapy BC subjects and 20 sex-matched healthy controls (HCs), were recruited from Taichung Veterans General Hospital in Taichung, Taiwan. Inclusion criteria for patients with BC were female sex, age between 20 and 55 years old, and histological confirmation of primary BC within 6 months after the completion of chemotherapy with standard chemotherapeutic agents (docetaxel and epirubicin); however, the GQI data of 1 of the BC subjects were not able to be analyzed. As a result, 20 HCs had rs-fMRI and GQI data and the rs-fMRI and GQI data of the chemo-brain group contained only 19 and 18 subjects, respectively. Among chemo-brain subjects, 4 received radiation therapy, 1 received hormone therapy, and 9 received surgical therapy. Five HCs and 5 chemo-brain subjects were postmenopausal. This study was approved by the Institutional Review Board of Taichung Veterans General Hospital (No. SF14185A). All participants participated in the study after providing informed consent, and all research was performed in accord with relevant guidelines and regulations.

2.2 | MRI acquisition

For all MRI examinations, subjects received a 1.5T MRI scan (Magnetom Aera; Siemens Medical Systems, Erlangen, Germany) with a standard, 8-channel head coil. T1-weighted 3D volume images were acquired with a magnetization prepared rapid gradient echo sequence (TR/TE/inversion time/flip angle [FA] = 9.11 ms/1.77 ms/450 ms/7°; 124 axial slices; and voxel size = $1.0 \times 1.0 \times 1.4$ mm^3). For fMRI measurements, a total of 256 volumes were acquired using a gradient echo-echo planar imaging sequence with 33 axial slices per volume (TR/TE/FA = 2000 ms/30 ms/90°; voxel size = $3.4 \times 3.4 \times 4.0$ mm^3; temporal resolution = 2 seconds). For GQI measurements, we performed a single-shot, diffusion-weighted spin echo-planar imaging sequence with the following parameters: magnetic field strength = 1.5T, repetition time = 7200 msec, echo time = 107 msec, field of view = 256 mm, matrix = 128×128, slice thickness = 4 mm, resolution = 2 ×

2×4 mm^3, b-values = 0, 1000, and 2000 sec/mm^2 in 129 noncollinear directions, number of excitations = 1, and the acquisition time was 16 minutes. During this examination, subjects were asked to lie down, stay still, keep their eyes closed, remain awake, and not think about anything in particular. To reduce motion and scanner noise, cushions and earmuffs were provided for participants. Other MRI pulse sequences, including axial T1-weighted imaging (T1WI), T2-weighted imaging (T2WI), fluid-attenuated inversion recovery (FLAIR), and coronal T2WI, were performed in both BC and HC groups. The additional postcontrast axial, coronal, and sagittal T1WI were performed in BC to rule out the possibility of brain metastasis, and the T2WI were performed for image coregistration and normalization.

2.3 | Resting-state functional MRI data analysis

Statistical Parametric Mapping 8 (SPM8; Wellcome Department of Cognitive Neurology, London, UK) based on MATLAB (The MathWorks, Inc., Natick, MA) was used for preprocessing. There were 7 steps during preprocessing, including slice-timing correction, motion correction, normalizing, spatial smoothing, nuisance regression, removing physiological noise, linear detrending, and bandpass temporal filtering. (1) Slice-timing correction was used for reducing the influence caused by different slice-timing in TR. (2) Following the slice-timing correction, motion correction was performed by selecting the center of each image and realigning the data to the first volume. (3) Next, affine transformation and resampling to isotropic 3-mm voxels on the data were performed for normalizing to Montreal Neurological Institute (MNI) standard space.[22] (4) For the purpose of better signal-to-noise ration gain, we spatially smoothed the data using a 6-mm full width at half maximum Gaussian kernel. (5) Nuisance regression was conducted using 6 head motion parameters as covariates. (6) The masks of whole brain, white matter, and cerebrospinal fluid were utilized for removing physiological noise. (7) Last, we used the Resting-State Data Analysis toolkit v1.8 (REST v1.8; Center for Cognition and Brain Disorders, Hangzhou Normal University, Zhejiang, China)[23] to further reduce physiological noise by performing linear detrending and bandpass temporal filtering (0.01–0.12 Hz) on the time series of each voxel.

After converting the time series of given voxel to frequency domain by fast Fourier transform, mfALFF was calculated in the frequency range of 0.01 to 0.12 Hz. The ALFF at a given voxel,[24] the square root of the power spectrum after averaging and normalizing across a predefined frequency interval, would be computed. The regional homogeneity (ReHo)[25] maps of subjects were calculated by giving each voxel a value corresponding to the Kendall's coefficient of concordance (KCC) of its time series with its nearest 26 neighboring voxels. Next, we adopted a mask for removing nonbrain tissues and noise on ReHo maps. Finally, we performed a standardization process on all subjects in which ReHo maps from each subject were divided by their own KCC for generating an individual mReHo.

2.4 | Generalized q-sampling imaging data analysis

For the Eddy current correction, we performed FSL Eddy correction (FMRIB Software Library, Oxford, UK). Following the Eddy current correction, we preprocessed our data using SPM8 (Wellcome Department of Cognitive Neurology). Affine transformation with 12 degrees of freedom and nonlinear warps were performed before mapping images to the standard MNI template. We then calculated 4 maps of GQI indices, including GFA, QA, NQA, and ISO maps, using DSI Studio (National Taiwan University, Taipei, Taiwan).[7]

2.5 | Machine-learning algorithms

As previously mentioned, we adopted multiple machine-learning models, including LR, CART, XGB, LDA, NB, KNN, SVM, MLP, and RF, for LOOCV. For observing the efficacy of standardization, we conducted the classification with and without standardization. The classification accuracy, confusion matrix, precision, recall, and f1-score were recorded. All machine-learning classifications were conducted using scikit-learn in Python. For results visualization, because of the several trials we did, we were not able to present all the results, but showed few models with better performances.

2.6 | Voxel-wise analysis

To acquire all values of each voxel belonging to gray matter in fMRI images, we first used all 90 regions from the Automated Anatomical Labeling (AAL) atlas for masking the MRI image. We then extracted the voxel values from the mask as our machine-learning features. White matter voxels in GQI maps were acquired though the masking of the International Consortium for Brain Mapping (ICBM) brain template[26] that divides white matter into 50 different regions. Likewise, all voxel values were extracted as our machine-learning features. As a result, 27,158 gray matter voxels within each rs-fMRI (mfALFF, mReHo) map and 30,344 white matter voxels within each GQI (GFA, ISO, NQA, and QA) map were selected to be independent features for our machine-learning models. The images from all of the imaging methods mentioned above cannot distinguish between the control group and chemo-brain group for physicians (Figure 1). Thus, we relied on machine-learning to identify chemo-brain.

FIGURE 1 GFA, NQA, and ISO of the HC group and chemo-brain group. A, GFA map of the HC, B, NQA map of the HC, C, ISO map of the HC, D, GFA map of the chemo brain group, E, NQA map of the chemo brain group, and F, ISO map of the chemo brain group. Differences between the HC group and chemo-brain group are virtually impossible to identify directly with human eyes. Thus, we tried to classify 2 groups using machine-learning models
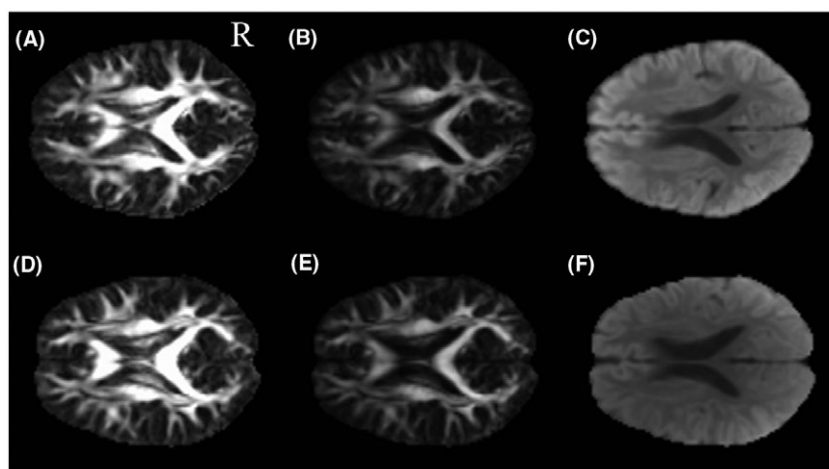
2.7 | Regional summation analysis

To address the issue of high-feature dimension that may have a bad influence on partial machine-learning models, we applied feature reductions to our data set. The AAL atlas helped us to segment whole gray matter and allowed us to segment the brain into 90 regions to sum all of the voxel values. In this case, the number of features was reduced to 90, and this could be a better method for some machine-learning models. White matter was also segmented into 50 regions and was summed up in every region using the ICBM template. Both voxel-wise analysis and regional summation were conducted using MATLAB (The MathWorks).

3 | RESULTS

3.1 | Participants and clinical characteristics

A total of 39 participants were recruited and enrolled: 19 female patients with BC (age range: 32–55 years, mean age:

43.8 ± 6.4 years; years of education: 13.9 ± 2.2; BC stage: I [n = 2], II [n = 14], and III [n = 3]) and 20 healthy female controls (age range: 43–55 years, mean age: 50.1 ± 2.5 years; years of education: 13.3 ± 2.3; Table 1).

3.2 | Voxel-wise analysis

In voxel-wise analysis (Table 2), LR, achieving over 70% accuracy, had an impressive performance on classifying chemo-brains and normal brains using GQI and rs-fMRI as input features. On a standardized set (Table 3), LR reached 84% accuracy with GFA, NQA, and QA. CART achieved 84% accuracy with GFA and over 70% accuracy with NQA and QA on both sets. XGB attained 84% accuracy with NQA and over 70% accuracy with GFA and QA. On the bottom row, we also showed the average accuracy of each machine-learning model versus every imaging method to compare the average performance of each machine-learning model. On both unstandardized and standardized sides, LR models reached the highest averaged accuracy.

TABLE 1 Demographic characteristics

Characteristics	Breast cancer patients after chemotherapy (n = 19)		Healthy controls (n = 20)		
	Mean or count	SD	Mean or count	SD	P value
Age (years)	43.8	6.4	50.1	2.5	0.001
Education (years)	13.9	2.2	13.3	2.3	0.435
Breast cancer stage (0, I, II, III, IV)	(0, 2, 14, 3, 0)	N/A	N/A	N/A	N/A
Chemotherapeutic drugs (docetaxel and epirubicin)	19	N/A	N/A	N/A	N/A
Radiation therapy	4	N/A	N/A	N/A	N/A
Hormonal treatment	1	N/A	N/A	N/A	N/A
Menopausal	5	N/A	5	N/A	N/A

Abbreviations: SD, standard deviation; N/A, not applicable.

TABLE 2 Classification accuracy of each machine-learning model versus each imaging method from leave-one-out validation of voxel-wise analysis

	LR	CART	XGB	LDA	NB	KNN	SVM	MLP	RF
mfALFF	64.1	30.77	41.03	64.1	64.1	58.97	7.69	53.85	61.54
mReHo	76.92	35.9	48.72	74.36	69.23	61.54	35.9	41.03	56.41
GFA	76.32	84.21	76.32	44.74	31.58	65.79	52.63	55.26	76.32
NQA	63.16	71.05	84.21	52.63	31.58	68.42	52.63	50	63.16
QA	60.53	76.32	78.95	39.47	31.58	55.26	52.63	57.89	52.63
ISO	71.05	50	55.26	65.79	31.58	65.79	52.63	39.47	47.37
Average	68.68	58.04	64.08	56.85	43.28	62.63	42.35	49.58	59.57

The features were the voxel values extracted directly from 6 maps. The bottom row is the average accuracy of each machine-learning model with every feature set.

TABLE 3 Classification accuracy of each machine-learning model versus each imaging method from standardized leave-one-out validation of voxel-wise analysis

	LR	CART	XGB	LDA	NB	KNN	SVM	MLP	RF
mfALFF	79.49	28.2	41.03	64.1	64.1	66.67	46.15	58.97	66.67
mReHo	79.49	51.28	48.72	74.36	69.23	61.54	58.97	71.79	61.54
GFA	84.21	84.21	76.32	44.74	31.58	68.42	57.89	68.42	55.26
NQA	84.21	76.32	84.21	55.26	31.58	73.68	55.26	65.79	50
QA	81.58	73.68	78.95	42.11	31.58	52.63	44.74	55.26	52.63
ISO	76.32	52.63	55.26	68.42	31.58	57.89	42.11	60.53	57.89
Average	80.88	61.05	64.08	58.17	43.28	63.47	50.85	63.46	57.33

Voxel values were standardized before classifying. The bottom row is the average accuracy of each machine-learning model with every feature set.

In addition to accuracy, we also recorded the confusion matrices (unstandardized: 16 of 20 HCs, 13 of 18 postchemo patients were correctly identified; standardized: 15 of 20 HCs, 17 of 18 postchemo patients were correctly identified) and the classification reports (unstandardized: precision = 0.76, recall = 0.76, f1-score = 0.76; standardized: precision = 0.86, recall = 0.84, f1-score = 0.84) of LR-GFA classification that was conducted using function in scikit-learn library.

To understand how machine-learning algorithms work, we extracted the coefficients from LR versus GQI indices in standardized voxel-wise analysis. We then aligned the coefficients back on MRI images for visualization (Figure 2). For brief visualization, we only presented the receiver operating characteristic curves and AUC of 3 GQI indices with LR, CART, and XGB in standardized voxel-wise analysis (Figure 3).

3.3 | Regional summation analysis

Unstandardized (Table 4) and standardized regional summation analyses (Table 5) were only 2 classifications showing an accuracy of over 80% and included the unstandardized and standardized LR-mReHo with an accuracy of 84.62% and 82.05%, respectively. In terms of the average of each machine-learning model versus each imaging method, LR also reached the highest average accuracy on both unstandardized and standardized sides in regional summation analysis.

Likewise, the confusion matrices (unstandardized: 16 of 20 HCs, 17 of 19 postchemo patients were correctly identified; standardized: 17 of 20 HCs, 15 of 19 postchemo patients were correctly identified) and the classification reports of LR-mReHo classifications (unstandardized: precision = 0.85, recall = 0.85, f1-score = 0.85; standardized: precision = 0.82, recall = 0.82, f1-score = 0.82) are shown.

4 | DISCUSSION

In the past decade, increased capabilities in the computer sciences have facilitated artificial intelligence and have excited researchers. In the field of medical imaging, machine-learning algorithms that have decision-making abilities are needed. The complex, subtle alternations that cannot be directly identified by physicians may now be detected by advanced machine-learning. Many scientific breakthroughs have exceeded current limitations. For example, 1 study of neural representations[27] adopted Gaussian naïve Bayes as machine-learning algorithms and fMRI image to classify subjects into HC groups and depression groups and reached 91% at the set of 17 suicidal ideators versus 17 controls. Former research[28] adopted
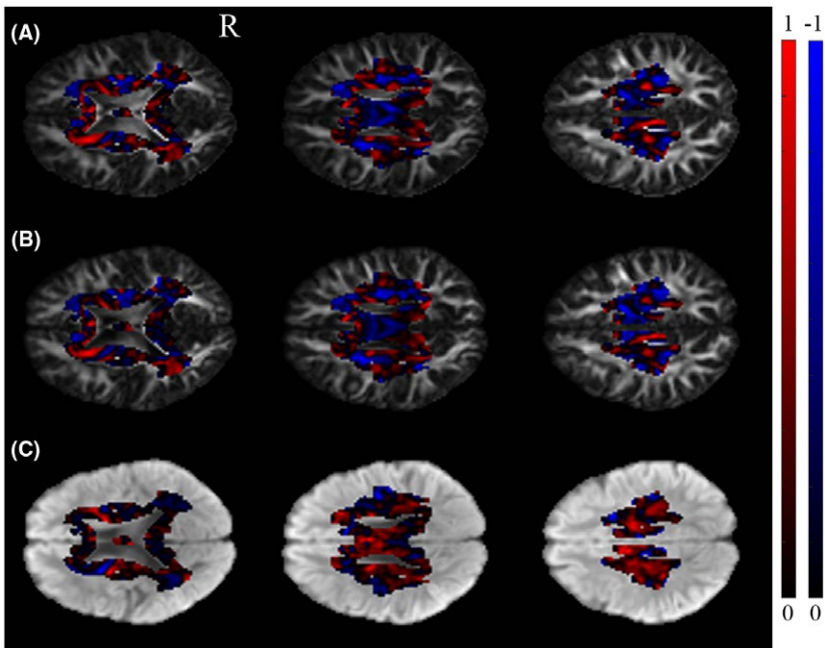
FIGURE 2 Coefficients of LR from A, LR-GFA, B, LR-NQA, and C, LR-ISO in standardized voxel-wise analysis. For brief visualization, we only show 3 different slides in this figure. The coefficients from the single model were separated into positive (red) and negative parts (blue). The positive and negative parts were rescaled between 0 and 1. We absolutized the negative part for color-coding and visualized both the positive and negative parts on GFA, NQA, and ISO maps
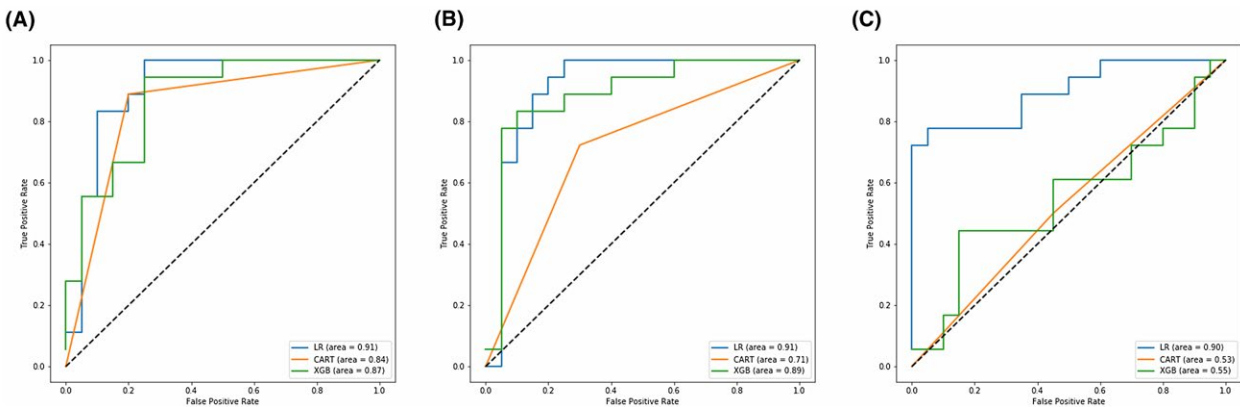

FIGURE 3 ROC curves and AUC of LR, CART, and XGB with A, GFA, B, NQA, and C, ISO. (AUC: LR-GFA = 0.91, CART-GFA = 0.84, XGB-GFA = 0.87, LR-NQA = 0.91, CART-NQA = 0.71, XGB-NQA = 0.89, LR-ISO = 0.90, CART-ISO = 0.53, and XGB-ISO = 0.55)

TABLE 4 Classification accuracy of each machine-learning model versus each imaging method from leave-one-out validation of regional summation analysis

	LR	CART	XGB	LDA	NB	KNN	SVM	MLP	RF
mfALFF	71.79	43.59	51.28	46.15	64.1	64.1	5.13	69.23	53.85
mReHo	84.62	56.41	66.67	56.41	76.92	61.54	23.08	48.72	74.36
GFA	73.68	39.47	52.63	50	42.11	73.68	52.63	42.11	50
NQA	63.16	52.63	47.37	47.37	50	50	52.63	57.89	55.26
QA	57.89	42.11	47.37	52.63	26.32	42.11	52.63	57.89	42.11
ISO	55.26	55.26	57.89	57.89	50	65.79	52.63	23.68	60.53
Average	67.73	48.25	53.87	51.74	51.58	59.54	39.79	49.92	56.02

The features are the regional summation of voxel values in each section of the brain. Depending on which template (MNI or ICBM template) map was used, there were 90 regions in the mfALFF and mReHo maps and 50 regions in GFA, NQA, QA, and ISO maps. The bottom row is the average accuracy of each machine-learning model with every feature set.

TABLE 5 Classification accuracy of each machine-learning model versus each imaging method from standardized leave-one-out validation of regional summation analysis

	LR	CART	XGB	LDA	NB	KNN	SVM	MLP	RF
mfALFF	69.23	48.72	51.28	46.15	64.1	58.97	56.41	64.1	43.59
mReHo	82.05	56.41	66.67	56.41	76.92	71.79	74.36	79.49	76.92
GFA	65.79	39.47	52.63	50	42.11	52.63	71.05	65.79	57.89
NQA	57.89	50	47.37	47.37	50	55.26	55.26	55.26	47.37
QA	60.53	47.37	47.37	52.63	26.32	44.74	44.74	57.89	50
ISO	63.16	55.26	57.89	57.89	50	52.63	55.26	60.53	52.63
Average	66.44	49.54	53.87	51.74	51.58	56.00	59.51	63.84	54.73

The regional summation data were standardized before classifying. The bottom row is the average accuracy of each machine-learning model with every feature set.

SVM classifiers to classify a data set of 20 chronic stroke participants who received brain-computer interface therapy and a resting-state functional connectivity scan at multiple time points (pre- and post-therapy groups) to analyze functional changes during the therapy. In LOOCV, an accuracy of 92.5% was achieved. Deep-learning, a specialized branch of machine-learning, is also a formidable model for classifying images into several groups. Convolution neutral network, a powerful image recognition tool based on deep learning, was used for Alzheimer's disease diagnoses because MRI images were the features.[29] An accuracy of 93.18% was achieved by their proposed model and showed the significance of machine-learning for an Alzheimer's disease diagnosis. For research on identifying chemo-brains, the study cited in the introduction of this paper[5] used DMN connectivity from rs-fMRI as features and SVM classifier for discriminating chemotherapy-treated (C^+) BC survivors from non-chemotherapy-treated (C^-) BC survivors and HCs. An accuracy of 91.23% and 90.74% were achieved in C^+ versus C^- and C^+ versus HC, respectively. More surprisingly, in their further work,[30] they constructed models which could accurately predict cognitive impairment at long-term follow-up with pretreatment rs-fMRI and RF. In comparison, we adopted GQI, voxel-wise analysis, regional summation, and multiple machine-learning models for revealing the effectiveness of these methods. We are hopeful that this research will benefit the establishment of clinical chemo-brain identification in postchemotherapy patients suffering from this functional impairment.

Before feeding data sets to our machine-learning models, we preprocessed our MRI images using various methods to optimize the prediction of our models, and this approach may be a better way to construct machine-learning models. For example, we used regional summation to reduce the dimension of features. We also tried other feature selection or reduction methods, such as principal component analysis (PCA), variance threshold, and select k best, among others. The utilization of feature selection is very common in constructing machine-learning models. Even though these feature selection methods worsened the results in our case, the application may

still be valid when constructing other machine-learning models. Recently, researchers developed a new machine-learning method to predict knee osteoarthritis progression on MRIs.[31] In their study, they adopted PCA for processing their feature set. Compared to our results, their results implied that PCA is effective for feature reduction and performance improvement. Similarly, PCA was used[32] to detect amyloid β-protein deposition on MRI images. We also tried feature combinations where multiple feature sets were fused into 1. For example, mfALFF and mReHo could be combined into 1 set for classifying; however, the results were not improved in this case and may be attributed to feature explosion caused by excess features.

Furthermore, we have determined that the choice of a model and preprocessing could be very influential to the outcome of machine-learning classifications. We suggest that researchers who have a small amount of data could try analyzing data with various models to find the most suitable model, considering that small data sets usually take less time to process. In a previous study,[33] researchers adopted multiple models and chest computed tomography images to predict whether a lung cancer patient would survive more than 2 years; they compared several machine-learning models with different feature selection methods to find the most suitable model for their study and found that RF had the highest average AUC. The SVM was the most accurate single classifier in another study on automatic sleep scoring[34] where researchers compared multiple classifier systems with many different single classifiers. The multiple classifier system used in that study surpassed the support vector machine and showed the potential of utilizing a multiple classifier system. In contrast, LR, CART, and XGB were the models that achieved the highest accuracy with the imaging methods in this study. In our study, all models maintained the default settings for reproducibility and simplicity. However, the information about tuning[9] is still discussed here. The GridSearchCV function in scikit-learn is a great function which tests every specified parameter values such as C value for an estimator. The C value and the penalty are the 2 parameters users have to tune for linear models such as LR and linear SVM. A high C value tends to classify all training examples correctly, and a low C value

tends to find a weight vector close to 0. For radial basis function kernel SVM, the default kernel for SVC in scikit-learn, the C value trades off misclassification of training examples against simplicity of the decision surface. A high C value also tends to classify all training examples correctly, and a low C value makes the decision surface smooth. With regard to penalty, L1 regularization is appropriate for the data sets have few truly discriminating features. Otherwise, the L2 regularization, the default setting in scikit-learn, should be used. The pre- and postpruning are the methods reducing overfitting for CART. So far, only the prepruning is available in scikit-learn. In its practical use, tuning of the max_depth could limit the size of the tree to prevent overfitting. General parameters, booster parameters, and learning task parameters are the 3 categories of parameters in XGB models. The tuning of XGB is a highly technical, complicated task. Therefore, we were not able to show all the contents here, but suggested that people who want to learn more about it view this detailed guide.[35] For other models, detailed information can be found in documentation of scikit-learn.

With regard to preprocessing, our results show that standardization may increase the accuracy of LR on both GQI and rs-fMRI sets. Other preprocessing steps, including rescaling or normalization, are also alternative methods that may be worth trying. A great issue often occurs when the ratio of training samples to dimensionality is low in overfitting, which is the exact problem most MRI researchers encounter. While overfitting, a model tends to learn the specific pattern of small samples rather than the general pattern it should learn. Applied in our research, LOOCV, which could reduce overfitting, is an appropriate method for a small data set. The reason why LOOCV or k-fold cross-validation reduces overfitting is that they are able to let models neglect data which might cause overfitting. Moreover, given that the difference in size between the training sets used in each fold and the entire data set is only a single pattern, LOOCV is approximately unbiased, which ensures the high reproducibility of LOOCV. The main disadvantage of LOOCV is long training time; however, it is not a great issue for small data sets such as ours. For features in voxel-wise analysis, GQI commonly showed better performance compared to rs-fMRI. Conversely, mReHo from the rs-fMRI performed better than GQI in regional analysis. Both imaging methods have the potential to identify chemo-brains in clinical use. Comparison between voxel-wise analysis and regional analysis showed that voxel-wise analysis was generally better for this study. According to our results, voxel-wise analysis is a better method to construct our model, and this result may be attributed to the over-reduction of regional summation.

4.1 | Limitations

The shortage of data was the main limitation in our study. The limitation of a low number of patients is common in medical imaging. However, in voxel-wise analysis, it can be neglected because different voxels can be treated as 1 patient (1 row in the final feature matrix). Some machine-learning methods require a large amount of training data to increase performance. Unfortunately, those methods would not have performed well in our case. We also did not set a prechemotherapy group to compare to the postchemotherapy group, and this may have been a more valid way to identify chemo-brain. Last, the use of partial models may have had a huge improvement in accuracy with some fine-tuning, and this was not achieved in our study because all default settings were used. The results of this study may be argued by researchers with better methods for model tuning.

5 | CONCLUSION

LR showed significant predictions on classifying rs-fMRI and GQI images into HCs and chemo-brain subjects. These results show the potential of machine-learning models for detecting chemo-brain that cannot be distinguished by the human eye directly. We support further research on machine-learning for identifying brain alternations in other diagnoses-needing situations.

REFERENCES

1. Hermelink K. Chemotherapy and cognitive function in breast cancer patients: the so-called chemo brain. *J Natl Cancer Inst Monogr*. 2015;2015:67–69.
2. Hislop JO. Yes, Virginia, chemo brain is real. *Clin Breast Cancer*. 2015;15:87–89.
3. Slowik AJ, Jablonski MJ, Michalowska-Kaczmarczyk AM, Jach R. Evaluation of quality of life in women with breast cancer, with particular emphasis on sexual satisfaction, future perspectives and body image, depending on the method of surgery. *Psychiatr Pol*. 2017;51:871–888.
4. Miao H, Chen X, Yan Y, et al. Functional connectivity change of brain default mode network in breast cancer patients after chemotherapy. *Neuroradiology*. 2016;58:921–928.
5. Kesler SR, Wefel JS, Hosseini SM, Cheung M, Watson CL, Hoeft F. Default mode network connectivity distinguishes chemotherapy-treated breast cancer survivors from controls. *Proc Natl Acad Sci U S A*. 2013;110:11600–11605.
6. Chen VC, Shen CY, Liang SH, et al. Assessment of brain functional connectome alternations and correlation with depression and anxiety in major depressive disorders. *PeerJ*. 2017;5:e3147.
7. Yeh FC, Wedeen VJ, Tseng WY. Generalized q-sampling imaging. *IEEE Trans Med Imaging*. 2010;29:1626–1635.
8. Panesar SS, Yeh FC, Jacquesson T, Hula W, Fernandez-Miranda JC. A quantitative tractography study into the connectivity,

segmentation and laterality of the human inferior longitudinal fasciculus. *Front Neuroanat*. 2018;12:47.

9. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: Machine Learning in Python. *J Mach Learn Res*. 2011;12:2825–2830.

10. Mannil M, Burgstaller JM, Held U, Farshad M, Guggenberger R. Correlation of texture analysis of paraspinal musculature on MRI with different clinical endpoints: Lumbar Stenosis Outcome Study (LSOS). *Eur Radiol*. 2018. doi: https://doi.org/10.1007/s00330-018-5552-6.

11. Dikaios N, Alkalbani J, Sidhu HS, et al. Logistic regression model for diagnosis of transition zone prostate cancer on multi-parametric MRI. *Eur Radiol*. 2015;25:523–532.

12. Wei C, Chia-Ping S, Ming-Jang C, et al. Epileptic EEG visualization and sonification based on linear discriminate analysis. *Conf Proc IEEE Eng Med Biol Soc*. 2015;2015:4466–4469.

13. Nair SR, Tan LK, Mohd Ramli N, Lim SY, Rahmat K, Mohd NH. A decision tree for differentiating multiple system atrophy from Parkinson's disease using 3-T MR imaging. *Eur Radiol*. 2013;23:1459–1466.

14. Griffis JC, Allendorfer JB, Szaflarski JP. Voxel-based Gaussian naive Bayes classification of ischemic stroke lesions in individual T1-weighted MRI scans. *J Neurosci Methods*. 2016;257:97–108.

15. Tahmassebi A, Gandomi AH, Schulte M, Goudriaan AE, Foo SY, Meyer-Baese A. Optimized naive-Bayes and decision tree approaches for fMRI smoking cessation classification. *Complexity*. 2018;2018:24.

16. Tahmassebi A, Gandomi A, McCann I, Schulte M, Goudriaan A. Deep learning in medical imaging: fMRI big data analysis via convolutional neural networks. In Proceedings of the Practice and Experience on Advanced Research Computing, Pittsburgh, PA, 2018. pp. 1–4.

17. Steenwijk MD, Pouwels PJ, Daams M, et al. Accurate white matter lesion segmentation by k nearest neighbor classification with tissue type priors (kNN-TTPs). *Neuroimage Clin*. 2013;3:462–469.

18. Ding X, Yang Y, Stein EA, Ross TJ. Multivariate classification of smokers and nonsmokers using SVM-RFE on structural MRI images. *Hum Brain Mapp*. 2015;36:4869–4879.

19. Weis C, Hess A, Budinsky L, Fabry B. In-vivo imaging of cell migration using contrast enhanced MRI and SVM based post-processing. *PLoS ONE*. 2015;10:e0140548.

20. Illan IA, Tahmassebi A, Ramirez J, et al. Machine learning for accurate differentiation of benign and malignant breast tumors presenting as non-mass enhancement. In Proceedings of the SPIE Commercial + Scientific Sensing and Imaging, Orlando, FL, 2018. p. 8.

21. Leuthardt EC, Guzman G, Bandt SK, et al. Integration of resting state functional MRI into clinical practice—a large single institution experience. *PLoS ONE*. 2018;13:e0198349.

22. Collins DL, Neelin P, Peters TM, Evans AC. Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space. *J Comput Assist Tomogr*. 1994;18:192–205.

23. Chao-Gan Y, Yu-Feng Z. DPARSF: a MATLAB toolbox for "pipeline" data analysis of resting-state fMRI. *Front Syst Neurosci*. 2010;4:13.

24. Yue Y, Jia X, Hou Z, Zang Y, Yuan Y. Frequency-dependent amplitude alterations of resting-state spontaneous fluctuations in late-onset depression. *BioMed Res Int*. 2015;2015:505479.

25. Zang Y, Jiang T, Lu Y, He Y, Tian L. Regional homogeneity approach to fMRI data analysis. *NeuroImage*. 2004;22:394–400.

26. Mazziotta J, Toga A, Evans A, et al. A probabilistic atlas and reference system for the human brain: International Consortium for Brain Mapping (ICBM). *Philos Trans R Soc Lond B Biol Sci*. 2001;356:1293–1322.

27. Just MA, Pan L, Cherkassky VL, et al. Machine learning of neural representations of suicide and emotion concepts identifies suicidal youth. *Nat Hum Behav*. 2017;1:911–919.

28. Mohanty R, Sinha AM, Remsik AB, et al. Machine learning classification to identify the stage of brain-computer interface therapy for stroke rehabilitation using functional connectivity. *Front Neurosci*. 2018;12:353.

29. Islam J, Zhang Y. Brain MRI analysis for Alzheimer's disease diagnosis using an ensemble system of deep convolutional neural networks. *Brain Inform*. 2018;5:2.

30. Kesler SR, Rao A, Blayney DW, et al. Predicting long-term cognitive outcome following breast cancer with pre-treatment resting state fMRI and random forest machine learning. *Front Hum Neurosci*. 2017;11:555.

31. Du Y, Almajalid R, Shan J, Zhang M. A novel method to predict knee osteoarthritis progression on MRI using machine learning methods. *IEEE Trans Nanobioscience*. 2018;17:228–236.

32. Yan F, Li Y, Zhu X, et al. Detection algorithm of amyloid beta-protein deposition in magnetic resonance image based on pixel feature learning method. [Article in Chinese; Abstract available in Chinese from the publisher]. *Sheng Wu Yi Xue Gong Cheng Xue Za Zhi*. 2017;34:431–438.

33. Parmar C, Grossmann P, Bussink J, Lambin P, Aerts HJ. Machine learning methods for quantitative radiomic biomarkers. *Sci Rep*. 2015;5:13087.

34. Gao V, Turek F, Vitaterna M. Multiple classifier systems for automatic sleep scoring in mice. *J Neurosci Methods*. 2016;264:33–39.

35. Jain A. Complete Guide to Parameter Tuning in XGBoost (with codes in Python). 2016; https://www.analyticsvidhya.com/blog/2016/03/complete-guide-parameter-tuning-xgboost-with-codes-python/. Accessed November 2, 2018.