Automated Classification of Types of Brain Tumor in T1-weighted MR Images: A Thorough Comparative Study

Lim Jia Qi[1, a)], Norma Alias[2, b)], Farhana Johar[3, c)]

1, 2, 3Department of Mathematical Sciences, Faculty of Science, 81310 UTM Johor Bahru, Johor, Malaysia

[a)] Corresponding author: ykj10@hotmail.com
[b)]norma@ibnusina.utm.my
[c)]farhanajohar@utm.my

Abstract. Undoubtedly, early detection and characterization of brain tumor is critical in clinical practices. Automated diagnosis using neuroimaging tool like MRI guided by machine learning approaches has been the focus of numerous researches. In this study, various feature extraction, dimensionality reduction and supervised classification models are explored, evaluated and compared under different finite number of features to identify the optimal pathway/pipeline for classification of types of brain tumor, namely meningioma, glioma and pituitary tumor. The performance metrics utilized include accuracy, Kappa statistic, sensitivity, precision, F-measure, training time and test time. Results show that RBF SVM (pairwise coupling) under 80 PLS features achieved the highest average accuracy ($95.02\% \pm 0.19\%$) among all other machine learning pipelines.

INTRODUCTION

Brain tumor, particularly malignant tumor is life threatening. According to latest statistics from [1], 5 years survival rate of patients diagnosed with malignant brain tumor and other central nervous system (CNS) is roughly 34% for men and 36% for women. In view of this, early detection and characterization of brain tumor is of paramount importance to conduct more accurate diagnosis, prognosis, treatment planning and prediction of therapeutic responses from patients.

With rapid advances in biomedical imaging technologies, vast amount of high resolution medical images, such as contrast enhanced magnetic resonance (MR) scan images [2] is nowadays a vital tool for studying human brain anatomy and disease detection, e.g. brain tumor detection, Alzheimer's disease, multiple sclerosis and other neurological disorders. Lately, there has been a surge in interest in the study in automated and semi-automated characterization of brain tumor with the aid of MR images, in conjunction with the rapid development of computer vision and machine learning algorithms. Besides offering high resolution multi-modality images, MRI scan also does not emit hazardous emission. Due to its superior soft tissue differentiation, MR images can provide useful semantic information on the types of tumor, and ultimately assist clinical experts and radiologists to make unbiased diagnosis.

In this study, we will be focusing on the brain tumor type classification problems, namely meningioma, glioma and pituitary tumor, in which meningioma (37%), glioma (27%) are two of the most common brain tumor, while pituitary tumor accounted for around 16% of all primary brain tumor. Previously, Zacharaki et. al [3] attempted the multiclass classification of different brain tumor grades, e.g. glioma grade II, glioma grade III, glioblastoma and metastasis using one versus all SVM voting scheme. In the work of [4], classification of multiple brain tumor types: astrocytoma, glioblastoma multiforme, childhood tumor-medulloblastoma, meningioma, secondary tumor metastatic, and normal regions were conducted. It was found that PCA-ANN achieved the highest overall accuracy of 85.23%. Multiple brain tumor types and grades classification using SVM had been proposed by [5]. An accuracy of 85% and 78.26% were achieved respectively when second order features was utilized. Jayachandran et. al [6] conducted multi

Proceedings of the 27th National Symposium on Mathematical Sciences (SKSM27)
AIP Conf. Proc. 2266, 080001-1–080001-16; https://doi.org/10.1063/5.0018056
Published by AIP Publishing. 978-0-7354-2029-8/$30.00

class brain tumor classification using hybrid structure descriptor and fuzzy logic based pair of RBF kernel SVM. . Classification accuracy of proposed system in meningioma is 98.6%, metastasis is 99.29%, gliomas grade II is 97.87% and gliomas grade III is 98.6%.

This task is more challenging compared to binary classification problem of normal and neoplastic brain MR images because it is a multiclass classification problem, in which some binary classifiers like logistic regression and support vector machine (SVM) cannot be readily implemented. In addition, instead of using whole images, the ambiguous tumor regions have to be identified and isolated by either human experts or image segmentation algorithm before feature extraction and subsequent classification can be carried out. A wide ranges of feature extraction, feature transform and reduction as well as pattern recognition algorithms will be experimented and their performances will be evaluated and compared in terms of accuracy, Kappa statistic, sensitivity, precision, F-measure, training time and test time. With extensive experimentation and its results interpretation, the best pathway for the development of automated computer aided diagnosis (CAD) can be identified.

MATERIALS AND METHODS

The overall pipeline of the research framework is outlined in Fig. 1.

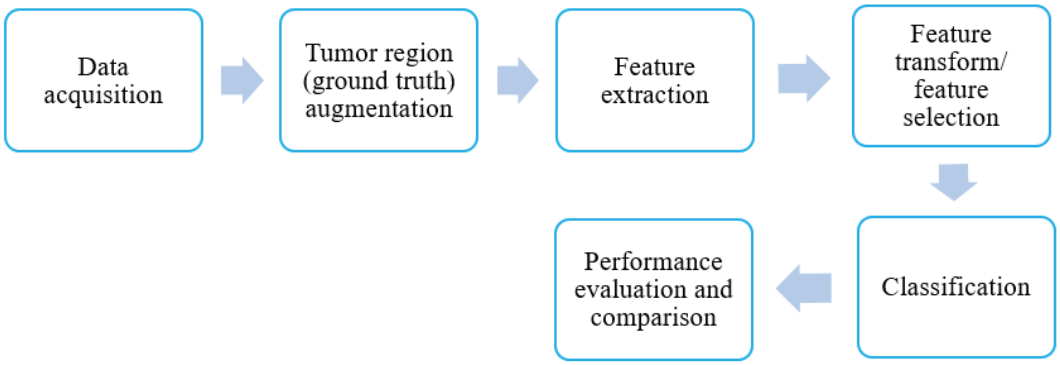

FIGURE 1. Brain tumor types classification framework.

Data acquisition

The brain MR images are downloaded from publicly available online database. A total of 3064 slices of T1-weighted contrast enhanced MR images from 233 patients was downloaded from https://figshare.com/articles/brain_tumor_dataset/1512427. There are three kinds of brain tumor in the MR images downloaded, namely meningioma (708 slices), glioma (1426 slices), and pituitary tumor (930 slices). The brain T1-weighed CE-MRI dataset was acquired from Nanfang Hospital, Guangzhou, China, and General Hospital, Tianjing Medical University, China, from year 2005 to 2010. The images have an in-plane resolution of 512×512 with pixel size 0.49×0.49 mm^2. The slice thickness is 6 mm and the slice gap is 1 mm [7, 8].

Tumor region augmentation

The 2D brain MR images comes with ground truth tumor region delineated manually by experienced radiologists. As pointed out in [9], tissues surrounding tumors can provide useful discriminative information about the types of tumor. Thus, augmentation of tumor region can be beneficial in extracting robust features. In this study, augmentation of tumor region is performed by morphological dilation with disk-shape structuring element with radius, R of 8.

Feature extraction

Simply speaking, feature extraction is a dimensionality reduction method, in which instead of utilizing pixels intensity in raw images, a set of attributes, also known as feature vector is constructed to represent a certain image. This stage is crucial as it can directly impacts the generalization performance of classification models [10]. Ideal set of feature should be relevant, non-redundant, provide low dimensional representation for visualization, and increase training and inference speed of learning algorithms [11]. In this paper, several feature extraction techniques are employed, including shape parameters, geometric moment invariants [12, 13], Zernike moments [14, 15], pseudo Zernike moments [16], histogram of oriented gradients (HOG) [17], linear binary pattern (LBP) [18], and bag of words (BOW) model [19]. It is worth noting that the image features will be extracted from the augmented tumor region. The number of features extracted from each methods are summarized in Table 1. Therefore, the whole feature matrix dimension is 3064×1711. This feature matrix will be linearly projected to new dimensional subspace. We spotted some predictors possess the same values throughout the data. As such, prior to feature selection methods, these non-discriminative features are removed.

Table 1. Feature extraction and its number of features.

Feature extraction methods	Number of features
Binary shape parameters	11
Geometric moment invariants	7
Zernike moments	12
Pseudo Zernike moments	15
HOG	900
LBP	266
Bag of words model	500

Binary shape parameters

Shape is definitely an important cue for human to identify and recognize the real world objects [20]. Shape features can be divided into 2 categories: contour based and region based method. Contour based method will be our focus because of low computation complexity and the shape descriptors are generally intuitive. These descriptors can only discriminate shapes with big differences, and thus has to be combined with other features for downstream classification tasks [21]. These shape attributes only requires binary image, in which the labeled tumor region has a value of '1' and '0' otherwise. Table 2 shows the list of shape parameters and their respective formula and details.

Table 2. List of binary shape parameters and its corresponding formula. x and y denote the indices of row and column of an image. I(x,y) denotes the binary image pixels.

Shape parameters	Formula	Details
Area	$A = \sum_{x=1}^{height} \sum_{y=1}^{width} I(x,y)$	Size of tumor.
Center of area	$\tilde{x} = \frac{1}{A} \sum_{x=1}^{height} \sum_{y=1}^{width} x.I(x,y)$ $\tilde{y} = \frac{1}{A} \sum_{x=1}^{height} \sum_{y=1}^{width} y.I(x,y)$	Row and column coordinates of the center of tumor.
Axis of least second moment	$\theta = \frac{1}{2}\tan^{-1} \frac{2\sum_{x=1}^{height}\sum_{y=1}^{width}(x-\bar{x})(y-\bar{y})I(x,y)}{\sum_{x=1}^{height}\sum_{y=1}^{width}(x-\bar{x})^2 I(x,y) - \sum_{x=1}^{height}\sum_{y=1}^{width}(y-\bar{y})^2 I(x,y)}$	Orientation of tumor
Perimeter		Counting the number of '1'

	$P_i = \frac{\pi}{4} \times Perimeter$	pixels with '0' pixels as neighbor. Apply Canny edge detector to obtain tumor border.
Thinness ratio	$T = 4\pi\left(\frac{A}{P_i^2}\right)$	Measure of roundness
Irregularity ratio	$IR = \frac{1}{T}$	Regularity of an object.
Aspect ratio	$AR = \frac{y_{max} - y_{min} + 1}{x_{max} - x_{min} + 1}$	Relative object spread in both vertical and horizontal directions.

Geometric moment invariants

Geometric moment invariants (GMI), also known as Hu's moment invariants was initially discovered by Hu [22], and had since being widely employed in various applications [23]. The most desirable properties of GMI is translation, scaling and rotation invariance, which is imperative in content based image retrieval as well as pattern recognition and object detection in images. However, it should be noted that the basis function of GMI is not orthogonal that may lead to information redundancy. Before the formula of seven Hu's moment invariants is introduced, the discrete implementation of $(p+q)^{th}$ order moments, central moment and normalized central moment are defined as below:

$$m_{pq} = \sum_{x=1}^{height} \sum_{y=1}^{width} x^p y^q f(x,y) \tag{1}$$

$$\mu_{pq} = \sum_{x=1}^{height} \sum_{y=1}^{width} (x-\bar{x})^p (y-\bar{y})^q f(x,y) \tag{2}$$

Where $f(x,y)$ is the binary tumor region, $\bar{x} = \frac{m_{10}}{m_{00}}$, $\bar{y} = \frac{m_{01}}{m_{00}}$

$$\nu_{pq} = \frac{\mu_{pq}}{\mu_{00}{}^{\gamma}}, \qquad where\ \gamma = \frac{p+q}{2} + 1 \tag{3}$$

Based on the normalized central moments, ν_{pq}, seven GMI are defined as follow:

$$\phi_1 = \nu_{20} + \nu_{02} \tag{4}$$

$$\phi_2 = (\nu_{20} - \nu_{02})^2 + 4\nu_{11}{}^2 \tag{5}$$

$$\phi_3 = (\nu_{30} - 3\nu_{12})^2 + (3\nu_{21} - \nu_{03})^2 \tag{6}$$

$$\phi_4 = (\nu_{30} + \nu_{12})^2 + (\nu_{21} + \nu_{03})^2 \tag{7}$$

$$\phi_5 = (\nu_{30} - 3\nu_{12})(\nu_{30} + \nu_{12})[(\nu_{30} + \nu_{12})^2 - 3(\nu_{21} + \nu_{03})^2] + (3\nu_{21} - \nu_{03})(\nu_{21} + \nu_{03})[3(\nu_{30} + \nu_{12})^2 - (\nu_{21} + \nu_{03})^2] \tag{8}$$

$$\phi_6 = (\nu_{20} - \nu_{02})[(\nu_{30} + \nu_{12})^2 - (\nu_{21} + \nu_{03})^2] + 4\nu_{11}(\nu_{30} + \nu_{12})(\nu_{21} + \nu_{03}) \tag{9}$$

$$\phi_7 = (3\nu_{21} - \nu_{03})(\nu_{30} + \nu_{12})[(\nu_{30} + \nu_{12})^2 - 3(\nu_{21} + \nu_{03})^2] - (\nu_{30} - 3\nu_{12})(\nu_{21} + \nu_{03}) \tag{10}$$

$$[3(\nu_{30}+\nu_{12})^2-(\nu_{21}+\nu_{03})^2]$$

Zernike moment (ZM) & pseudo Zernike moment (PZM)

Zernike moment initiated by the work of Teague [24] is another moment techniques with the use of Zernike polynomials as basis function for the moment calculation. Unlike GMI, Zernike polynomial used is orthogonal, which can reduce information between moments. There are several properties of ZM that make it a very effective global image descriptors, including rotational invariance, ease of image reconstruction from ZM [25], tolerance to noise, expression efficiency and multi-level representation [26]. The conventional way of calculating Zernike moments suffer from discretization errors as pointed out by [27], thus accurate calculation of ZM through geometric moments is applied in this research. Zernike moments can be defined in terms of geometric moments according to [23]:

$$A_{pq}=\frac{p+1}{\pi}\sum_{k=|q|}^{p}B_{p|q|k}\sum_{l=0}^{u}\sum_{m=0}^{|q|}w^m\binom{u}{l}\binom{|q|}{m}G_{k-2l-m,2l+m} \tag{11}$$

Where $u=\frac{1}{2}(k-|q|)$; $w=-j$ when $q>0$ while $w=+j$ when $q\leq 0$

$$B_{pqk}=\frac{(-1)^{\left(\frac{p-k}{2}\right)}\left(\frac{p+k}{2}\right)!}{\left(\frac{p-k}{2}\right)!\left(\frac{k+q}{2}\right)!\left(\frac{k-q}{2}\right)!} \tag{12}$$

$$G_{nm}=\sum_{x=0}^{height-1}\sum_{y=0}^{width-1}f(x,y)g_{nm,ik} \tag{13}$$

Where $g_{nm,ik}=\frac{1}{(n+1)(m+1)}(a_{i+1}^{n+1}-a_i^{n+1})(b_{i+1}^{n+1}-b_i^{n+1})$

$$a_i=\frac{2i-N}{N},\qquad b_k=\frac{2k-N}{N} \tag{14}$$

Similar to ZM, pseudo Zernike moment (PZM) also belongs to the class of orthogonal moment. However, PZM can produce twice the number of moments in comparison to ZM provided the same moment order [16]. In this paper, fast calculation of PZM using q recursive method will be employed [28].

HOG

HOG proposed by [17] is a local feature descriptor that is able to capture edge gradient structure with high tolerance to local geometric and photometric (illumination) transformation [29] and simultaneously maintain high selectivity [30]. These features is well known for preserving local second-order interaction between pixels [31]. In this research, we first resize the augmented tumor images to be 192×192 before applying Sobel filter kernel on to get the gradients. Histogram of gradient are calculated in each 32×32 cells and normalized with 64×64 blocks.

LBP

Ojala et. al [18] introduced a grayscale and rotation invariant feature, known as local binary pattern (LBP). LBP is a simple yet efficient operator in depicting local image pattern and has been utilized as features for various

application [32].Aside from the implementation of original LBP, an improved version of locally rotation invariant LBP [33] is applied for better tradeoff between discriminative power and robustness.

Bag of words (BOW) model

BOW techniques was originally introduced in text document analysis and retrieval and had extended its application in feature extraction in computer vision domain [34]. Prior to BOW, a set of local descriptors have to be defined for each images. In this paper, local image patches of size 5×5 is used. In BOW model, there are two main steps in converting local descriptors into image representation: coding and pooling [35]. Coding can be defined as an operation whereby local descriptors are encoded by codebook or vocabulary and the response of feature on this codebook is generated either through probability function or sparse coding [36]. On the other hand, pooling is the process of transforming feature response into image representation that preserves vital information.[37]. Sparse coding and max pooling was chosen as the former can mitigate quantization loss of vector quantization algorithm, which lead to better class discrimination, whereas the latter reconstructs more discriminative local features, unlike average pooling which might be affected by noisy local features [36].

Feature reduction

Since the feature data is of high dimensionality and thus can incur high computational cost, dimensionality reduction is necessary to remove irrelevant features. The main goal of feature reduction is to prevent overfitting [38], a condition where the learning models do not generalize well to unseen data (not part of the training dataset) by removing noisy and redundant features. In order to reduce the number of features, we employ feature transform methods (e.g. Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), Partial Least Square (PLS), and Independent Component Analysis (ICA)) and feature selection methods (e.g, Relieff algorithm [39], Discriminative Least Square Regression (DLSR) [40], and LW-index [41].

Classification models

Classification models attempt to derive relationship between the set of input variables (normally in the form of input vector) and its corresponding categorical target/label. In this research, the classifiers are going to deal with multi-class classification problem (identify and detect different types of brain tumor (e.g. meningioma, glioma and pituitary tumor)). The classifiers employed in this study are, Naïve-Bayes (NB) classifier, kNN classifier, weighted kNN [42], artificial neural network (ANN), support vector machines (SVM), and extreme learning machine (ELM) [43]. Table 3 presents the implementation of the classification algorithms mentioned above.

. Table 3. Implementation details of various classifiers

Classification models	Implementation details
kNN	10-fold cross validation is used to determine optimum parameter, k (number of nearest neighbors) in the range from 1 to 100. Euclidean distance measures is used.
Weighted kNN	Inversion kernel function is to calculate the weights of nearest neighbors.
NB	Gaussian probability distribution function is used to quantify the class conditional probability of each features.
ANN	Double layer hidden nodes: first one consists of 100 hidden neurons, while second one consists of 50 hidden neurons. Learning algorithm: batch gradient descent. Activation function: hyperbolic tangent.
SVM	Two SVMs: linear and RBF. RBF SVM is applied because it is capable of learning non-linear decision boundary. The sigma and C parameters in RBF SVM are optimized by grid search method[44]. Three classification schemes under one versus one SVM model will be explored: majority voting, directed acyclic graph (DAG)[45], and pairwise coupling [46]. One

	versus one SVM is chosen because of its superior performance reported in [47].
ELM	500 hidden nodes is used throughout the experiment. Optimum C parameters is found using 10-fold cross-validation method in the range of $\{2^{-24}, 2^{-23}, \ldots, 2^{25}\}$[48].

Performance evaluation and analysis

In order to evaluate the performance of each machine learning approaches applied, 7 commonly accepted performance measures are utilized, including test accuracy (%), Kappa statistic, sensitivity (%), precision (%), F-measure (%), training time (s) and test time (s). Since different schemes of feature reduction and classifiers will be experimented for 30 times using different sets of training data, the mean and standard deviation of performance metrics can be computed for unbiased analysis and comparison among methods employed. Some notations are shown in Table 4 before the performance measures together with its formula and definitions are displayed in Table 5.

Table 4. Notations in performance measures formula

Notations	Definitions
TP	The number of true positives (tumorous images correctly categorized).
FP	The number of false positives (tumorous images incorrectly categorized).
TN	The number of true negatives (normal images correctly categorized).
FN	The number of false negatives (normal images incorrectly categorized).
P(A)	The relative observed agreement among raters.
P(E)	The hypothetical probability of chance agreement.

Table 5. Performance measures and its definitions

Performance measures	Formula	Definition
Accuracy	$Acc = \frac{TP}{TP + FP + TN + FN}$	Overall efficiency and generalizability of classifier [49]. However, its use in performance comparison among classifiers is limited [50].
Kappa statistic	$Kappa = \frac{P(A) - P(E)}{1 - P(E)}$	Measure the degree of agreement between the predicted labels and the ground truth [51].
Sensitivity	$Sensitivity = \frac{TP}{TP + FN}$	Compute the proportion of samples of class 'A' label that are correctly predicted as class 'A' in the test phase.
Precision	$precision = \frac{TP}{TP + FP}$	Indicate the proportion of test samples that are predicted to be class 'A' that match the known true class label.
F-measure/F-score	$F - score = \frac{2TP}{2TP + FP + FN}$	Harmonic mean of precision and sensitivity.
Training time	-	Quantify the convergence speed of training.
Test time	-	The rate at which the output labels of test samples are generated.

RESULTS AND DISCUSSIONS

Construction of different feature sets and experiments involved

Different types of feature reduction and classification algorithms will then be applied. For feature transform method, 3, 5, 10, 20 and 50 features will be experimented. Each experiment will be conducted 30 times using stratified sampling strategy for unbiased statistical analysis. Fig. 2 shows the schematic diagram of the processed involved. In every machine learning trials, the dataset are split into 60% training data and 40% test data. In the later discussion, number of features is abbreviated as n.o.f.

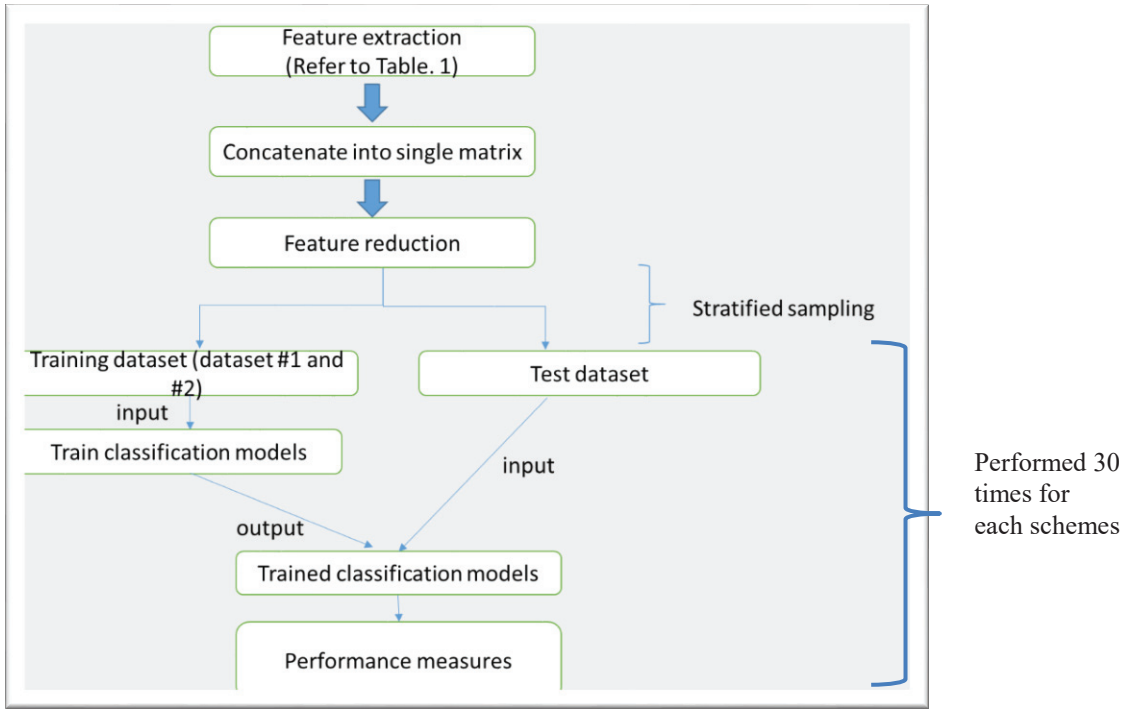

Figure 2. Flow of experiments conducted.

Performance evaluation

There are 2 types of dimensionality reduction techniques used in this study: 1) feature transform (in most literature, it is called feature extraction); 2) feature selection. Feature transform refers to PCA, PLS, LDA and ICA while feature selection denotes Relieff, DLSR and SFS-LW. We different groups of finite number of features (n.o.f) for these 2 types of dimensionality reduction: 10,20,30,40,50,60,70 and 80 features (feature transform); 50,100,150,200,250,300 features (feature selection). Thus, it is unreasonable to compare these 2 methods, but comparative study can be conducted in between methods in each dimensionality reduction techniques.

Performances of classifiers under PLS scheme

Based on Fig. 3, RBF SVM with pairwise coupling is the classifiers with the highest average accuracy (95.02%) and Kappa statistics (0.922) when n.o.f is 80. However, statistically significant difference of accuracy cannot be established in between all the RBF SVM and linear SVM (pairwise coupling) under n.o.f=80 as there is overlap in confidence interval (CI). Empirical observation on Fig. 3 also reveals that not all classification models' accuracy increases with inclusion of more features. kNN, weighted kNN and ANN are a few classifiers that achieves their highest accuracy when n.o.f=30, but performances deteriorate with inclusion of more features. This shows that these classifiers are sensitive to noises. The accuracy achieved by both linear and RBF SVM are similar, with pairwise

coupling performs slightly better than other methods. The sensitivity, precision and F-score of class A (meningioma) are generally lower than class B (glioma) and class C (pituitary tumor). This may be due to lower prior probability (number of samples) of class A during training phase.

Performances of classifiers under PCA scheme

Based on Fig. 4, RBF SVM (majority voting scheme) recorded the highest average accuracy (85.68%) and Kappa statistic (0.774) when n.o.f is 70. It should be noted here that we cannot rule out the accuracy of RBF SVM under 60 to 80 PCA features is similar due to overlap in CI. It is obviously shown in Fig. 4 that the accuracy of RBF SVM is higher than other classifiers by fairly large margin in all n.o.f settings. Interestingly, the accuracy of ANN plunges with the increase in n.o.f, which probably caused by overfitting issue. Similar to PLS, the sensitivity, precision and F-score of class A (meningioma) are generally lower than class B (glioma) and class C (pituitary tumor), and these performance measures of class A come with lower value compared to PLS. This shows that more samples of meningioma has been misclassified as glioma and pituitary tumor.

Performances of classifiers under LDA scheme

According to Fig. 5, RBF SVM with pairwise coupling achieved the highest average accuracy (87.48%) and Kappa statistic (0.803) when n.o.f is 80. There is no significant difference among linear SVM and RBF SVM under 70 and 80 LDA features. SVM is clearly better classifiers with significantly higher accuracy than other classifiers based upon empirical observation on Fig. 5. NB classifiers recorded average accuracy of 23.1% when n.o.f is 40, which in other words predicts all test samples as class A (meningioma).

Performances of classifiers under ICA scheme

According to Fig. 6, linear SVM (DAG scheme) recorded the highest accuracy and Kappa statistic of 47.07% and 0.0126 respectively. The accuracy of classification models with ICA feature as input are lower compared to the aforementioned feature transform approach. This is primarily due to the prediction of majority of test samples as class B (glioma). In other words, ICA features do not possess useful information in discriminating the tumor types.

Performances of classifiers under all three feature selection schemes

Under Relieff, DLSR, SFS-LW, the maximum accuracy is attained by RBF SVM (pairwise coupling): 89.52%, 89.63% and 86.35% respectively when n.o.f is 300. For all three feature selection methods, RBF SVM is the optimal classifiers. If we inspect bars representing RBF SVM in Fig. 7, 8 and 9, there is no statistical difference in accuracy for all three variants of one versus one RBF SVM. For these three SVM variants, the average accuracy increases with increase in n.o.f (in the case of Relieff); the average accuracy are roughly the same for all n.o.f settings (in the case of DLSR); the average accuracy are roughly the same for n.o.f in between 100 until 300 (in the case of SFS-LW). The optimal machine learning paradigms under different dimensionality reduction methods are shown in Table 6.

It is noted that kNN and weighted kNN have approximately the same accuracy under feature transform approaches. However, weighted kNN has higher accuracy compared to kNN under feature selection methods. This suggests that feature selection can better preserve distance measures, in contrast to feature transform since weighted kNN utilizes the sum of kernel function of nearest neighbors belonging to each class in making class prediction. On top of that, the accuracy of ANN decreases as n.o.f increases in all 3 cases, which suggest overfitting. Pairwise coupling is better methods in generating models outputs, particularly in linear SVM as shown by higher accuracy and Kappa statistic compared to other techniques. It is also noticed that ELM recorded comparable accuracy to linear SVM. Interestingly, NB has the worst accuracy (23.1%), in other words predict all test samples as class A (meningioma) when SFS-LW feature selection is implemented. The possible explanation is that the feature subsets constructed is

either dependent or has minimum mutual information (low entropy feature distribution) as explained in the work of [52].

In a nutshell, these are the findings that can be concluded from this research:
a) All in all, almost every machine learning paradigms investigated can achieve an average accuracy higher than 85% as long as the n.o.f is sufficient, given that the classifier used is SVM. This condition does not apply when ICA features are implemented, with accuracy lower than 48% for very combination of n.o.f and classifiers.
b) The best machine learning pipeline found was RBF SVM (pairwise coupling) + 80 PLS features, which achieves an average accuracy of 95.02%, which is the highest accuracy obtained thus far in contrast to other findings in literature.
c) Generally speaking, ANN performs poorly in all dimensionality reduction scheme. This situation becomes more evident as the n.o.f included increases. We suspect that the proposed architecture of ANN is suboptimal and cannot acquire semantic information from the features input. NB is the worst performing classifiers, which under certain settings of n.o.f and feature selection method only recorded 23.1% in accuracy, indicating that the classifier has lost its discriminating power.
d) Weighted kNN generally performs similarly as compared to kNN in majority of machine learning pipelines, with higher test time with the increase in the number of features.
e) Of the three SVM variants, there is no statistically significant difference in between them when RBF SVM is employed as there is overlap in the confidence intervals shown in error bars in the following figures. However, when linear SVM is used as classifier, pairwise coupling produces better generalization performance when PLS features and feature subsets selected by all the 3 feature selection methods are used. The test time of majority voting is lower compared to DAG and pairwise coupling. The difference is around 2 order of magnitude.
f) The training time can be arranged in ascending order as follow: NB < ELM < linear SVM < RBF SVM < ANN when n.o.f is below 200. Otherwise, the training time of linear SVM is lower than ELM.
g) The superior performance of RBF SVM under PLS scheme is due to the ability to identify meningioma correctly, as can be observed in relatively high sensitivity, precision and F-score of class A.
h) According to Table 7, our proposed machine learning approaches had achieved better accuracy than other proposed methods, showing its feasibility in classification of tumor types using T1-weighted contrast enhanced MR images.

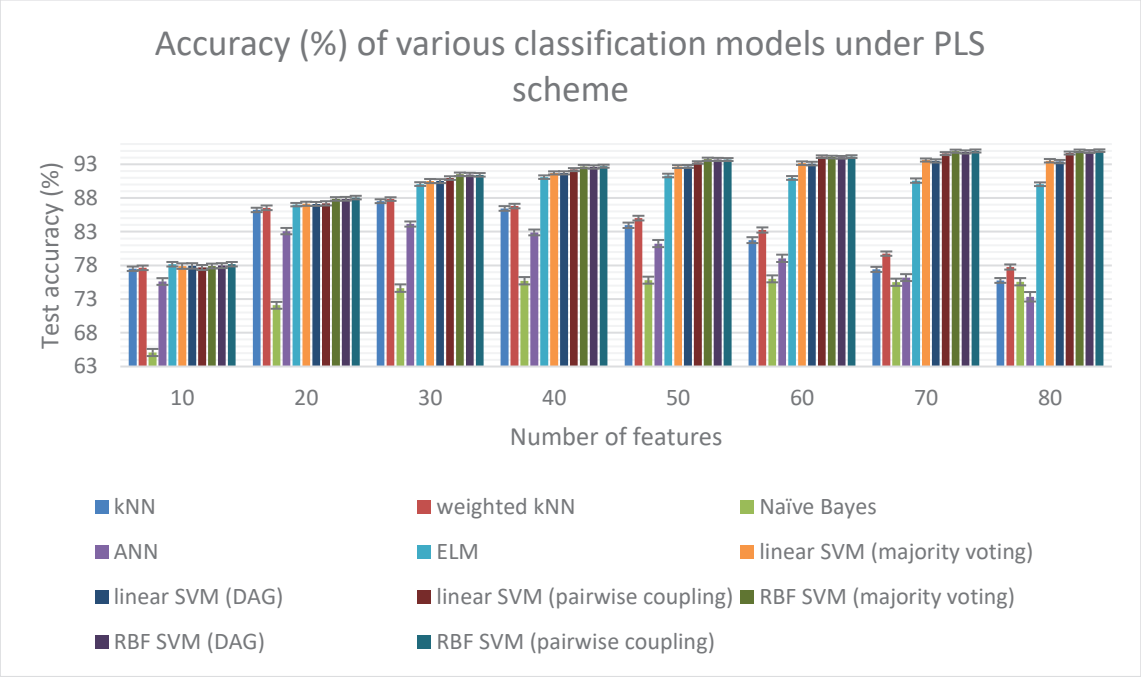

Figure 3. Classification accuracy of PLS + different classifiers. The error bars are constructed by formula: average accuracy ± (t(standard deviation))/ √n, where t represents t multiplier for student's t distribution. The value of t is 2.0452 (degree of freedom=n-1, probability=0.05). n represents the number of samples. The ranges of accuracy

represents 95% of confidence intervals. The highest accuracy is (94.83%,95.21%) attained by RBF SVM (pairwise coupling).

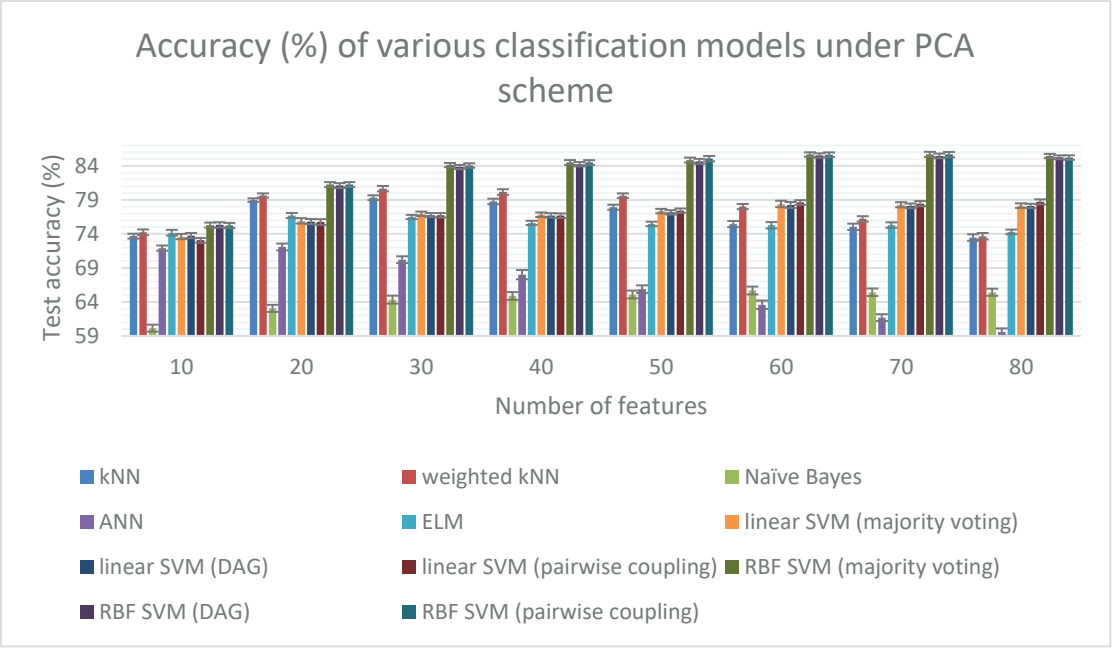

Figure 4. Classification accuracy of PCA + different classifiers. 95% confidence interval for the highest accuracy is (85.30%,86.07%) attained by RBF SVM (majority voting).

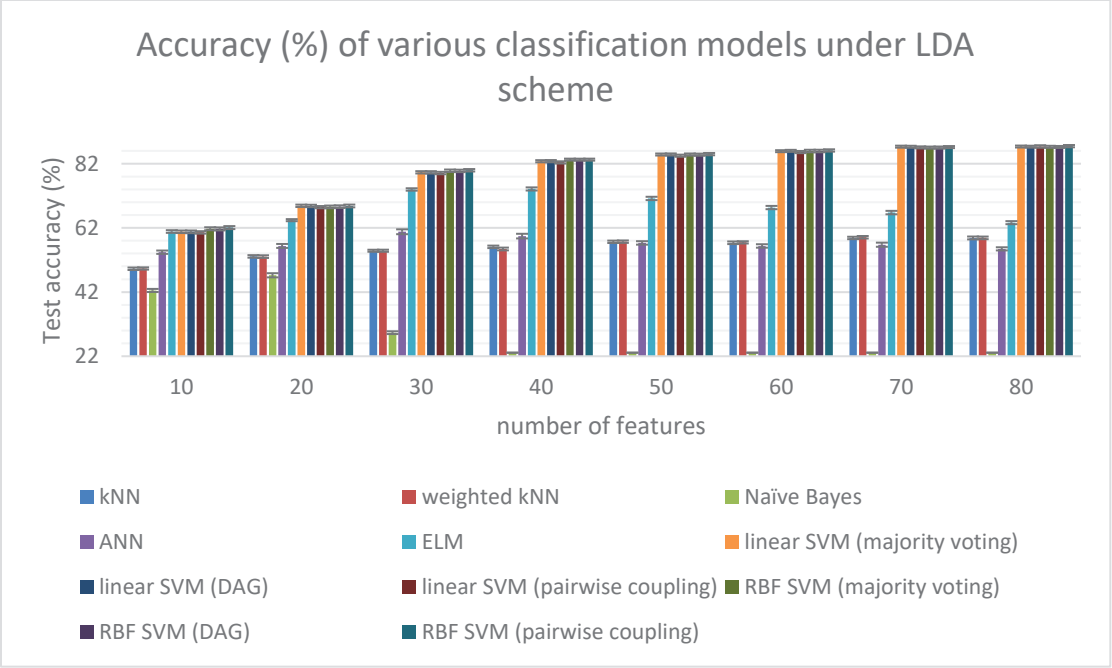

Figure 5. Classification accuracy of LDA + different classifiers. 95% confidence interval for the highest accuracy is (87.12%,87.83%) attained by RBF SVM (pairwise coupling).

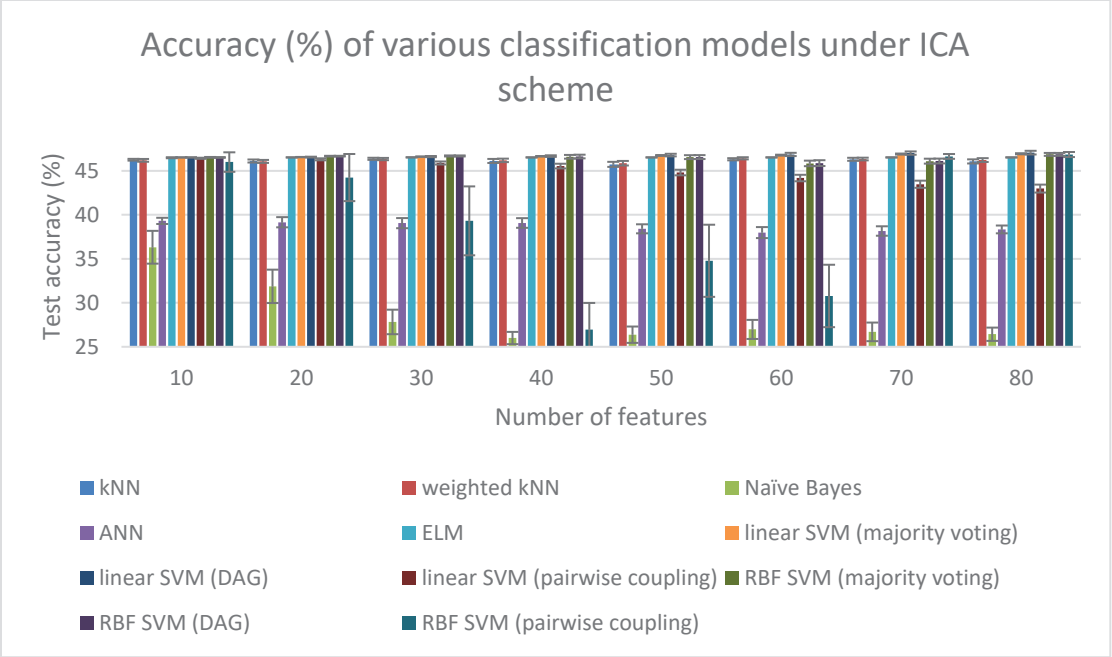

Figure 6. Classification accuracy of ICA + different classifiers. 95% confidence interval for the highest accuracy is (46.99%,47.15%) attained by linear SVM (DAG).

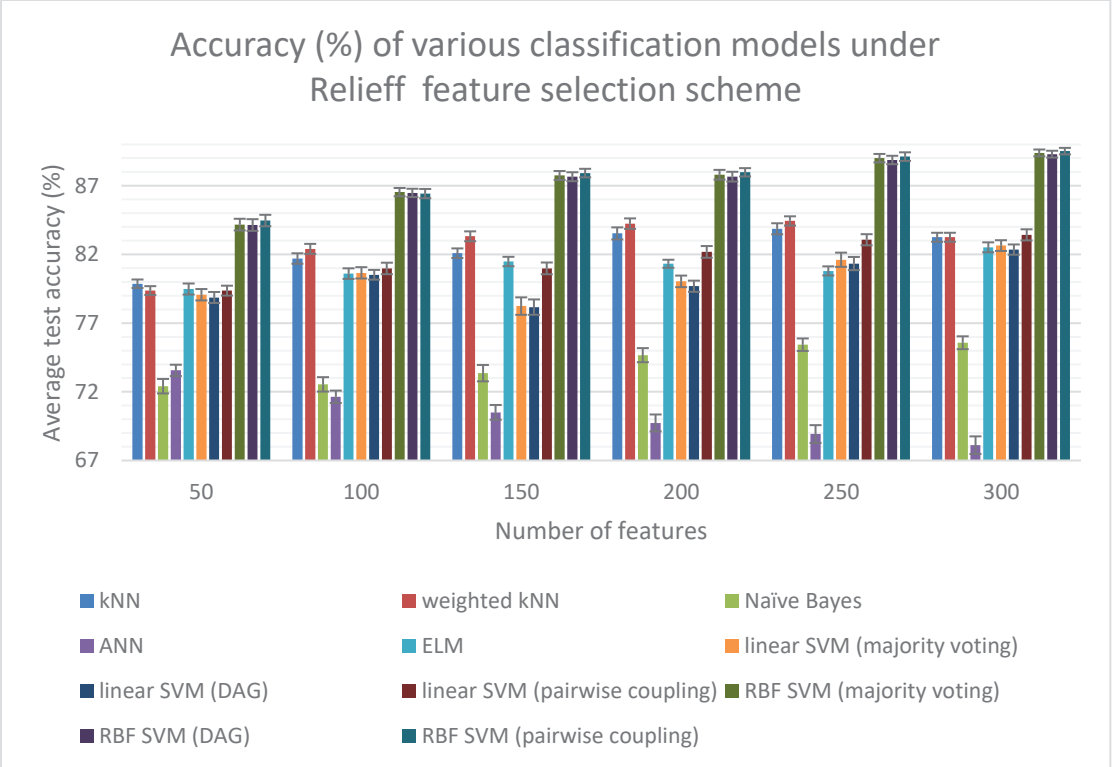

Figure 7. Classification accuracy of Relieff + different classifiers. 95% confidence interval for the highest accuracy is (89.28%,89.75%) attained by RBF SVM (pairwise coupling).

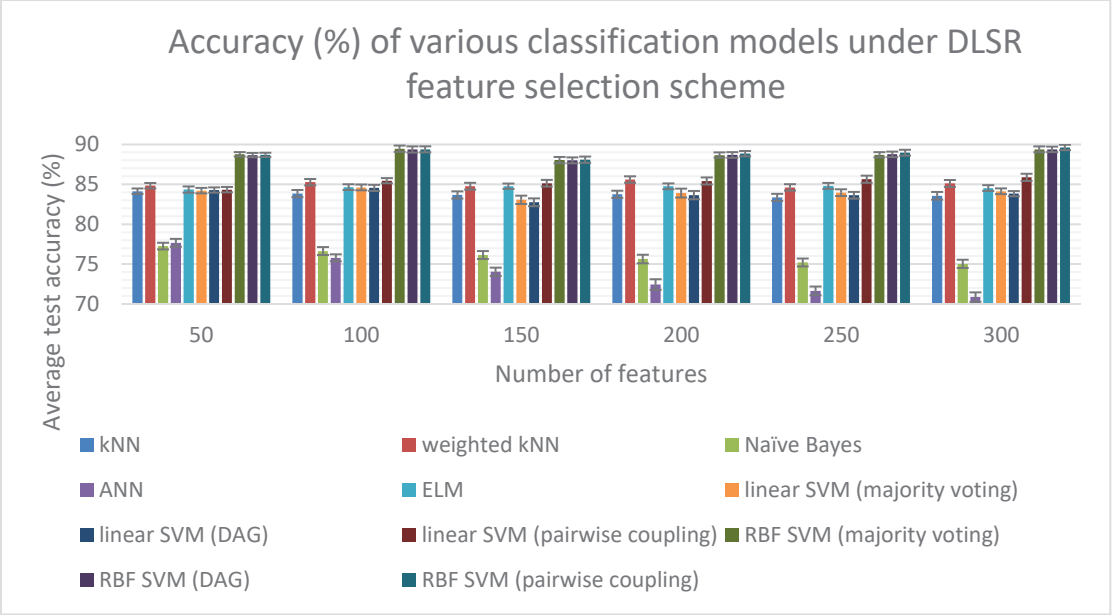

Figure 8. Classification accuracy of DLSR + different classifiers. 95% confidence interval for the highest accuracy is (88.72%,90.54%) attained by RBF SVM (pairwise coupling).

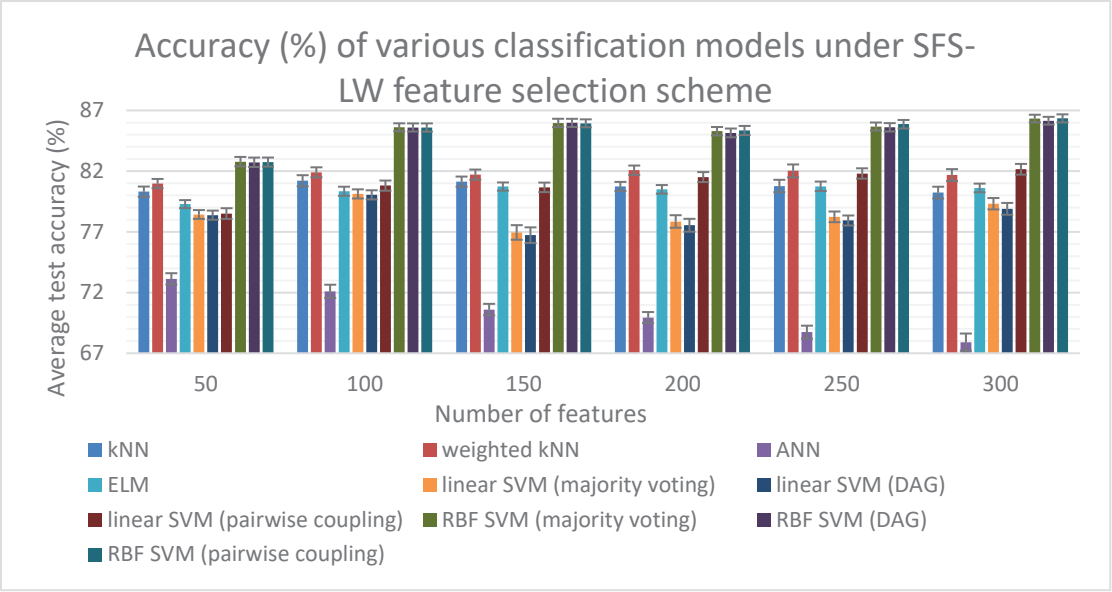

Figure 9. Classification accuracy of SFS-LW + different classifiers. 95% confidence interval for the highest accuracy is (86.02%,86.68%) attained by RBF SVM (pairwise coupling). The accuracy of NB classifier is not shown as all the accuracy is 23.01%.

Table 6. Overview of the optimal supervised classification models under each dimensionality reduction methods. The values in this table is shown in the form of: mean ± standard deviation. A denotes meningioma; B denotes glioma; C denotes pituitary tumor.

Dimensionality reduction methods	Classifiers/number of features	Performance indicators												
		Accuracy	Kappa statistics	Sensitivity			Precision			F-measure			Training time	Test time
				A	B	C	A	B	C	A	B	C		
PLS	RBF SVM (pairwise coupling) / 80	**95.02 ± 0.51**	92.18 ± 0.80	88.61 ± 1.38	97.22 ± 0.75	96.54 ± 0.73	91.92 ± 1.61	96.26 ± 0.62	95.42 ± 1.13	90.22 ± 1.09	96.74±0.4	95.97 ± 0.65	0.14±0.03	0.36±0.02
PCA	RBF SVM (majority voting) / 70	85.68 ± 1.03	77.36 ± 1.63	67.95 ± 2.97	90.98 ± 1.30	91.06 ± 1.41	79.03 ± 2.59	87.26 ± 1.14	87.53 ± 1.38	73.03 ± 2.20	89.08 ± 0.82	89.25 ± 1.00	0.25±0.01	0.04±0.00
LDA	RBF SVM (pairwise coupling) / 80	87.48 ± 0.95	80.29 ± 1.49	77.27 ± 2.52	91.12 ± 1.22	89.67 ± 1.52	80.41 ± 2.38	88.75 ± 0.93	90.76 ± 1.26	78.77 ± 1.77	89.91 ± 0.83	90.20 ± 0.96	0.77±0.05	6.94±0.19
ICA	Linear SVM (DAG) / 80	47.07 ± 0.21	1.26±0.43	0.00	99.77 ± 0.17	2.14±0.65	0.00	46.81 ± 0.10	82.53 ± 9.82	0.00	63.72 ± 0.11	4.17±1.24	0.08±0.01	0.45±0.03
DLSR	RBF SVM (pairwise coupling) / 300	89.63 ± 0.91	83.68 ± 1.43	77.34 ± 2.77	93.56 ± 1.08	92.96 ± 1.17	82.51 ± 2.48	92.08 ± 1.05	90.94 ± 1.23	79.80 ± 1.96	92.81 ± 0.80	91.93 ± 0.82	0.13±0.00	1.68±0.03
Relieff	RBF SVM (pairwise coupling) / 300	89.52 ± 0.64	83.50 ± 1.01	75.91 ± 2.49	93.75 ± 1.03	93.38 ± 1.15	82.36 ± 1.69	92.27 ± 0.93	90.29 ± 1.09	78.97 ± 1.51	92.99 ± 0.57	91.80 ± 0.68	0.15±0.00	1.81±0.03
SFS-LW	RBF SVM (pairwise coupling) / 300	86.35 ± 0.89	78.44 ± 1.43	69.00 ± 2.94	91.96 ± 1.22	90.94 ± 1.27	77.53 ± 2.17	88.74 ± 1.29	88.53 ± 1.11	72.98 ± 2.15	90.32 ± 0.91	89.71 ± 0.67	0.14±0.00	1.74±0.04

Table 7. Comparison with other findings in literature using same dataset and performing the same classification tasks.

Authors	Classifiers	Feature input	Accuracy
[53]	Capsule networks	Downsampled 64×64 raw image patches.	86.56%
[7]	SVM	BoW model with region partitioning	91.28%
Our work	RBF SVM	PLS features extracted augmented tumor region	95.02%

CONCLUSION AND FUTURE WORK

We had devised and implemented several pattern recognition paradigms under different finite n.o.f in an effort to classify multi class brain tumor using T1-weighted contrast enhanced MR images. These studies can be crucial for the development of highly demanded automated CAD system. It was found that RBF SVM (pairwise coupling) under 80 PLS features achieved the highest average accuracy ($95.02\% \pm 0.19\%$) among all other machine learning pipelines. Of all the dimensionality reduction strategies, ICA features do not offer useful information in classification of brain tumor types as the accuracy is no more than 48%. Other than that, ANN and Naïve Bayes classifiers can suffer from overfitting problem in some dimensionality reduction methods if the architecture is inappropriate (ANN) and dependence between features exist (NB).

Future work can potentially orienting towards other non-linear dimensionality reduction methods like Isomap, kernel PCA and etc. Classification algorithms like decision tree and fuzzy classification can be employed. Real online dataset (it should be noted that the data used in this research does not represent the real proportion of tumor types), which has the realistic ratio of tumor types should be considered in future study.

ACKNOWLEDGEMENTS

A special thanks to Universiti Teknologi Malaysia (UTM) for the opportunity to carry out the research and Ministry of Education (MOE) for financial support. This project was supported by Research University Grant [Vot Number: 19H03] initiated by UTM and MOE.

REFERENCES

[1] Q. T. Ostrom, H. Gittleman, G. Truitt, A. Boscia, C. Kruchko, and J. S. Barnholtz-Sloan, "CBTRUS Statistical Report: Primary Brain and Other Central Nervous System Tumors Diagnosed in the United States

in 2011-2015," *Neuro Oncol,* vol. 20, pp. iv1-iv86, Oct 1 2018.
[2] J. Lohrke, T. Frenzel, J. Endrikat, F. C. Alves, T. M. Grist, M. Law*, et al.*, "25 Years of Contrast-Enhanced MRI: Developments, Current Challenges and Future Perspectives," *Advances in therapy,* vol. 33, pp. 1-28, 2016.
[3] E. I. Zacharaki, S. Wang, S. Chawla, D. Soo Yoo, R. Wolf, E. R. Melhem*, et al.*, "Classification of brain tumor type and grade using MRI texture and shape in a machine learning scheme," *Magn Reson Med,* vol. 62, pp. 1609-18, Dec 2009.
[4] J. Sachdeva, V. Kumar, I. Gupta, N. Khandelwal, and C. K. Ahuja, "Segmentation, Feature Extraction, and Multiclass Brain Tumor Classification," *Journal of Digital Imaging,* vol. 26, pp. 1141-1150, 05/04 2013.
[5] K. M. Priya, S. Kavitha, and B. Bharathi, "Brain tumor types and grades classification based on statistical feature set using support vector machine," in *2016 10th International Conference on Intelligent Systems and Control (ISCO)*, 2016, pp. 1-8.
[6] A. Jayachandran and R. Dhanasekaran, "MULTI CLASS BRAIN TUMOR CLASSIFICATION OF MRI IMAGES USING HYBRID STRUCTURE DESCRIPTOR AND FUZZY LOGIC BASED RBF KERNEL SVM," *Iranian Journal of Fuzzy Systems,* vol. 14, pp. 41-54, 2017.
[7] J. Cheng, W. Huang, S. Cao, R. Yang, W. Yang, Z. Yun*, et al.*, "Enhanced Performance of Brain Tumor Classification via Tumor Region Augmentation and Partition," *PLoS One,* vol. 10, p. e0140381, 2015.
[8] J. Cheng, W. Yang, M. Huang, W. Huang, J. Jiang, Y. Zhou*, et al.*, "Retrieval of Brain Tumors by Adaptive Spatial Pooling and Fisher Vector Representation," *PLoS One,* vol. 11, p. e0157112, 2016.
[9] W. Yang, Q. Feng, M. Yu, Z. Lu, Y. Gao, Y. Xu*, et al.*, "Content-based retrieval of brain tumor in contrast-enhanced MRI images using tumor margin information and learned distance metric," *Med Phys,* vol. 39, pp. 6929-42, Nov 2012.
[10] S. A. Medjahed, *A Comparative Study of Feature Extraction Methods in Images Classification* vol. 7, 2015.
[11] I. Guyon, Andr, and Elisseeff, "An introduction to variable and feature selection," *J. Mach. Learn. Res.,* vol. 3, pp. 1157-1182, 2003.
[12] L. Keyes and A. Winstanley, "Using moment invariants for classifying shapes on large-scale maps," *Computers, Environment and Urban Systems,* vol. 25, pp. 119-130, 2001/01/01/ 2001.
[13] H. Zhihu and L. Jinsong, "Analysis of Hu's moment invariants on image scaling and rotation," in *2010 2nd International Conference on Computer Engineering and Technology*, 2010, pp. V7-476-V7-480.
[14] A. Khotanzad and Y. H. Hong, "Invariant image recognition by Zernike moments," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 12, pp. 489-497, 1990.
[15] D. v. P. P. Bhaskara Rao, Ch.pavan Kumar, "Feature extraction using zernike moments," *International Journal of Latest Trends in Engineering and Technology* vol. 2, March 2013 2013.
[16] C. Singh and R. Upneja, "Accurate calculation of high order pseudo-Zernike moments and their numerical stability," *Digital Signal Processing,* vol. 27, pp. 95-106, 2014/04/01/ 2014.
[17] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, 2005, pp. 886-893 vol. 1.
[18] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 24, pp. 971-987, 2002.
[19] D. Cai, H. Bao, and X. He, "Sparse concept coding for visual analysis," in *CVPR 2011*, 2011, pp. 2905-2910.
[20] D. Tian, *A review on image feature extraction and representation techniques* vol. 8, 2013.
[21] M. Yang, K. Kpalma, and J. Ronsin, *A Survey of Shape Feature Extraction Techniques* vol. 15, 2007.
[22] H. Ming-Kuei, "Visual pattern recognition by moment invariants," *IRE Transactions on Information Theory,* vol. 8, pp. 179-187, 1962.
[23] C. Teh and R. T. Chin, "On image analysis by the methods of moments," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 10, pp. 496-513, 1988.
[24] M. R. Teague, "Image analysis via the general theory of moments*," *Journal of the Optical Society of America,* vol. 70, pp. 920-930, 1980/08/01 1980.
[25] M. A Vorobyov, "Shape Classification Using Zernike Moments," 2011.
[26] W.-Y. Kim and Y.-S. Kim, "A region-based shape descriptor using Zernike moments," *Signal Processing: Image Communication,* vol. 16, pp. 95-102, 2000/09/01/ 2000.
[27] C. Singh, E. Walia, and R. Upneja, "Accurate calculation of Zernike moments," *Information Sciences,* vol. 233, pp. 255-275, 2013/06/01/ 2013.
[28] C. Singh, E. Walia, Pooja, and R. Upneja, "Analysis of algorithms for fast computation of pseudo Zernike

moments and their numerical stability," *Digital Signal Processing,* vol. 22, pp. 1031-1043, 2012/12/01/ 2012.
[29] M. Annalakshmi, S. M. M. Roomi, and A. S. Naveedh, "A hybrid technique for gender classification with SLBP and HOG features," *Cluster Computing,* 2018/01/30 2018.
[30] J. J. DiCarlo, D. Zoccolan, and N. C. Rust, "How does the brain solve visual object recognition?," *Neuron,* vol. 73, pp. 415-34, Feb 9 2012.
[31] H. Bristow and S. Lucey, *Why do linear SVMs trained on HOG features perform so well?*, 2014.
[32] Z. Guo, L. Zhang, and D. Zhang, "A Completed Modeling of Local Binary Pattern Operator for Texture Classification," *IEEE Transactions on Image Processing,* vol. 19, pp. 1657-1663, 2010.
[33] F. Lu and J. Huang, "An improved local binary pattern operator for texture classification," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2016, pp. 1308-1311.
[34] A. Bosch, X. Muñoz, and R. Martí, "Which is the best way to organize/classify images by content?," *Image and Vision Computing,* vol. 25, pp. 778-791, 2007/06/01/ 2007.
[35] M. T. Law, N. Thome, and M. Cord, "Bag-of-Words Image Representation: Key Ideas and Further Insight," in *Fusion in Computer Vision: Understanding Complex Visual Content*, B. Ionescu, J. Benois-Pineau, T. Piatrik, and G. Quénot, Eds., ed Cham: Springer International Publishing, 2014, pp. 29-52.
[36] C. Wang and K. Huang, "How to use Bag-of-Words model better for image classification," *Image and Vision Computing,* vol. 38, pp. 65-74, 2015/06/01/ 2015.
[37] Y. Boureau, F. Bach, Y. LeCun, and J. Ponce, "Learning mid-level features for recognition," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2010, pp. 2559-2566.
[38] J. Tang, S. Alelyani, and H. Liu, *Feature selection for classification: A review*, 2014.
[39] I. Kononenko, M. Robnik-Sikonja, and S. Uros Pompe, *ReliefF for estimation and discretization of attributes in classification, regression, and ILP problems*, 2000.
[40] S. Xiang, F. Nie, G. Meng, C. Pan, and C. Zhang, "Discriminative Least Squares Regression for Multiclass Classification and Feature Selection," *IEEE Transactions on Neural Networks and Learning Systems,* vol. 23, pp. 1738-1754, 2012.
[41] C. Liu, W. Wang, Q. Zhao, X. Shen, and M. Konan, "A new feature selection method based on a validity index of feature subset," *Pattern Recognition Letters,* vol. 92, pp. 1-8, 2017/06/01/ 2017.
[42] K. Hechenbichler and K. Schliep, *Weighted k-Nearest-Neighbor Techniques and Ordinal Classification* vol. 399, 2004.
[43] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing,* vol. 70, pp. 489-501, 2006/12/01/ 2006.
[44] C. W. Hsu, C. C. Chang, and C. J. Lin, *A Practical Guide to Support Vector Classification* vol. 101, 2003.
[45] P. Chen and S. Liu, "An Improved DAG-SVM for Multi-class Classification," in *2009 Fifth International Conference on Natural Computation*, 2009, pp. 460-462.
[46] T. Hastie and R. Tibshirani, "Classification by pairwise coupling," *Ann. Statist.,* vol. 26, pp. 451-471, 1998/04 1998.
[47] H. Chih-Wei and L. Chih-Jen, "A comparison of methods for multiclass support vector machines," *IEEE Transactions on Neural Networks,* vol. 13, pp. 415-425, 2002.
[48] G.-B. Huang, X. Ding, and H. Zhou, "Optimization method based extreme learning machine for classification," *Neurocomputing,* vol. 74, pp. 155-163, 2010/12/01/ 2010.
[49] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Information Processing & Management,* vol. 45, pp. 427-437, 2009/07/01/ 2009.
[50] F. J. Provost, T. Fawcett, and R. Kohavi, "The Case against Accuracy Estimation for Comparing Induction Algorithms," presented at the Proceedings of the Fifteenth International Conference on Machine Learning, 1998.
[51] Z. Zhang, "Introduction to machine learning: k-nearest neighbors," *Annals of translational medicine,* vol. 4, pp. 218-218, 2016.
[52] I. Rish, *An Empirical Study of the Naïve Bayes Classifier* vol. 3, 2001.
[53] P. Afshar, A. Mohammadi, and K. Plataniotis, *Brain Tumor Type Classification via Capsule Networks*, 2018.