

Article

# Multi-Class Disease Classification in Brain MRIs Using a Computer-Aided Diagnostic System

Muhammad Faisal Siddiqui <sup>1,2</sup>, Ghulam Mujtaba <sup>3,4</sup>, Ahmed Wasif Reza <sup>5,\*</sup> and Liyana Shuib <sup>3,\*</sup>

<sup>1</sup> Faculty of Engineering, Department of Electrical Engineering, University of Malaya, Kuala Lumpur 50603, Malaysia; mfaisal\_bce@yahoo.com

<sup>2</sup> Department of Electrical Engineering, Faculty of Engineering, COMSATS Institute of Information Technology, Islamabad 45550, Pakistan

<sup>3</sup> Department of Information Systems, Faculty of Computer Science and Information Technology, University of Malaya, Kuala Lumpur 50603, Malaysia; mujtaba@iba-suk.edu.pk

<sup>4</sup> Department of Computer Science, Sukkur Institute of Business Administration, Sukkur 65200, Pakistan

<sup>5</sup> Department of Computer Science and Engineering, Faculty of Science and Engineering, East West University, Dhaka 1212, Bangladesh

\* Correspondence: awreza98@yahoo.com (A.W.R.); liyanashuib@um.edu.my (L.S.); Tel.: +880-2-09666775577-395 (A.W.R.)

Academic Editor: Angel Garrido

Received: 29 November 2016; Accepted: 1 March 2017; Published: 8 March 2017

**Abstract:** Background: An accurate and automatic computer-aided multi-class decision support system to classify the magnetic resonance imaging (MRI) scans of the human brain as normal, Alzheimer, AIDS, cerebral calcinosis, glioma, or metastatic, which helps the radiologists to diagnose the disease in brain MRIs is created. Methods: The performance of the proposed system is validated by using benchmark MRI datasets (OASIS and Harvard) of 310 patients. Master features of the images are extracted using a fast discrete wavelet transform (DWT), then these discriminative features are further analysed by principal component analysis (PCA). Different subset sizes of principal feature vectors are provided to five different decision models. The classification models include the J48 decision tree, k-nearest neighbour (kNN), random forest (RF), and least-squares support vector machine (LS-SVM) with polynomial and radial basis kernels. Results: The RF-based classifier outperformed among all compared decision models and achieved an average accuracy of 96% with 4% standard deviation, and an area under the receiver operating characteristic (ROC) curve of 99%. LS-SVM (RBF) also shows promising results (i.e., 89% accuracy) when the least number of principal features was used. Furthermore, the performance of each classifier on different subset sizes of principal features was (80%–96%) for most performance metrics. Conclusion: The presented medical decision support system demonstrates the potential proof for accurate multi-class classification of brain abnormalities; therefore, it has a potential to use as a diagnostic tool for the medical practitioners.

**Keywords:** computer aided diagnostic system; neuroimaging; brain magnetic resonance imaging (MRI); multi-classification; medical imaging

---

## 1. Introduction

In this modern era, different advanced imaging modalities (e.g., X-rays, computerized tomography (CT) scans, positron emission tomography (PET), single-photon emission computerized tomography (SPECT), and magnetic resonance imaging (MRI)) are used in neurology and basic neuroscience fields. In X-rays and CT scans, the patients are exposed in ionizing radiation waves which may increase the risk of developing cancers, whereas PET and SPECT use radioactive tracers, with minimal exposure to harmful radiation. However, MRI is a non-invasive, dominant, and flexible modality to investigate

the pathological conditions of the brain and other body parts. The common practice to identify the brain abnormalities is done by MRI. MRI scans provide high contrast and high spatial resolution images, which enables to differentiate the characteristics of the soft tissues. Magnetic resonance (MR) image texture is used to distinguish between the healthy and diseased anatomy. Brain MR images having any disease (such as Alzheimer, AIDS dementia, cerebral calcinosis, glioma, or metastasis) are categorized by large cells and high contrast, which can be identified by abrupt changes in the images [1,2]. In recent years, machine learning techniques have been widely employed in the medical domain to support decision-making [3–10]. Moreover, medical decision support systems are in high demand to automatically detect these abrupt changes properly and classify the brain MRI as normal or any class of disease [3,5]. The very large amount of MR imaging data generates complexity to interpret the full pattern of atrophy by the existing visual inspection method. Therefore, it generates the requirement of computer-aided diagnosis (CAD) system to identify the specific condition of the brain MRI and enhance the diagnostic capabilities of the medical personnel. The radiologists can use these automated systems as an instrument for diagnosis, pre-surgical, and post-surgical procedures [4,11–22].

Generally, supervised classification methods are used instead of unsupervised methods, because of better accuracy, for brain MRI classification. There are, commonly, three phases involved in the implementation of a classifier for medical images: (1) features extraction; (2) feature reduction; and (3) training/testing of classification models. In the first step, discriminative features are extracted from brain MR images. Then, in the second step, these features are processed by some feature reduction technique to reduce the dimensionality of the features, such as PCA, linear discriminant analysis (LDA), etc. Finally, the reduced principal features are used to train the classifier model and classify the query images on the bases of these features. Widely-used decision models include naïve Bayes (NB), J48, random forest (RF), k-nearest neighbor (kNN), and support vector machine (SVM).

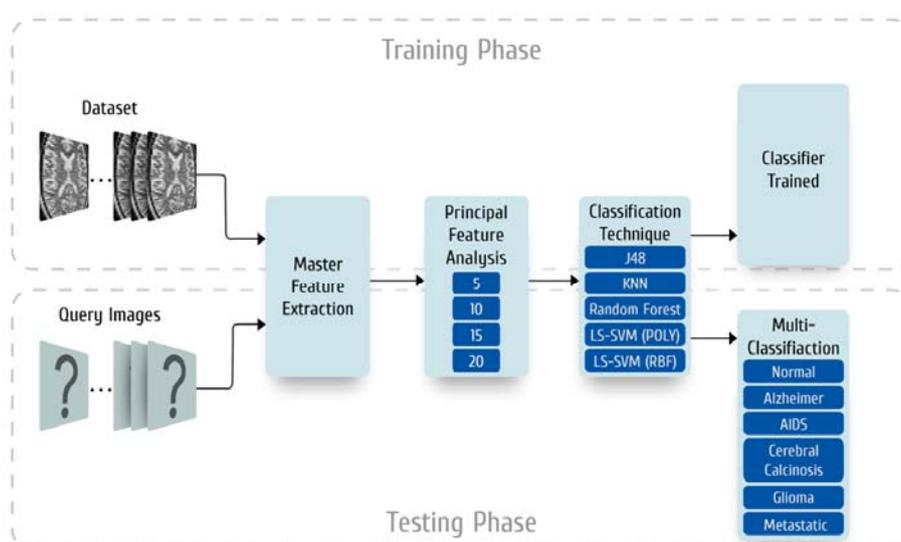
Recently, various feature selection schemes and machine learning decision models for brain MRI classification have been proposed. In [6,14–19], the authors have used 2D-DWT (two-dimensional discrete wavelet transform) and principal component analysis (PCA) for a features extraction and selection, respectively. Zhang et al. [14–17] proposed different advanced decision models based solutions with DWT and PCA techniques in their research and achieved promising results with some limitations. They used forward neural networks (FNN) with scaled chaotic artificial bee colony (SCABC) [14], back propagation with conjugate gradient method [15], kernel support vector machines (KSVM) with Gaussian radial basis function (GRB) [16] and particle swarm optimization (PSO) [17] as decision models to improve the efficiency of the brain MRI classifier. The schemes proposed in [18,19], have used feed-forward back propagation artificial neural networks (FP-ANN), kNN, feedback pulse-coupled neural network (FBPNN) and achieved an average accuracy of 99% for binary classification of brain MR images. Recently, [20–22] have proposed numerous complex feature engineering techniques with SVM- and NB-based decision models to enhance the classifier performance. In [20,21], the authors proposed Ripplet transform and discrete wavelet packet transform, respectively, instead of DWT for feature extraction. Whereas, the authors proposed wavelet entropy method in [22] for feature reduction. Wang et al. [23] proposed a different classification scheme by using dual-tree complex wavelet transform and twin support vector machine for the classification of pathological brain detection. In [23], the authors have achieved average accuracy of 99.57%. However, usage of a small number of cases and limited number of disease categories in their datasets are the main limitation of these works and also their performance is significantly reduced when large datasets are used. Regardless, the advanced complex methods are used in [20–23], which increases the complexity of the classifier and consumed relatively higher computational time, are not able to perform well on large datasets. Conversely, the technique proposed in [6] achieve accurate results for larger datasets; when using DWT, PCA, and LS-SVM (RBF) for feature extraction, feature reduction and classification, respectively. In addition, all these schemes [6,14–23] have proposed for binary classification and only capable to predict normal and abnormal anatomy of a brain MRI. Furthermore, a multi-class brain MRI classifier has proposed by Zacharaki et al. [24]. The authors extended the scope of the features

and included age, tumour shape, and ROI (region of interest) as a part of feature sets. This technique is semiautomatic because ROIs needs to trace manually. In [24], the authors have tested three different disease classes (i.e., meningioma, glioma, and metastasis) and achieved a mean accuracy about 90%. The comparison between linear discriminant analysis (LDA), kNN, and non-linear SVM-based decision models were shown. The main limitation of this technique is that it needs human intervention for classification. Therefore, the research gap in the development of fully automatic multi-class classifier with significant accuracy is generated.

The main motivation behind this study is to develop an accurate multi-class brain MRI classifier, which is capable to diagnose the diseases class in brain MRIs. The proposed multi-class brain MRI classifier has a potential to classify five different brain diseases. These brain diseases include Alzheimer, AIDS dementia, cerebral calcinosis, glioma, and metastatsis. The proposed system technique is composed of three sub-models; master feature extraction, principal feature analysis and decision models. Fast discrete wavelet transform (DWT) used for extracting the master features from the brain MR images. The principal feature analysis was done by PCA and different subsets were used to calculate the efficiency of the multi-class classifier system. In addition, PCA analysis reduced the dimension of the master features, which also decreases the classification time and complexity of the classifier. For a comprehensive comparison of decision models' performance on the multi-classification of brain MRIs, the proposed research compared five different decision models (J48, kNN, RF, and LS-SVM with polynomial (Poly) and radial basis functions (RBF)). For comparative analysis with the proposed system, some of the other published methods from recent literature [6,18,20,21] were also tested using the same large datasets.

## 2. Materials and Methods

The proposed classifier for multi-classification is composed of master feature extraction, principal feature analysis, and classification model blocks, as shown in Figure 1, which illustrates the methodology of the proposed system. The classifier is constructed and evaluated using two phases: (1) a training phase, and (2) a testing phase. In the training phase, the classifier is trained by randomly selected images from the datasets. Once the classifier is trained, then it is capable to classify the query images. In the testing phase, the query image(s) is/are fed to the trained classifier to classify the image(s) as normal, Alzheimer, AIDS, cerebral calcinosis, metastatic, or glioma. Furthermore, a five-fold cross-validation is used in this work to minimize the generalization error.

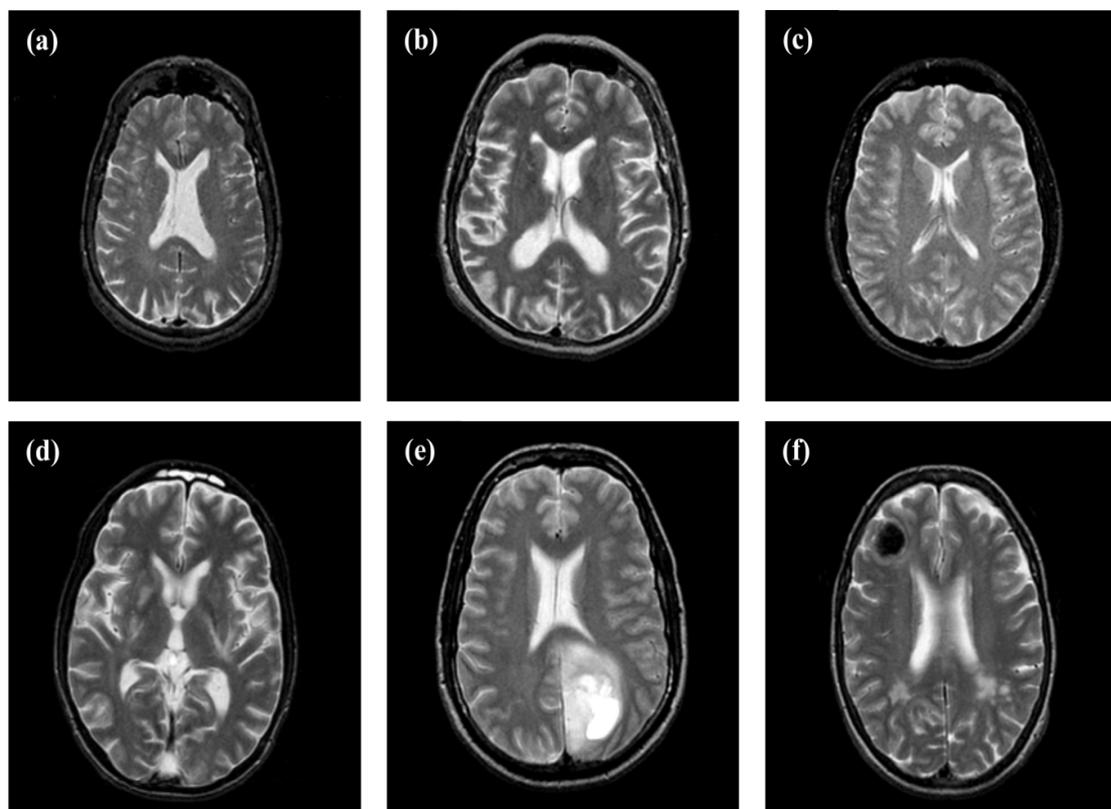


**Figure 1.** Methodology of the proposed classifier.

### 2.1. Dataset Collection

The benchmark MRI dataset used in this research was collected from ‘Open Access Series of Imaging Studies (OASIS)’ and ‘Harvard Medical School’ MRI databases to validate the proposed classification system. This database consists of human brain MRI images in the axial plane. These datasets were acquired using the following scan parameters: Voxel res:  $1.0 \times 1.0 \times 1.25$  (mm<sup>3</sup>), Rect. FOV: 256/256, TR: 9.7 (ms), TE: 4.0 (ms), TI: 20.0 (ms), and flip angle: 10°. The dimensions of the image are  $256 \times 256$  in a plane-resolution. Three hundred and ten patients’ (men and women) brain MRI scans were involved to formulate this database.

The brain MR image dataset is composed of healthy and abnormal images. The abnormal image database has five types of different brain diseases. The abnormal MRI scan images having the following diseases: Alzheimer’s disease, AIDS dementia, cerebral calcinosis, glioma and metastatic dementia. A sample image of each class of the images included in the benchmark dataset is shown in Figure 2.



**Figure 2.** The sample images of healthy and abnormal magnetic resonance imaging (MRI) (a) normal/healthy; (b) Alzheimer’s disease; (c) AIDS dementia; (d) cerebral calcinosis; (e) glioma; and (f) metastatic dementia.

The dataset is comprised of 310 brain MR images having 70 healthy (normal), 70 Alzheimer, 50 AIDS, and 40 each for cerebral calcinosis, glioma, and metastasis. The distribution of training and testing images is shown in Table 1. The ratio of training and testing images, i.e., 70% of the dataset is used for training and the remaining 30% of the dataset is used for testing purposes. Training images were used to construct the classifier, whereas testing images were used to evaluate the performance of the multi-class classifier. In addition, the testing images were unknown to the constructed classifier for the sake of unbiased evaluation.

**Table 1.** The distribution of training and testing images.

Class	Total No. of Images	Total No. of Training Images	Total No. of Testing Images	Distribution (%)
Normal	70	49	21	22.58
Alzheimer	70	49	21	22.58
Aids	50	35	15	16.13
Cerebral Calcinosi	40	28	12	12.90
Glioma	40	28	12	12.90
Metastasis	40	28	12	12.90

## 2.2. Master Feature Extraction

A MATLAB (R2013a, The Mathworks, Inc. Natick, MA, United States) script was written, using discrete wavelet transform, to extract the main features of the brain MR images. To improve the efficiency of a classifier, the master features in the MRI image is needed to be identified properly. In recent literature [6,14–22,24–27], there are many different algorithms (such as DWT and Ripplet transform) used to extract the main features of the images. DWT has some advantages over RT, being less computationally complex and also due to the characteristics of brain MRIs. The sparse nature of MRIs provides an opportunity to identify the major contributed features of the MR image by representing it in some sophisticated domains (such as wavelet domains) [28]. Thus, DWT provides master features, having rich knowledge of the input MR image pattern with less complex implementation. The master features extracted from the MRI database using DWT has a potential to increase the capability of the decision making power and complexity of the classifier. A three-level “Haar” DWT was used to extract the master features of the images in this paper.

## 2.3. Preparation of the Principal Feature Vector

The main characteristics of any robust and accurate classifier are a selection of the discriminative features from the dataset and reduce the dimensions of the dataset. Large databases increase the feature dimensions, which eventually increase the complexity of the classifier and demands excessive time to classify. Therefore, different feature reduction schemes are used by the researchers to remove the curse of dimensionality problems in the classifier system [19,29–31].

In this article, the PCA technique was applied on the discriminative features of MR image to further reduce the dimension of the master features extracted by DWT. PCA preserved the variance by extracting the linear lower-dimensional representation of the MR image features [32,33]. Therefore, it extracts the major components of the image (data) and forms the principal feature vector. This leads to an increase the efficiency of the classifier system.

### 2.3.1. Feature Subset Sizes

Principal feature vectors are used for decision modelling. However, subsets of the principal feature vectors were introduced to check the performance trend of the classifier. Subset sizes of 5, 10, 15, and 20 principal components were used to compute the results of the proposed multi-class classifier.

## 2.4. Classifier Models

Weka toolkit (Version 3.8, University of Waikato, Hamilton, New Zealand) was used to develop decision models. Principal feature vectors were exported from MATLAB in comma separated values (CSV) format. Afterwards the CSV data were loaded into Weka for further analysis. Five different decision models were constructed for performance measures. These five classifiers were J48, k-nearest neighbour (kNN), random forest (RF), least-square support vector machine with polynomial kernel (LS-SVM (Poly)), and least-squares support vector machine with radial basis function kernel (LS-SVM (RBF)).

#### 2.4.1. J48 Classifier (J48)

J48 is a kind of decision tree algorithm [34]. J48 utilizes the entropy to compute the homogeneity of a sample. If entropy is zero, then it means that the sample is completely homogeneous and if the sample is unequally divided, then it has entropy of one. The relative entropy of a given dataset  $X$  having positive and negative class instances is mathematically defined as:

$$E(X) = -p(P) \log_2 p(P) - p(N) \log_2 p(N) \quad (1)$$

where  $p(P)$  and  $p(N)$  are related probabilities of positive and negative class, respectively.

#### 2.4.2. K-Nearest Neighbor (kNN)

kNN is also known as a lazy learning non-parametric algorithm. kNN is the simplest classification algorithm that stores all training instances and uses a Euclidean distance function to classify new instances (shown in Equation (2)) [35,36]:

$$\sqrt{\sum_{i=1}^k (x_i - y_i)^2} \quad (2)$$

where  $x$  and  $y$  are two vectors (trained instance vector and a query vector for classification), and  $k$  represents the number of attributes.

#### 2.4.3. Random Forest (RF)

RF was proposed by UC Berkeley visionary Leo Breiman in 1999 [37]. This algorithm works as a large collection of decorrelated decision trees using a bagging technique. The RF creates various sub-training sets from a super training set. A decision tree classifier is constructed from each sub-training set. At the time of testing, each input vector of the test set is classified by all of the decision trees in a forest and, finally, the forest is responsible for choosing the classification results; using either majority votes or averaging the predictions using the equation given below [38]:

$$f = \frac{1}{B} \sum_{b=1}^B f_b(x) \quad (3)$$

where  $B$  represents the samples/trees,  $f_b$  is a predictor, and  $x$  corresponds to the test point.

#### 2.4.4. Least Squares-Support Vector Machine (LS-SVM)

The SVM classifier is highly influenced by advances in statistical learning theory [39–41]. SVM plays a vital role in the application of object detection [42], face detection [43], handwriting recognition [44], medical imaging classification [6], and bioinformatics [45]. SVM learns from training examples. An improved version of SVM, i.e., LS-SVM, was used in this article because of its robustness and efficiency. Each training instance consists of  $n$  number of attributes  $(x_1, x_2, \dots, x_n)$  with a corresponding class label. The nonlinear function estimation can be mathematically presented as:

$$y = \text{sign} [ W \varphi(x) + b ] \quad (4)$$

where the high dimensional feature space is represented by  $\varphi(x)$ , the weight vector is defined by  $W$ , and the bias term is denoted by  $b$ . Then, the LS-SVM solution of such an optimization problem can be obtained as follows (for a deeper introduction of this method, readers can refer to [46–48]):

$$y(x) = \text{sign} \left[ \sum_{i=1}^N \alpha_i y_i K(x, x_i) + b \right] \quad (5)$$

Table 2 provides some of the choices of kernel functions  $K(x_k, x_l)$ .

**Table 2.** Least squares-support vector machine (LS-SVM) kernel functions.

Kernel	Expression
Linear	$K(x, y) = x^T y$
Polynomial	$K(x, y) = \left(1 + \frac{x^T y}{\sigma^2}\right)^d$
RBF	$K(x, y) = \exp\left\{-\frac{\ x - y\ ^2}{\sigma^2}\right\}$

### 2.5. Performance Measures

Recall (sensitivity), precision, F-measure, accuracy, and area under the receiver operating characteristic (ROC) curve are widely used metrics to determine the performance of the classifiers [49]. The possible outcomes of the proposed classifier can be described as:

*TP* (True Positive): Number of images correctly diagnosed under any specific class;

*TN* (True Negative): Number of images correctly rejected by the classifier;

*FP* (False Positive): Number of images incorrectly identified by the classifier;

*FN* (False Negative): Number of images incorrectly discarded by the classifier.

For multi-class classification, macro-averaged recall, macro-averaged precision and macro-averaged F-measure are used to validate the performance of the classifier [49].

*Recall<sub>M</sub>* is the average of the each class recall (i.e., the probability of the test finding the positive cases among all the positive cases of the respective class):

$$Recall_M = \frac{\sum_{i=1}^C \frac{TP_i}{TP_i + FN_i}}{C} \quad (6)$$

*Precision<sub>M</sub>* is the average of the each class precision (i.e., the probability of the test correctly diagnosed as positive cases given that the number of cases labelled by the system as positive):

$$Precision_M = \frac{\sum_{i=1}^C \frac{TP_i}{TP_i + FP_i}}{C} \quad (7)$$

*F-Measure<sub>M</sub>* (macro-averaged F-measure) is a weighted combination of the *Recall<sub>M</sub>* and *Precision<sub>M</sub>*. Mathematically, it is defined as:

$$F-Measure_M = \frac{(\beta^2 + 1) Recall_M \times Precision_M}{\beta^2 Precision_M + Recall_M} \quad (8)$$

**Average Accuracy** is the fraction of test results predicted as correct among all the classes:

$$Accuracy_{Avg} = \frac{\sum_{i=1}^C \frac{TP_i + TN_i}{TP_i + FN_i + TN_i + FP_i}}{C} \quad (9)$$

**Area under the ROC curve (AUC)** is the area occupied by the receiver operating characteristic curve of each class. It is used to analyse how good any classification model predicts the specific class versus all other classes:

$$AUC = \frac{1}{2} \left( \frac{TP}{TP + FN} + \frac{TN}{TN + FP} \right) \quad (10)$$

where  $C$  represents the total number of classes. i.e.,  $C = 6$ .  $M$  index represents to macro-averaging.  $\beta = 1$  was used in this research.

### 2.6. Experimental Setup

Separate experiments were conducted on training and testing datasets. PCA with DWT was applied to extract the discriminative principal feature vectors with four different subset sizes (5, 10, 15, and 20). Four plus four (total of eight) principal feature vector sets were extracted from training datasets and testing datasets, respectively. To evaluate the performance of the proposed multi-class classifier using performance metrics; a total of 40 ( $8 \times 5$ ) analyses were performed by applying each of these 8 feature sets to five different classifier models (J48, kNN, RF, LS-SVM (Poly), and LS-SVM (RBF)).

## 3. Results and Discussion

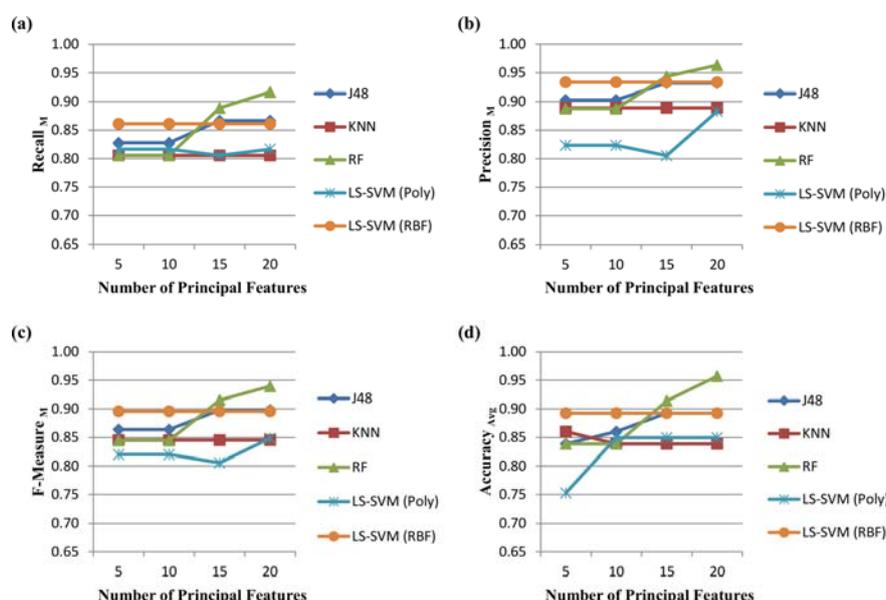
### 3.1. Feature Reduction

In order to extract discriminative and reduced features, fast DWT with PCA was used in this research. The fast DWT only computes the approximation component of the wavelet features, which includes the major information of the MR image pattern. By only computing the approximation component of DWT decomposition, it decreases the size of the MRI images, which eventually reduces the computation time and complexity of the classifier. Initially, MRI images were  $256 \times 256$  in size. After applying DWT with a three-level Haar wavelet decomposition (approximation component only) it changes to  $32 \times 32$ . Then PCA was applied on these reduced master feature sets, which allows a further decrease in the size of the feature sets by extracting the high variance components. In this article, four different feature subset sizes (5, 10, 15, and 20) were used. For classification purposes, the classifier used only 0.076%, 0.015%, 0.023%, and 0.031% of the original MR image in preparation of principal feature of size 5, 10, 15, and 20, respectively. Therefore, the proposed classification system achieved approximately 99.969% feature reduction while retains the accuracy of the classifier.

### 3.2. Performance Evaluation

The performance of the proposed multi-class classifier was evaluated in terms of macro-averaged recall, macro-averaged precision, macro-averaged F-measure, overall accuracy, and AUC of each class. Figure 3 shows the comparison of different decision models' performance against the number of principal features used.

Figure 3a illustrates the macro-averaged recall for each classifier model with respect to the feature subset sizes. A majority of the classifier models achieved  $\text{recall}_M$  values greater than 81% for any number of principal components were used. To observe the effect of the feature subset sizes, the results indicate that the LS-SVM (RBF) classifier model produced fixed 86%  $\text{recall}_M$  without dependence on feature subset sizes. However, RF and J48 models increase  $\text{recall}_M$  values as the number of principal feature subset sizes increase and the attained macro-averaged recall values increase to 96% and 87%, respectively, whereas the remaining classifier models (kNN and LS-SVM (Poly)) were not able to increase their performance in terms of  $\text{recall}_M$ .



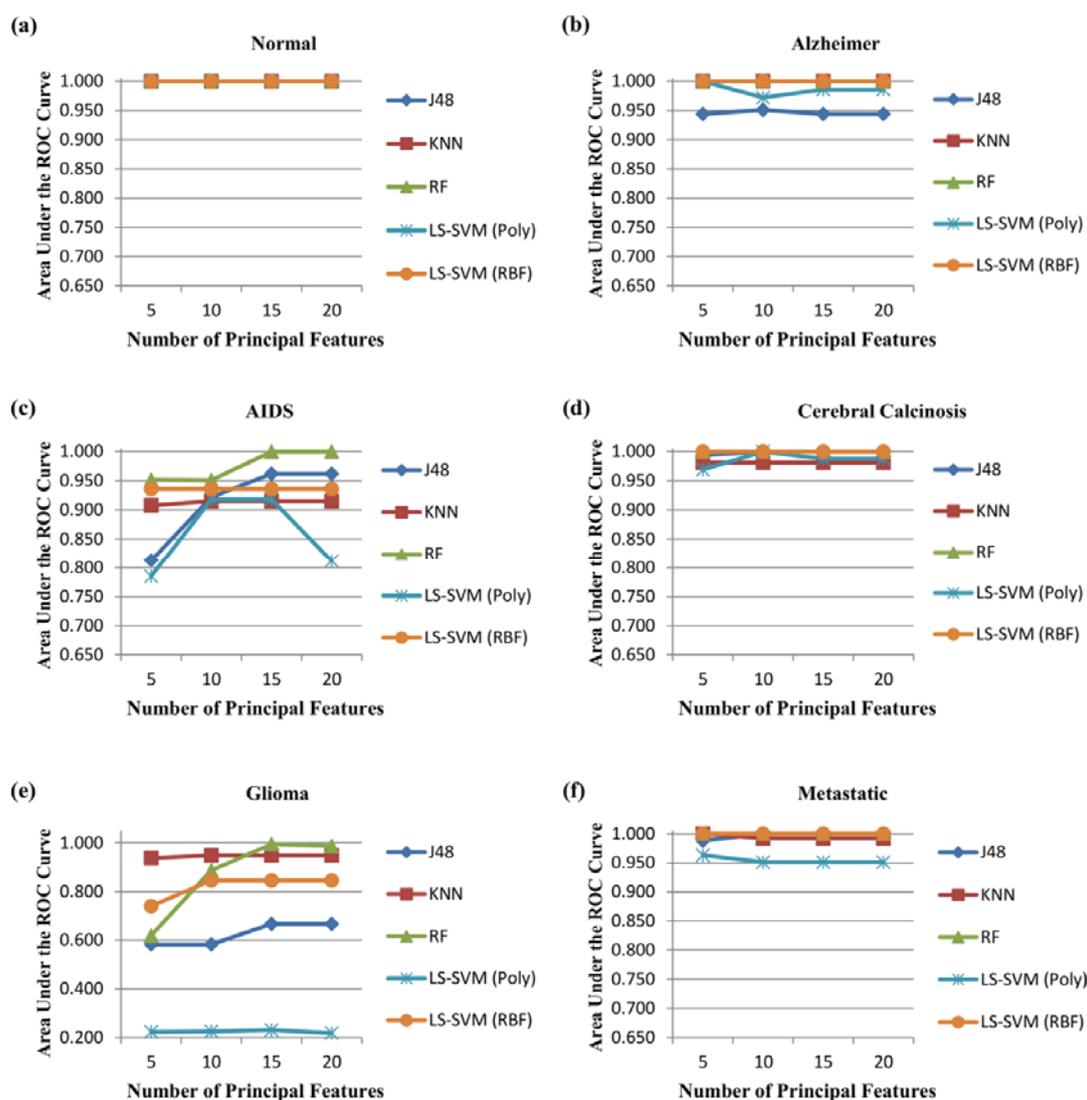
**Figure 3.** Performance measures of the proposed multi-class classifiers: (a) macro-averaged recall; (b) macro-averaged precision; (c) macro-averaged f-measure; and (d) average accuracy.

Figure 3b shows the performance of each classifier model with respect to the number of principal feature components in terms of macro-averaged precision. From the results, it is observed that a feature subset size of 10 or more produced precision<sub>M</sub> greater than 90% for all five classifier models. However, the RF model outperformed and achieved precision<sub>M</sub> values up to 96% using a feature subset size of 20. The lowest precision<sub>M</sub> was observed in LS-SVM (Poly) for any number of given principal features.

The performance evaluation in terms of macro-averaged F-measure, with respect to the number of principal components used by the classifiers, is shown in Figure 3c. The results revealed that the F-measure<sub>M</sub> generally exceeded 90% for RF and J48 when feature subset size used 10 or more. However, LS-SVM (RBF) achieved 90% F-measure<sub>M</sub> values for any combination of feature subsets. Furthermore, kNN and LS-SVM (Poly) could not able to improve the efficiency significantly in terms of F-measure<sub>M</sub> even the number of features was increased.

The overall accuracy of each classifier model was compared (Figure 3d). The average accuracy of each classifier model exceeded 84% for maximum number of principal components was used. However, RF improved the average accuracy with increasing the number of features and achieved the highest accuracy rate (i.e., 96%, standard deviation = 4%) when a plateau of 20 features was reached. The results again show that LS-SVM (RBF) overall average accuracy was stable and not associated with increased feature subset sizes. LS-SVM (RBF) provided the best results when the least number of principal features was used for classification. Furthermore, we found that no significant accuracy improvement was achieved, for any feature subset sizes, in the case of kNN and LS-SVM (Poly) decision models.

In Figure 4, the area under the ROC curve for each class was estimated. The results reveal that all five classifier models achieved AUC 100% with 0% standard deviation for the “Normal” class as shown in Figure 4a. RF, LS-SVM (RBF), and kNN produced significant results (i.e., 100% AUC with 0% standard deviation) for the “Alzheimer” class, as depicted in Figure 4b. However, J48 and LS-SVM (Poly) has a fluctuating trend of AUC for the “Alzheimer” class. The comparison of AUC for the “AIDS” class is shown in Figure 4c. Only the RF decision model achieved AUC 100% for the “AIDS” class when using equal to, and more than, a 15 feature subset. Moreover, the remaining four classifier models exceeded AUC 78% for different sizes of principal feature sets.



**Figure 4.** Area under the receiver operating characteristic (ROC) curve for each class: (a) normal; (b) Alzheimer; (c) AIDS; (d) cerebral calcinosis; (e) glioma; and (f) metastasis.

In Figure 4d,e, a majority of decision models attained  $AUC > 96\%$  with  $< 4\%$  standard deviation for classes “cerebral calcinosis” and “metastasis”, respectively. However, RF and LS-SVM (RBF) showed  $AUC$  100% with 0% standard deviation for any feature subset order. Figure 4c shows the  $AUC$  performance of the “glioma” class for all analyses. With the exception of RF and kNN models, all  $AUC$  measures were below 85% for any number of principal features and decision models. However, the  $AUC$  values measured for the RF and kNN models were greater than 95% when keeping the feature subset size more than equal to 15. The maximum  $AUC$  value achieved for “glioma” class by the RF model was 99% with 1% standard deviation.

The overall comparison results of decision models with different number of principal features subset showed that RF performed significantly better than other four decision models (J48, kNN, LS-SVM (Poly), and LS-SVM (RBF)). However, RF has not performed well when the lesser number of features is used, but its performance gradually increases as the size of the feature sets increased. It is also notable that increasing the number of principal features may not always be worthy because the complexity of the machine learning classifier may be increased by using larger feature subset sizes. Therefore, from the results it is observed that RF required at least 15 features to achieve better performance in terms of accuracy, precision<sub>M</sub>, recall<sub>M</sub>, F-measure<sub>M</sub>, and  $AUC$ . On the other hand,

the overall comparison also reveals that LS-SVM (RBF) achieved significant performance regardless of the feature set size. LS-SVM (RBF) achieved a constant performance trend for any number of principal features used, which leads to a decrease in the computational time and complexity of the multi-class classifier.

### 3.3. Comparison with Existing State-of-the-Art Classification Schemes

Different classification schemes were compared to evaluate the proposed multi-class classifier performance, which were examined for the same MRI dataset and on the same platform for multi-class brain MRI classification. Initially, these methods [6,18,20,21] were proposed for binary classification of brain MR images. The average accuracy results (when using 20 principal features) were gathered for each scheme, as presented in Table 3.

**Table 3.** Performance comparison with different classification schemes.

Scheme	Proposed in	Average Accuracy (%)
DWT + PCA + kNN	[18]	83.87
RT + PCA + LS-SVM (RBF)	[20]	86.02
DWPT + GEPSVM	[21]	88.92
DWT + PCA + LS-SVM (RBF)	[6]	89.25
DWT + PCA + RF	this paper	95.70

The results reveal that the highest accuracy (i.e., 95.7%) for multi-class brain MRI classification is achieved by the proposed scheme with RF as a decision model. LS-SVM (RBF) classification scheme also provided promising results with 89.25% accuracy rate. The scheme proposed by Das et al. [20] for binary classification attained an average accuracy of 86.02% when applied to multi-class brain MRI classification, regardless of the complex algorithm used (i.e., Ripplet transform) for feature extraction. Furthermore, the complex method used in [21] is not able to increase the accuracy rate up to 90% or more and managed to achieve a correctness rate of 88.92%. The average accuracy of the kNN-based scheme (i.e., 83.87%) is the worst performance among all of the compared state-of-the-art schemes examined for multi-class brain MRI classification. In addition, the method proposed in [24] includes age, tumour shape and ROI as feature sets for multi-class classification of brain MRI diseases. This scheme needs to trace ROIs manually, which makes this scheme semiautomatic. However, our proposed multi-class classifier is automatic and has no need of human intervention for decision-making purposes. In [24], the authors have achieved mean accuracy as 81.1%, 89.8%, and 91.2% for LDA, kNN, and non-linear SVM based decision models, respectively. Regardless of human intervention involved in feature extraction scheme, which is an additional cost, the average accuracy claimed in [24] is almost 4% less than our proposed work.

It is observed from the results and comparisons that the proposed classifier performance is quite remarkable as compared to the existing state-of-the-art techniques. Furthermore, a comprehensive study of different decision models' performances on MRI brain multi-class classification is shown. The comparison of decision models suggests that RF, LS-SVM (RBF), and J48 are more accurate than kNN and LS-SVM (POLY). RF achieved the highest accuracy rate when 20 features were used. Moreover, LS-SVM (RBF) maintained its performance for any number of features, which leads to being more advantageous when the least number of principal features are available. The feature engineering scheme used in this study proves that it reduced the number of discriminating features, which eventually reduces the classifier complexity and enhances its accuracy. Furthermore, the proposed classifier has the potential to classify various disease classes accurately using brain MRIs. The limitation of this study is that the experiments only involved brain MR images. However, the proposed approach has the potential to produce accurate results for different body parts' MR images as well.

#### 4. Conclusions

In this article, a multi-class classifier has been developed to classify brain MR slices as normal, Alzheimer, AIDS, cerebral calcinosis, glioma, or metastasis. It is composed of fast DWT, PCA, and five different decision models. The proposed medical decision support system yielded better performance in terms of macro-averaged recall, macro-averaged precision, macro-averaged F-measure, overall accuracy and AUC for each class, when compared to the other state-of-the-art schemes. This study provides a comprehensive comparison of different decision models performance, which concludes that RF work more accurately than other classification models (J48, kNN, LS-SVM (POLY), and LS-SVM (RBF)). It is evident from the results that the proposed classifier has the potential to classify the brain MR images accurately. Furthermore, the promising results indicate that the general practitioners can use this automated multi-class classifier as a second opinion, which assist them to reach the final decision more quickly. In future, the proposed method can be extended for automated classification of different pathological conditions and disease types, which are manually identified by the MRI scans. Moreover, this work can be employed on other imaging modalities (such as CT-scan, PET, and SPECT) datasets as well.

**Acknowledgments:** Muhammad Faisal Siddiqui likes to thank the University of Malaya Bright Sparks program. This research work was partially supported by Faculty of Computer Science and Information Technology, University of Malaya under a special allocation of Post Graduate Fund. This research was also partially funded by the University Malaya Research Grant (No: RP028F-14AET).

**Author Contributions:** Muhammad Faisal Siddiqui and Ghulam Mujtaba conceived and designed the experiments; Muhammad Faisal Siddiqui and Ghulam Mujtaba performed the experiments; Muhammad Faisal Siddiqui, Ghulam Mujtaba, Ahmed Wasif Reza, and Liyana Shuib analysed the data; Muhammad Faisal Siddiqui, Ghulam Mujtaba, Ahmed Wasif Reza, and Liyana Shuib contributed reagents/materials/analysis tools; Muhammad Faisal Siddiqui wrote the paper.

**Conflicts of Interest:** The authors declare no conflict of interest.

#### References

1. McKhann, G.; Drachman, D.; Folstein, M.; Katzman, R.; Price, D.; Stadlan, E.M. Clinical diagnosis of Alzheimer's disease report of the nincls-adrda work group\* under the auspices of department of health and human services task force on Alzheimer's disease. *Neurology* **1984**, *34*, 939. [[CrossRef](#)] [[PubMed](#)]
2. Sahu, O.; Anand, V.; Kanhangad, V.; Pachori, R.B. Classification of magnetic resonance brain images using bi-dimensional empirical mode decomposition and autoregressive model. *Biomed. Eng. Lett.* **2015**, *5*, 311–320. [[CrossRef](#)]
3. Prasad, P.V. *Magnetic Resonance Imaging: Methods and Biologic Applications*; Springer Science & Business Media: New York, NY, USA, 2006; Volume 124.
4. Maji, P.; Chanda, B.; Kundu, M.K.; Dasgupta, S. Deformation correction in brain MRI using mutual information and genetic algorithm. In Proceedings of the International Conference on Computing: Theory and Applications, Kolkata, India, 5–7 March 2007; pp. 372–376.
5. Scapaticci, R.; Di Donato, L.; Catapano, I.; Crocco, L. A feasibility study on microwave imaging for brain stroke monitoring. *Prog. Electromagn. Res. B Pier B* **2012**, *40*, 305–324. [[CrossRef](#)]
6. Siddiqui, M.F.; Reza, A.W.; Kanesan, J. An automated and intelligent medical decision support system for brain MRI scans classification. *PLoS ONE* **2015**, *10*, e0135875.
7. Mujtaba, G.; Shuib, L.; Raj, R.G.; Rajandram, R.; Shaikh, K. Automatic text classification of ICD-10 related CoD from complex and free text forensic autopsy reports. In Proceedings of the 5th IEEE International Conference on Machine Learning and Applications (ICMLA), Anaheim, CA, USA, 18–20 December 2016; pp. 1055–1058.
8. Mujtaba, G.; Shuib, L.; Raj, R.G.; Rajandram, R.; Shaikh, K.; Al-Garadi, M.A. Automatic ICD-10 multi-class classification of cause of death from plaintext autopsy reports through expert-driven feature selection. *PLoS ONE* **2017**, *12*, e0170242. [[CrossRef](#)] [[PubMed](#)]

9. Jin, Y.; Wee, C.Y.; Shi, F.; Thung, K.H.; Ni, D.; Yap, P.T.; Shen, D. Identification of infants at high-risk for autism spectrum disorder using multiparameter multiscale white matter connectivity networks. *Hum. Brain Mapp.* **2015**, *36*, 4880–4896. [[CrossRef](#)]
10. Huang, L.; Jin, Y.; Gao, Y.; Thung, K.-H.; Shen, D.; Initiative, A.S.D.N. Longitudinal clinical score prediction in Alzheimer's disease with soft-split sparse regression based random forest. *Neurobiol. Aging* **2016**, *46*, 180–191. [[CrossRef](#)] [[PubMed](#)]
11. Mwangi, B.; Ebmeier, K.P.; Matthews, K.; Steele, J.D. Multi-centre diagnostic classification of individual structural neuroimaging scans from patients with major depressive disorder. *Brain* **2012**, *135*, 1508–1521. [[CrossRef](#)] [[PubMed](#)]
12. Klöppel, S.; Stonnington, C.M.; Barnes, J.; Chen, F.; Chu, C.; Good, C.D.; Mader, I.; Mitchell, L.A.; Patel, A.C.; Roberts, C.C. Accuracy of dementia diagnosis—A direct comparison between radiologists and a computerized method. *Brain* **2008**, *131*, 2969–2974. [[CrossRef](#)] [[PubMed](#)]
13. Faisal, A.; Parveen, S.; Badsha, S.; Sarwar, H.; Reza, A.W. Computer assisted diagnostic system in tumor radiography. *J. Med. Syst.* **2013**, *37*. [[CrossRef](#)] [[PubMed](#)]
14. Zhang, Y.; Wu, L.; Wang, S. Magnetic resonance brain image classification by an improved artificial bee colony algorithm. *Prog. Electromagn. Res. Pier* **2011**, *116*, 65–79. [[CrossRef](#)]
15. Zhang, Y.; Dong, Z.; Wu, L.; Wang, S. A hybrid method for MRI brain image classification. *Expert Syst. Appl.* **2011**, *38*, 10049–10053. [[CrossRef](#)]
16. Zhang, Y.; Wu, L. An MR brain images classifier via principal component analysis and kernel support vector machine. *Prog. Electromagn. Res. Pier* **2012**, *130*, 369–388. [[CrossRef](#)]
17. Zhang, Y.; Wang, S.; Ji, G.; Dong, Z. An MR brain images classifier system via particle swarm optimization and kernel support vector machine. *Sci. World J.* **2013**, *2013*. [[CrossRef](#)] [[PubMed](#)]
18. El-Dahshan, E.-S.A.; Hosny, T.; Salem, A.-B.M. Hybrid intelligent techniques for MRI brain images classification. *Digit. Signal Process.* **2010**, *20*, 433–441. [[CrossRef](#)]
19. El-Dahshan, E.-S.A.; Mohsen, H.M.; Revett, K.; Salem, A.-B.M. Computer-aided diagnosis of human brain tumor through MRI: A survey and a new algorithm. *Expert Syst. Appl.* **2014**, *41*, 5526–5545. [[CrossRef](#)]
20. Das, S.; Chowdhury, M.; Kundu, M.K. Brain MR image classification using multiscale geometric analysis of Ripplet. *Prog. Electromagn. Res. Pier* **2013**, *137*, 1–17. [[CrossRef](#)]
21. Zhang, Y.; Dong, Z.; Wang, S.; Ji, G.; Yang, J. Preclinical diagnosis of magnetic resonance (MR) brain images via discrete wavelet packet transform with Tsallis entropy and generalized eigenvalue proximal support vector machine (GEPSVM). *Entropy* **2015**, *17*, 1795–1813. [[CrossRef](#)]
22. Zhou, X.; Wang, S.; Xu, W.; Ji, G.; Phillips, P.; Sun, P.; Zhang, Y. Detection of pathological brain in MRI scanning based on wavelet-entropy and naive Bayes classifier. In Proceedings of the Bioinformatics and Biomedical Engineering, Granada, Spain, 15–17 April 2015; pp. 201–209.
23. Wang, S.; Lu, S.; Dong, Z.; Yang, J.; Yang, M.; Zhang, Y. Dual-tree complex wavelet transform and twin support vector machine for pathological brain detection. *Appl. Sci.* **2016**, *6*, 169. [[CrossRef](#)]
24. Zacharaki, E.I.; Wang, S.; Chawla, S.; Soo Yoo, D.; Wolf, R.; Melhem, E.R.; Davatzikos, C. Classification of brain tumor type and grade using MRI texture and shape in a machine learning scheme. *Magn. Reson. Med.* **2009**, *62*, 1609–1618. [[CrossRef](#)] [[PubMed](#)]
25. Chaplot, S.; Patnaik, L.; Jagannathan, N. Classification of magnetic resonance brain images using wavelets as input to support vector machine and neural network. *Biomed. Signal Process. Control* **2006**, *1*, 86–92. [[CrossRef](#)]
26. Zhang, Y.; Wang, S.; Wu, L. A novel method for magnetic resonance brain image classification based on adaptive chaotic pso. *Prog. Electromagn. Res. Pier* **2010**, *109*, 325–343. [[CrossRef](#)]
27. Maitra, M.; Chatterjee, A. A slantlet transform based intelligent system for magnetic resonance brain image classification. *Biomed. Signal Process. Control* **2006**, *1*, 299–306. [[CrossRef](#)]
28. Lustig, M.; Donoho, D.L.; Santos, J.M.; Pauly, J.M. Compressed sensing MRI. *IEEE Signal Process. Mag.* **2008**, *25*, 72–82. [[CrossRef](#)]
29. Blum, A.L.; Langley, P. Selection of relevant features and examples in machine learning. *Artif. Intell.* **1997**, *97*, 245–271. [[CrossRef](#)]
30. Kohavi, R.; John, G.H. Wrappers for feature subset selection. *Artif. Intell.* **1997**, *97*, 273–324. [[CrossRef](#)]
31. Wettschereck, D.; Aha, D.W.; Mohri, T. A review and empirical evaluation of feature weighting methods for a class of lazy learning algorithms. *Artif. Intell. Rev.* **1997**, *11*, 273–314. [[CrossRef](#)]

32. Sengur, A. An expert system based on principal component analysis, artificial immune system and fuzzy k-NN for diagnosis of valvular heart diseases. *Comput. Biol. Med.* **2008**, *38*, 329–338. [[CrossRef](#)] [[PubMed](#)]
33. Duda, R.O.; Hart, P.E.; Stork, D.G. *Pattern Classification*; John Wiley & Sons: New York, NY, USA, 2012.
34. Zhao, Y.; Zhang, Y. Comparison of decision tree methods for finding active objects. *Adv. Space Res.* **2008**, *41*, 1955–1959. [[CrossRef](#)]
35. Bao, Y.; Ishii, N.; Du, X. Combining multiple k-nearest neighbor classifiers using different distance functions. In Proceedings of the International Conference on Intelligent Data Engineering and Automated Learning, Exeter, UK, 25–27 August 2004; pp. 634–641.
36. Fukunaga, K. *Introduction to Statistical Pattern Recognition*; Academic Press: Salk Lake City, UT, USA, 2013.
37. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
38. Liaw, A.; Wiener, M. Classification and regression by random forest. *R News* **2002**, *2*, 18–22.
39. Vapnik, V. *The Nature of Statistical Learning Theory*; Springer Science & Business Media: New York, NY, USA, 2013.
40. Patil, N.; Shelokar, P.; Jayaraman, V.; Kulkarni, B. Regression models using pattern search assisted least square support vector machines. *Chem. Eng. Res. Des.* **2005**, *83*, 1030–1037. [[CrossRef](#)]
41. Wang, F.-F.; Zhang, Y.-R. The support vector machine for dielectric target detection through a wall. *Prog. Electromagn. Res. Pier Lett.* **2011**, *23*, 119–128. [[CrossRef](#)]
42. Chen, G.-C.; Juang, C.-F. Object detection using color entropies and a fuzzy classifier. *IEEE Comput. Intell. Mag.* **2013**, *8*, 33–45. [[CrossRef](#)]
43. Magalhães, F.; Sousa, R.; Araújo, F.M.; Correia, M.V. Compressive sensing based face detection without explicit image reconstruction using support vector machines. In Proceedings of the 10th International Conference on Image Analysis and Recognition, Berlin, Germany, 26–28 June 2013; pp. 758–765.
44. Dasgupta, J.; Bhattacharya, K.; Chanda, B. A holistic approach for off-line handwritten cursive word recognition using directional feature based on arnold transform. *Pattern Recogn. Lett.* **2016**, *79*, 73–79. [[CrossRef](#)]
45. Komiyama, Y.; Banno, M.; Ueki, K.; Saad, G.; Shimizu, K. Automatic generation of bioinformatics tools for predicting protein–ligand binding sites. *Bioinformatics* **2015**, *32*, 901–907. [[CrossRef](#)] [[PubMed](#)]
46. Cristianini, N.; Shawe-Taylor, J. *An introduction to SVM*; Cambridge University Press: Cambridge, UK, 1999.
47. Suykens, J.A.; Vandewalle, J. Least squares support vector machine classifiers. *Neural Process. Lett.* **1999**, *9*, 293–300. [[CrossRef](#)]
48. Van Gestel, T.; Suykens, J.A.; Baesens, B.; Viaene, S.; Vanthienen, J.; Dedene, G.; De Moor, B.; Vandewalle, J. Benchmarking least squares support vector machine classifiers. *Mach. Learn.* **2004**, *54*, 5–32. [[CrossRef](#)]
49. Sokolova, M.; Lapalme, G. A systematic analysis of performance measures for classification tasks. *Inf. Process. Manag.* **2009**, *45*, 427–437. [[CrossRef](#)]

