A hybrid time-wavelet domain approach for early-stage diagnosis of Parkinson's disease using multichannel EEG signals

Shivani Dhok[a,*], Varad Pimpalkhute[a], Ankit A. Bhurane[a], Manish Sharma[b], U. Rajendra Acharya[c,d,e]

[a]*Department of Electronics and Communication, Indian Institute of Information Technology, Nagpur (IIITN), India*
[b]*Department of Electrical Engineering, Institute of Infrastructure, Technology, Research and Management (IITRAM), Ahmedabad, India*
[c]*Department of Electronics and Computer Engineering, Ngee Ann Polytechnic, Singapore 599489, Singapore*
[d]*Department of Biomedical Engineering, School of Science and Technology, SUSS, Singapore.*
[e]*School of Medicine, Faculty of Health and Medical Sciences, Taylors University, 47500 Subang Jaya, Malaysia*

Abstract

Parkinson's disease is the second most common neurodegenerative disorder with difficult early stage detection. The specific diagnosis techniques for PD are not available. It is diagnosed on the basis of the symptoms reviewed by the patients. These symptoms arise when 60-80% of the dopamine-producing cells are dead. Due to this fact, the early-stage detection of PD is important. The reduction in neurons affect the EEG, and minor abnormalities can be observed. A hybrid time-wavelet domain approach for early-stage detection of PD using inter-channel self-similarity and multi-resolution features extracted in the time domain and wavelet domain respectively has been proposed in this paper. In order to extract inter-channel self-similarity, a set of correlation coefficients is extracted in the time domain. Rényi entropy and Kraskov entropy features are extracted in the wavelet domain. Optimal biorthogonal wavelet filter bank (OBWFB) has been used to carry out . The wavelet transform allows the utilization of the predominant quadrature bandpass frequency components, resulting in the enhancement of the

*Corresponding author
Email addresses: shivanid17499@gmail.com (Shivani Dhok), pimpalkhutevarad@gmail.com (Varad Pimpalkhute), ankit.bhurane@gmail.com (Ankit A. Bhurane), manishsharma.iitb@gmail.com (Manish Sharma), aru@np.edu.sg (U. Rajendra Acharya)

entropy features. In this paper, we have examined each feature as well as the features in conjunction. The classification is carried out using the support vector machine of polynomial degree 3 (CSVM). 10-fold cross-validation is exercised to reduce the over-fitting phenomenon. The proposed method provides excellent results with accuracy, specificity and sensitivity as 99.56%, 99.24% and 99.87% respectively. The proposed method can be further used for diagnosis of various neurological diseases like epilepsy, sleep disorder etc.

Keywords: Parkinsons disease, Electroencephalogram, wavelet transform, optimal biorthogonal wavelet filter bank, correlation coefficients, Rényi entropy, Kraskov entropy

---

1. Introduction

Parkinson's disease (PD) is a neurodegenerative disorder in which the dopamine-producing brain cells die, resulting in the progressive retrogression of the motors function. It is the second most common neurodegenerative disorder, first being Alzheimer. Dopamine acts as a neurotransmitter between substantia nigra and the corpus striatum, which is responsible for smooth and controlled movements[1, 2, 3]. Lack of dopamine results in the failure of transmission between the substantia nigra and corpus striatum, leading to movement impairment [4]. The people affected by PD are mostly in the age group of 60+, but many in the age group of 40-60 are also seen to be affected. Every year, around 1 million people are affected by PD in India alone [5].

Diagnosis of PD is carried out on the basis of the reviews of the symptoms given by the patient and their medical history. These symptoms include tremors, rigidity, bradykinesia (slow movement), postural instability, speech and writing changes [6]. The symptoms arise when 60-80% of dopamine-producing cells are dead [7]. Over the time, these symptoms turn from mild to severe. Although many imaging techniques like magnetic resonance imaging (MRI), computed tomography (CT) scan, positron emission tomography (PET) scan and DaTscans can be used for PD detection, these techniques are expensive and require expertise. However, specific tests for PD diagnosis are not available. A low cost and non-invasive technique need to be devised for

Table 1: Abbreviations

Abbreviation	Definition
PD	Parkinson's Disease
EEG	Electroencephalogram
RE	Rényi Entropy
OBWFB	Optimal Biorthogonal Wavelet Filter Bank
SVM	Support Vector Machine
QSVM	Quadratic Support Vector Machine
CSVM	Cubic Support Vector Machine

early-stage PD detection.

PD affected subject's Electroencephalogram (EEG) signals might contain some minuscule abnormalities which cannot be observed by naked eyes. In this paper, these abnormalities in the 1-D multi-channel EEG signals are used for classification of normal and PD subjects. The main contribution of this paper are:

- The method develops a hybrid time-wavelet domain approach for detection of Parkinson's disease using EEG signals.
- For, wavelet decomposition, optimal biorthoganal wavelet filter banks have been used. Two entropy features, Rényi entropy and Kraskov entropy have been calculated for the wavelet coefficients.
- For analyzing the inter channel self-similarity, correlation coefficients have been calculated in the time domain.
- The method analyzes each extracted feature independently as well as in conjunction.

2. Notations and Abbreviations

Table 1 show the abbreviations that we have used in this paper.

3. Related Work

In recent years, many methods for PD detection were devised. These methods can be broadly classified on the basis of the signals used for diagnosis. The signals used mainly consisted of voice signals, gait signals and EEG signals. A parallel neural network approach was developed by Astrom et al. to utilize the voice signals for detection of PD [8]. The authors have reported the accuracy of $91.20 \pm 1.60\%$. Voice signals were found to be widely used for PD detection. Another method was proposed by Chen et al., who developed a PD detection method using fuzzy k-nearest neighbourhood (FKNN) classifier [9]. The method proposed by them resulted in an average accuracy of 96.02% with specificity and sensitivity as 95.53% and 96.22% respectively. The accuracy by FKNN was enhanced by Zuo et al. by introducing swarm optimization [10]. This step resulted in increased accuracy of 97.47%. Another method that has used voice signals was implemented by Ma et al. [11]. The authors deployed a kernel based extreme learning machine (KELM) that resulted in an average accuracy of 99.49%. Another class of signals used is gait signals. Daliri used the gait signals in order to identify the PD patients [12]. The Chi-square distance kernel of gaits method used by Daliri resulted in an accuracy of 91.20% with specificity and sensitivity 89.92% and 91.71% respectively. Some of the methods developed made use of EEG signals, one such method was devised by Yuvraj et al. [13]. The authors used bispectrum features for classification of PD affected EEG signals, resulting in classification accuracy of 99.62%. Recently, Oh et al. employed a 13 layer convolutional neural networks (CNN) for PD detection, reporting an average accuracy of 88.25% with specificity and sensitivity being 91.77% and 84.71% respectively[13].

Although, EEG signals have been used for detection of various other diseases like sleep disorder, epilepsy etc [14, 15, 16, 17], it can be observed that very limited work for PD detection has been developed using EEG signals. In this work we try to capture the minuscule abnormalities in the EEG of the PD affected people to distinguish the normal and PD affected EEG signals. Entropy features have been widely used for analysis of biomedical signals[18, 19, 20, 21, 22, 23, 24, 25]. Permutation Rényi entropy was used to differentiate interictal states from ictal states by implementing spatial tem-

poral analysis of the EEG signals by Mammone et al. [21]. Sharma et al. have used various entropy parameters like average Shannon entropy, average Rényi entropy, average approximate entropy, average sample entropy and average phase entropy to classify the focal and non-focal EEG signals [18]. The authors have decomposed EEG signals using the empirical mode decomposition (EMD) to extract intrinsic mode functions (IMF). To differentiate the normal and the coronary artery disease (CAD) patients, Kumar et al. have extracted the KNN entropy and fuzzy entropy from the FAWT decomposed sub-bands of the HRV [23]. Bhattacharya et al. have implemented a multi-scale entropy measure technique by extracting Kraskov entropy (K-nearest neighbour entropy) for characterization of epileptic seizure, non-seizure and normal EEG signals using the tunable-Q wavelet transform. We have extracted Rényi entropy and Kraskov entropy for predominant principal frequency components of the EEG signals in the wavelet domain to classify them as normal and PD affected. For wavelet decomposition, optimal biorthogonal wavelet filter banks have been employed. To enhance the accuracy further, we have determined the inter-channel self-similarity of the EEG signals by extracting the correlation coefficients as the time domain feature. Correlation coefficients have been used for performance evaluation as well as a classification feature many times [26, 27]. In this method we have explored correlation coefficient as a feature for characterization of EEG signals. In this proposed method, we have used a hybrid time-wavelet domain approach for classification of EEG signals. The paper is further organized as follows: the method proposed and the data used in this paper is described in detail in section 4. The results obtained are stated in section 5 which are further discussed in section 6. The conclusions are stated in section 7.

4. Materials and Proposed Method

4.1. Database used

EEG signals for 20 (9 men and 11 women) normal condition and 20 (10 men and 10 women) PD affected subjects were collected for a duration of 5 minutes with a sampling rate of 128 Hz. In order to remove eye blinking artefacts, a threshold technique to exclude amplitudes greater than $100\mu V$. Sixth order Butterworth bandpass filter

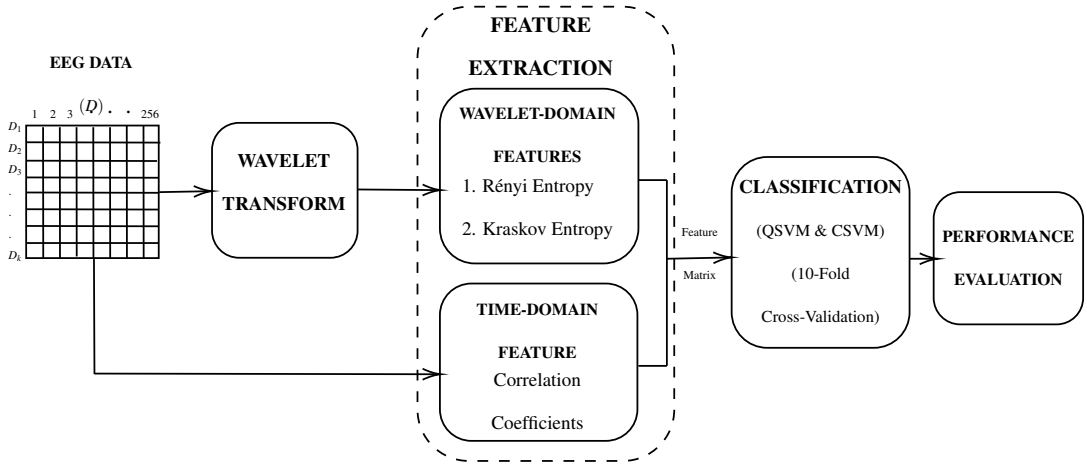

Figure 1: Work-flow for the proposed method

with forward reverse filtering method was implemented to filter a frequency range of 1-49Hz [28]. The PD affected subjects were in the age range of 45 to 65 years with an average period of PD of 5.75 ± 3.52 years. The exclusion specification includes the presence of neurological ailments like epilepsy, depression etc. The normal condition subjects had no past records of mental or neurological disorders. The mean age of the normal subjects was 58.10 ± 2.95 years.

The signal recorded for 5 minutes was divided in chunks of two seconds duration which resulted in the total 3159 sets of EEG signals. In these sets, 1588 were the normal condition signals and 1571 were PD affected signals.

4.2. Proposed Approach

The proposed approach is implemented in four stages. It begins with the preparation of input data for the feature extraction stage, which is the stage one. The stage two includes feature extraction and determination of feature matrix which is used for classification. In the third stage, the data signals are classified using the obtained features. The fourth stage is the parameters calculation. The complete work-flow of the method is described in Figure 1.

4.2.1. Data Preparation

Let D be the data matrix of size $l \times k$, where k is the number of channels and l be the number of samples of the EEG signal. For this method, $k = 14$ and $l = 256$. Let D_k be the channel-wise vector, where $k \in [1, 14]$. A total of 3159 signal-matrices are available.

The features extracted in this method are in two domains, time domain and wavelet domain. The feature correlation coefficient is extracted in the time domain whereas the features Kraskov entropy and Rényi entropy are extricated in the wavelet domain.

4.2.2. Feature extraction

The subsequent discussion explains the time domain and the wavelet domain features.

Time Domain Features:. Correlation coefficients is the feature which is extracted in the time domain.

- Correlation Coefficients:

  Correlation coefficients is a numerical measure of the statistical relationship between two random variables [29, 30, 31].

$$\rho(A,B) = \frac{1}{N-1} \sum_{i=1}^{n} \left(\frac{A_i - \mu_A}{\sigma_A} \right) \left(\frac{B_i - \mu_B}{\sigma_B} \right) \tag{1}$$

where mean of A is μ_A and standard deviation of A is σ_A. Similarly, the mean and standard deviation of B is μ_B and σ_B respectively. Further, the correlation coefficients in terms of the covariance of A and B are:

$$\rho(A,B) = \frac{cov(A,B)}{\sigma_A \sigma_B} \tag{2}$$

For this experiment, to ascertain the inter-channel self-similarities, the auto-correlation of the input matrix, $\rho(D', D')$ is calculated, where D' is the transpose of normalized input matrix (D). As $\rho(D', D')$ is symmetric, it can be represented with either strictly upper triangular or strictly lower triangular matrix, as shown in Figure 2. These elements are arranged in a row vector ρ, which is considered as the first feature for classification. This can be mathematically represented as,

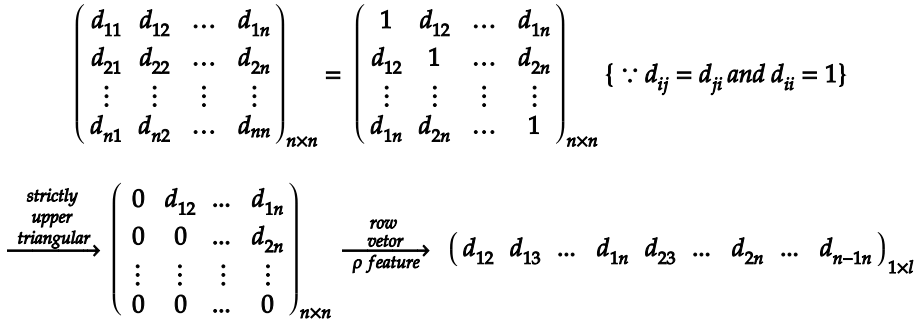

Figure 2: Formation of correlation coefficients feature

$$\rho(D', D') = [d_{ij}]_{n\times n}$$
$$\rho = [d_{ij}]_{1\times l}, \forall\, i > j \tag{3}$$

where, $l = \frac{n^2-n}{2}$.

Wavelet Domain Features:. Wavelet transform have been widely used in analysis of the biomedical signals [32, 33, 34, 35, 36, 37, 25, 20, 23]. Wavelet transform have been used for diagnosis of various neurological diseases like epilepsy [34, 25, 20], depression [35] etc.. We have implemented the wavelet transform using optimal biorthogonal wavelet filter bank for the diagnosis of PD. Two features are extracted in the wavelet domain, namely Rényi entropy and Kraskov entropy.

Optimal Biorthogonal Wavelet Filter Banks (OBWFB):. In order to perform wavelet transform, analysis filters were constructed using the optimal biorthogonal filter banks (OBWFB). Due to relaxation in the orthogonality condition, biorthogonal filters are symmetric and hence allow linear phase conditions as well as the integrity of the signal. The halfband constraint was applied to construct analysis *type-1* low pass filter of the OBWFB [38]. The perfect reconstruction condition include:

$$F_0(z)H_0(-z) + F_1(z)H_1(-z) = 0 \tag{4}$$

$$F_0(z)H_0(z) + F_1(z)H_1(z) = 2z^{-k}, k \in \mathbb{Z} \tag{5}$$

Where, H_0 and H_1 are the analysis low pass and high pass filters respectively and F_0 and F_1 are the synthesis low pass and high pass filters respectively. The product filter is defined as $P(z) = z^k H_0(z)F_0(z)$. The PR conditions and the product filter results into the condition for halfband filters, given by:

$$P(z) + P(-z) = 2 \tag{6}$$

Along with PR conditions, optimization of filters is necessary. To construct optimized filter, objective function is defined as,

$$\begin{aligned} \phi &= a^T \{\alpha_1 T + \alpha_2 F + \alpha_3 P + \alpha_4 S\} a \\ &= a^T R a \end{aligned} \tag{7}$$

Where, α_j are the weighting function ($\{\alpha_j : 0 \leq \alpha_j \leq 1, \sum_j \alpha_j = 1\}$); $T, F, P, S, R \in \mathbb{R}^{(N+1)\times(N+1)}$ are positive definite matrices and a is defined as:

$$a = [h(0)\sqrt{h(1)} \ldots \sqrt{2}h(N)]^T_{1\times N}$$

Let $h_0(n)$ be the analysis lpw pass filter and f_0 be the synthesis low pass filter, such that their lengths are $2P+1$ and $2Q+1$ respectively. Let $h \in \mathbb{R}^{(P+1)}$ and $f \in \mathbb{R}^{(Q+1)}$ be the optimization vectors, then the optimization problem for analysis low pass filter is given by,

- $h^T R h$: Cost function
- $Ah = 0$: Regularity constraint
- $Bh = 0$: Halfband constraint
- $h^T h = 1$: Unit norm constraint

The optimization problem for synthesis low pas filter is

- $f^T R f$: Cost function

- $Af = 0$: Regularity constraint
- $Bf = 0$: PR constraint
- $f^T f = 1$: Unit norm constraint

Where, $A \in \mathbb{R}^{(Q+1)\times M_s}$, $B \in \mathbb{R}^{(N+1)\times(P-1)/2}$ and $T \in \mathbb{R}^{(Q+1)\times(P+Q-1)/2}$ and A,B and T are given as:

$$[A]_{k,l} = \begin{cases} 1 & ;k,l = 0 \\ \sqrt{2}(l)^{2k}(-1)^l & ;otherwise \end{cases}$$

$$[B]_{k,l} = \begin{cases} \frac{1}{\sqrt{2}} & ;k = 0,1,\ldots,N; l = 2,4,\ldots,(N-1) \\ 0 & ;otherwise \end{cases}$$

$$[T]_{k,l} = \begin{cases} h_0[2k+2] & ;l = 0 \\ h_0[2(k+1)-l] + h_0[2(k+1)+l] & ;1 \le l \le Q \end{cases}$$

where, $k \in [0, \frac{P+Q-1}{2}]$. Using the specified conditions and constraints, $h_0(n)$ and $f_0(n)$ were designed. The $h_0(n)$ and $f_0(n)$ are of the order 14 and 28 respectively. The obtained filter coefficients are as shown in Table 2.

Wavelet Transform:. Wavelet decomposition of order three was calculated using the filters from OBWFB. The transform results in three detailed coefficients and one approximate coefficient which is used for calculation of the entropy features.

- Rényi Entropy:

  Rényi entropy generalizes Shannon entropy [39]. It is also known as alpha-order entropy. It is calculated as,

$$RE(X_k) = \left(\frac{1}{1-\alpha}\right) \log\left(\sum_{n=1}^{N} |X_k(n)|^{\alpha}\right) \tag{8}$$

  where, $N = 256$.

  Here, $\alpha \neq 1$. For the presented method, Rényi entropy of order 2 is calculated ($\alpha = 2$).

Table 2: Filter Coefficients

Index (n)	$h_0(n)$	$f_0(n)$
$\pm$ 0	-0.737373124493960	-0.768532189913231
$\pm$ 1	-0.457126071272320	-0.410562165837711
$\pm$ 2	0	0.0939230137118358
$\pm$ 3	0.127903440731053	0.115379716323613
$\pm$ 4	0	-0.0808062261565235
$\pm$ 5	-0.0516553232613650	-0.0472748305157014
$\pm$ 6	0	0.0642848894041939
$\pm$ 7	0.0121913915556523	0.0113064384044617
$\pm$ 8	-	-0.0300802689713363
$\pm$ 9	-	0
$\pm$ 10	-	0.00718058219124597
$\pm$ 11	-	0
$\pm$ 12	-	-0.00157367235354061
$\pm$ 13	-	0
$\pm$ 14	-	0.000186935505391014

So, the equation of Rényi entropy becomes,

$$RE = -\log\Big(\sum_{n=1}^{N} |X_k(n)|^2\Big) \tag{9}$$

Here, X_k determine the coefficients obtained after wavelet decomposition of D_k.

- Kraskov entropy

  Kraskov entropy is the estimate for the Shannon entropy[40]. It can be calculated as:

$$H(\hat{X_k}) = -\psi(k) + \psi(N) + \log(c) + \frac{d}{N}\sum_{i=1}^{N} \log(\varepsilon(i)) \tag{10}$$

  Where, ψ is the diagamma function, c is the volume of a d-dimensional unit ball,$\varepsilon(i)$ is twice the distance of x_i to its k^{th} neighbour, N is the number of bivariate measurements $(N = 256)$ and X_k, input signal is the random variable in the metric space[41]. Here, X_k are the coefficients obtained from wavelet decomposition of D_k. The entropy features are obtained at each level of wavelet decomposition and are further concatenated to form a vector.

The features obtained for all the fourteen channels are concatenated to form a feature vector. The features obtained for each data matrix D together form the feature matrix which is further sent to the classification stage.

4.2.3. Classification

After extracting the features, they are used for further classification. The classification is effectuated using the cubic support vector machine (CSVM)[42]. Given a labeled training data set (supervised learning) to the SVM model, the training algorithm builds a model which outputs a plane which classifies the new examples[43, 44]. A d^{th} degree kernel polynomial $K(p,q)$ is defined as: $K(p,q) = (p^T q + c)^d$, where $c \geq 0$ is a free parameter and p and q are the vectors. For CSVM, $d = 3$.

For k-fold cross-validation, the data set is divided into k groups, or folds, of approximately equal size [45]. Of the k folds, the first is taken as a validation set while the remaining $k-1$ sets are used for training the model[46]. This paper uses a 10-fold cross validation, that is, $k = 10$.

4.2.4. Parameters Calculation

For performance evaluation, various parameters are calculated from the confusion matrix obtained after classification stage. These parameters include accuracy, specificity, sensitivity, precision, false positive rate, negative predictive value, false discovery rate, F-1 score and Mathews correlation coefficient [47].

5. Results

The presented approach was implemented on a personal computer equipped with Intel (R) Core (TM) CPU @2.30GHz 2.30GHz, 8GB RAM and Windows 10 (64 bit) operating system. The classification was carried out using MATLAB R2018a software, with no other processes running in parallel. In order to easily capture the complex relationships between the data, CSVM classifier was used for the classification. Moreover, to reduce the over-fitting phenomenon, 10-fold cross-validation was implemented.

The training time for this method was observed to be 2.8234s. The Table 3 lists the training time and prediction rate. Classification results into confusion matrix. The confusion matrices are tabulated in Table 4. The confusion matrices are used for calculating parameters, which are used for performance evaluation. The performance, in terms of classification parameters, is listed in Table 5 for both the classification machines. We observe that the proposed method results in appreciable classification accuracy. The classification error was observed to be 0.44%. Low classification error and increased training time indicate the complexity of the classifier machine.

In order to evaluate the contribution of each feature, feature-wise performance is determined which indicate the ascendancy of *RE* feature over the other two. The Table 6 prognosticates the feature-wise performance for both the classifiers. The method also studies the cumulative effect of features on classification parameters, which are tabulated in Table 7.

Table 3: Training Time and Prediction Rates

Training Time (s)	Prediction Rate (obs/s)
2.8234	26000

Table 4: Confusion matrix

Classifier: CSVM		Predicted Label	
		Normal	PD
True Label	Normal	1576	12
	PD	2	1569

Table 5: Parameters' values obtained after classification

Parameters	**Values %**
Accuracy	99.56
Specificity	99.24
Sensitivity	99.87
Precision	99.24
False Positive Rate	0.76
False Negative Rate	0.13
Negative Predictive Value	99.87
False Discovery Rate	0.76
F1 Score	99.56
Mathews Correlation Coefficient (%)	99.12

Table 6: Feature-wise parameters

Parameters	ρ	$\hat{H}$	RE
Accuracy (%)	97.59	98.96	99.34
Specificity (%)	97.40	98.79	99.11
Sensitivity (%)	97.79	99.12	99.56
Precision (%)	97.42	98.80	99.12
False Positive Rate (%)	2.60	1.21	0.89
False Negative Rate (%)	2.21	0.88	0.44
Negative Predictive Value (%)	97.77	99.11	99.55
False Discovery Rate (%)	2.58	1.20	0.88
F1 Score (%)	97.60	97.91	99.34
Mathews Correlation Coefficient (%)	95.19	96.84	98.67

Table 7: Cumulative effect of features on evaluated parameters

Parameters	RE	$RE+\hat{H}$	$RE+\hat{H}+\rho$
Accuracy (%)	99.34	99.27	99.56
Specificity (%)	99.11	99.17	99.24
Sensitivity (%)	99.56	99.37	99.87
Precision (%)	99.12	99.18	99.24
False Positive Rate (%)	0.89	0.83	0.76
False Negative Rate (%)	0.44	0.63	0.13
Negative Predictive Value (%)	99.36	99.11	99.87
False Discovery Rate (%)	0.88	0.82	0.76
F1 Score (%)	99.34	99.28	99.56
Mathews Correlation Coefficient (%)	98.67	98.54	99.12

6. Discussion

The stated method implements the classification model for detection of Parkinson's disease using a hybrid time-wavelet domain approach. Spectral features extracted in the wavelet domain include Rényi entropy (RE) and Kraskov entropy $(\hat{H})$. The spatial features include correlation coefficients (ρ).

To evaluate the performance of the features, we have implemented a feature ranking algorithm based on the class-separability criterion [48, 49, 50]. The feature vector of size 1×203 is formed after combining the RE, $\hat{H}$ and ρ vectors. All 203 features are ranked and the performance of the proposed method is analyzed using the top 10%, 20%, ..., 100% is features. The accuracy trends for the same are pictorially represented in Figure 3. We observe that the accuracy increases with the increase in the number of features, which indicates that each and every feature contributes significantly to the classification of the medical data set. It is also observed that the rate of increase in the accuracy decreases with the increase in the number of features. Moreover, the dominance of RE, $\hat{H}$ and ρ features in the top-ranked features is determined and listed in Table 8. It is observed that the feature RE shows dominance over the other features followed by $\hat{H}$ and ρ. The results obtained from the feature ranking methodology

Table 8: Dominance of RE, $\hat{H}$ and ρ in the top features

Percentage of Features (%)	RE (%)	$\hat{H}$ (%)	ρ (%)
10	45.00	35.00	20.00
20	37.50	32.50	30.00
30	41.67	31.67	26.67
40	43.21	34.58	22.22
50	43.56	34.65	21.78
60	36.36	32.23	31.41
70	32.39	31.69	35.92
80	33.33	28.40	38.27
90	29.67	28.02	42.31
100	27.59	27.59	44.83

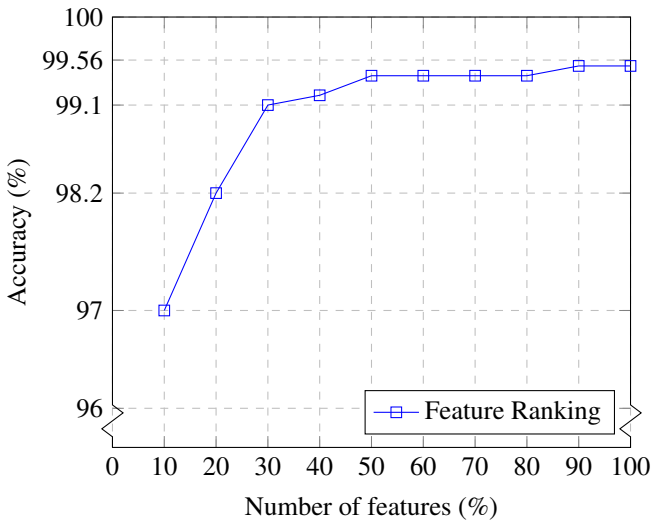

Figure 3: Accuracy (%) versus number of features obtained using feature ranking technique

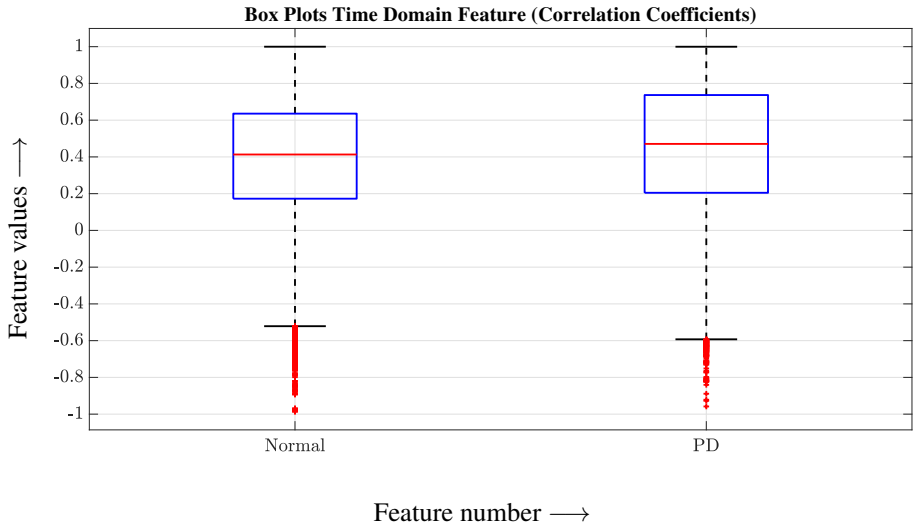

Figure 4: Box plot for time domain feature (correlation coefficients)

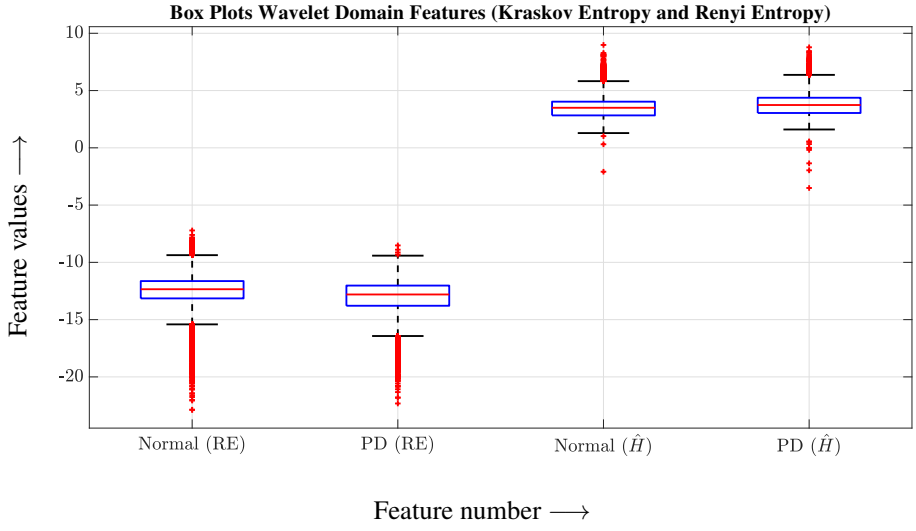

Figure 5: Box plot for wavelet domain feature (Kraskov entropy and Rényi entropy)

Table 9: Comparison with some state-of-the-art methods

Author	Technique	Validation	Accuracy (%)	Specificity (%)	Sensitivity (%)
Astrom et al.	Parallel neural network [8]	40% holdout validation	91.20	93.00	90.50
Zuo et al.	Particle swarm optimization enhanced fuzzy k-nearest neighbor (FKNN) [10]	10-fold cross validation	97.47	96.57	98.16
Daliri	Chi-square distance kernel of the gaits [12]	2-fold cross validation	91.20	89.92	91.71
Chen et al.	Fuzzy k-nearest neighbor (FKNN) [9]	10-fold cross validation	95.87	95.53	96.22
Ma et al.	Kernel-based extreme learning machine (KELM) [11]	10-fold cross validation	96.35	95.89	95.72
Oh et al.	Convolutional Neural Networks (CNN) [28]	10-fold cross validation	88.25	91.77	84.71
Yuvraj et al.	Bispectrum features [13]	10-fold cross validation	99.62	99.25	100.00
Present Work	Multiresolution and Correlation Coefficients (QSVM)	10-fold cross validation	**99.56**	**99.24**	**99.87**

are found to be in synchronization with the results obtained from feature-wise classification, listed in Table 6. Among the wavelet domain features, Rényi entropy feature shows higher classification accuracy $> 99\%$. The same can be observed from the box plot for the *RE* feature (refer Figure 5). The box plot for normal subjects is placed higher as compared to the PD. This can be concluded by looking at the upper and lower whiskers and the median. This distinguishing pattern is better for *RE* feature. The box plots for time domain and wavelet domain features are as shown in figures 4 and 5. The feature-wise evaluation parameters, without considering the feature ranking, are listed in Table 6. The results are listed in the increasing order of feature-wise classification performance. Further, the Kraskov entropy feature shows a better performance, followed by correlation coefficients. As mentioned earlier, the wavelet domain features outperform the time domain feature, correlation coefficients (ρ), but ρ contributes quite significantly in increasing the classification accuracy. In short, each feature contributes significantly to the classification process. The progressive effect of the features are prognosticated in Table 7. The Figure 6 shows the parallel co-ordinate plot obtained for the top 10 features after classification.

Moreover, the time required for classification is quite low. The training time was found to be 2.8234s with prediction rate of 26000 obs/s. The same is listed in Table 3. The presented model is compared with some of the state-of-the-art methods in Table

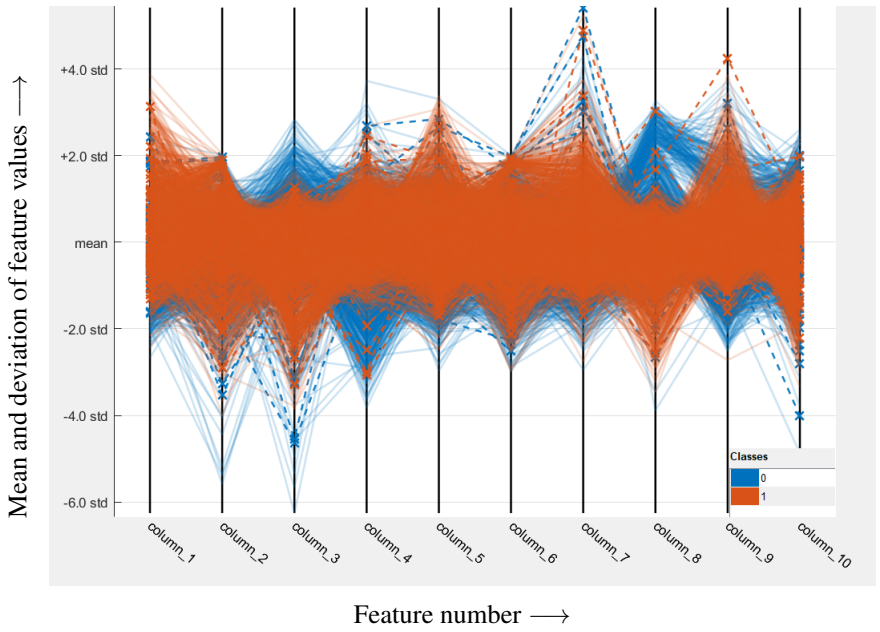

Figure 6: Parallel coordinate plots of normal (blue) and PD (orange) class.

9 and it is observed to be performing better than most of the previously implemented methods.

Following are the salient features of the proposed method:

- The method uses OBWFB for carrying out wavelet decomposition.
- We have studied the performance of wavelet domain (RE and $\hat{H}$) and time domain (ρ) features. The feature RE shows significant performance giving a performance accuracy $> 99\%$. Further, the method also observes significant contributions from the other two features.
- Feature ranking algorithm using class separability criterion is implemented to evaluate performance of the features.
- We have examined the progressive contribution of features with and without considering the feature ranking.
- The method does not have any data dependency.

This method can be considered as an effective tool for PD diagnosis but, it has the following limitations:

- Modelling of support vector machine is a time-consuming process.
- The classification requires large training samples. Memory requirements increase with the increase in data.

7. Conclusions

In this paper, we have proposed a hybrid time-wavelet domain method to diagnose PD disease using multi-channel EEG signals. We have explored the effect of multidomain features on the classification accuracy. Three features were extracted, in which the features RE and $\hat{H}$ were extracted in the wavelet domain, whereas, ρ was extracted in the time domain.The feature ρ was successfully found to be extracting the inter-channel self-similarities. Whereas, the performance of the RE and $\hat{H}$ in wavelet domain was found to be incremented by the employment of the quadrature bandpass filter coefficients. The wavelet decomposition was carried out using a set of optimal biorthogonal filters. The classification was carried out using support vector machines of polynomial degrees three (CSVM). 10-fold cross validation was exercised in order to reduce the problem of over-fitting. The method gives promising results with the classification accuracy of 99.56%. We have evaluated the performance of individual features. The Rényi entropy extracted in wavelet domain was found to be the most promising feature. Further, the features were combined and their progressive performances were studied.

The propounded method provides competitive classification accuracy, which is necessary while the diagnosis of a disease in a person and hence, this method is an efficient model which can be used for diagnosis of PD using EEG signals. Further, we intend to implement this methodology for early-stage detection of diseases like epilepsy, sleep disorder etc.

References

[1] N. I. of Neurological Disorders, Stroke, Parkinson's disease information page, last accessed 23 November 2018.
URL https://www.ninds.nih.gov/Disorders/All-Disorders/Parkinsons-Disease-Information-Page

[2] D. J. Surmeier, J. N. Guzman, J. Sanchez-Padilla, J. A. Goldberg, Chapter 4 - what causes the death of dopaminergic neurons in parkinsons disease?, in: A. Bjrklund, M. A. Cenci (Eds.), Recent Advances in Parkinsons Disease: Basic Research, Vol. 183 of Progress in Brain Research, Elsevier, 2010, pp. 59 – 77. doi:https://doi.org/10.1016/S0079-6123(10)83004-3.
URL http://www.sciencedirect.com/science/article/pii/S0079612310830043

[3] J. Silver, M. Schwab, P. Popovich, Central nervous system regenerative failure: Role of oligodendrocytes, astrocytes, and microglia, Cold Spring Harbor perspectives in biology 7 (3). doi:10.1101/cshperspect.a020602.

[4] C. P. D. Sietske N. Heyn, Parkinson's Disease symptoms, causes, stages, treatment, and life expectancy, last accessed 25 November 2018 (2018).
URL https://www.medicinenet.com/parkinsons_disease/article.htm

[5] WHO., Neurological disorders: public health challenges., World Health Organization, 2006.
URL https://www.who.int/mental_health/neurology/neurological_disorders_report_web.pdf

[6] M. C. Staff, Parkinson's Disease patient health care information, last accessed 25 November 2018 (2018).
URL https://www.mayoclinic.org/diseases-conditions/parkinsons-disease/symptoms-causes/syc-20376055

[7] Parkinson's disease center: Symptoms, treatments, causes, tests, diagnosis, and prognosis, last accessed 23 November 2018 (Nov. 2018).
URL https://www.webmd.com/parkinsons-disease/default.htm

[8] F. strm, R. Koker, A parallel neural network approach to prediction of parkinsons disease, Expert Systems with Applications 38 (10) (2011) 12470 – 12474. doi:https://doi.org/10.1016/j.eswa.2011.04.028.
URL http://www.sciencedirect.com/science/article/pii/S0957417411005446

[9] H.-L. Chen, C.-C. Huang, X.-G. Yu, X. Xu, X. Sun, G. Wang, S.-J. Wang, An efficient diagnosis system for detection of parkinsons disease using fuzzy k-nearest neighbor approach, Expert Systems with Applications 40 (1) (2013) 263 – 271. doi:https://doi.org/10.1016/j.eswa.2012.07.014.
URL http://www.sciencedirect.com/science/article/pii/S095741741200869X

[10] W.-L. Zuo, Z.-Y. Wang, T. Liu, H.-L. Chen, Effective detection of parkinson's disease using an adaptive fuzzy k-nearest neighbor approach, Biomedical Signal Processing and Control 8 (4) (2013) 364 – 373. doi:https://doi.org/10.1016/j.bspc.2013.02.006.
URL http://www.sciencedirect.com/science/article/pii/S1746809413000359

[11] J. H.-L. X.-H. Ma, Chao;Ouyang, An efficient diagnosis system for parkinsonx2019;s disease using kernel-based extreme learning machine with subtractive clustering features weighting approach, Computational and Mathematical Methods in Medicine 2014. doi:10.1155/2014/985789.
URL http://dx.doi.org/10.1155/2014/985789

[12] M. R. Daliri, Chi-square distance kernel of the gaits for the diagnosis of parkinson's disease, Biomedical Signal Processing and Control 8 (1) (2013) 66 – 70. doi:https://doi.org/10.1016/j.bspc.2012.04.007.

URL http://www.sciencedirect.com/science/article/pii/S1746809412000511

[13] Y. RAJAMANICKAM, U. R. Acharya, Y. Hagiwara, A novel parkinsons disease diagnosis index using higher-order spectra features in eeg signals, Neural Computing and Applicationsdoi:10.1007/s00521-016-2756-z.

[14] M. Kaur, G. Singh, Classification of seizure prone eeg signal using amplitude and frequency based parameters of intrinsic mode functions, Journal of Medical and Biological Engineering 37 (4) (2017) 540–553. doi:10.1007/s40846-017-0275-8.
URL https://doi.org/10.1007/s40846-017-0275-8

[15] A. K. Jaiswal, H. Banka, Local transformed features for epileptic seizure detection in eeg signal, Journal of Medical and Biological Engineering 38 (2) (2018) 222–235. doi:10.1007/s40846-017-0286-5.
URL https://doi.org/10.1007/s40846-017-0286-5

[16] G. Chen, W. Xie, T. D. Bui, A. Krzyżak, Automatic epileptic seizure detection in eeg using nonsubsampled wavelet–fourier features, Journal of Medical and Biological Engineering 37 (1) (2017) 123–131. doi:10.1007/s40846-016-0214-0.
URL https://doi.org/10.1007/s40846-016-0214-0

[17] X. Wei, L. Zhou, Z. Chen, L. Zhang, Y. Zhou, Automatic seizure detection using three-dimensional cnn based on multi-channel eeg, BMC Medical Informatics and Decision Making 18. doi:10.1186/s12911-018-0693-8.

[18] R. Sharma, R. B. Pachori, U. R. Acharya, Application of entropy measures on intrinsic mode functions for the automated identification of focal electroencephalogram signals, Entropy 17 (2) (2015) 669–691. doi:10.3390/e17020669.
URL http://www.mdpi.com/1099-4300/17/2/669

[19] D. Mateos, J. M. Diaz, P. W. Lamberti, Permutation entropy applied to the characterization of the clinical evolution of epileptic patients under pharmacological-

treatment, Entropy 16 (11) (2014) 5668–5676. doi:10.3390/e16115668.
URL http://www.mdpi.com/1099-4300/16/11/5668

[20] R. Sharma, R. B. Pachori, U. R. Acharya, An integrated index for the identification of focal electroencephalogram signals using discrete wavelet transform and entropy measures, Entropy 17 (8) (2015) 5218–5240. doi:10.3390/e17085218.
URL http://www.mdpi.com/1099-4300/17/8/5218

[21] N. Mammone, J. Duun-Henriksen, T. W. Kjaer, F. C. Morabito, Differentiating interictal and ictal states in childhood absence epilepsy through permutation rnyi entropy, Entropy 17 (7) (2015) 4627–4643. doi:10.3390/e17074627.
URL http://www.mdpi.com/1099-4300/17/7/4627

[22] G. Zhu, Y. Li, P. Paul Wen, S. Wang, M. Xi, Epileptogenic focus detection in intracranial eeg based on delay permutation entropy, AIP Conference Proceedings 1559 (2013) 31–36. doi:10.1063/1.4824993.

[23] M. Kumar, R. B. Pachori, U. Rajendra Acharya, An efficient automated technique for cad diagnosis using flexible analytic wavelet transform and entropy features extracted from hrv signals, Expert Syst. Appl. 63 (C) (2016) 165–172. doi:10.1016/j.eswa.2016.06.038.
URL https://doi.org/10.1016/j.eswa.2016.06.038

[24] A. Bhattacharyya, R. B. Pachori, U. R. Acharya, Tunable-q wavelet transform based multivariate sub-band fuzzy entropy with application to focal eeg signal analysis, Entropy 19 (3). doi:10.3390/e19030099.
URL http://www.mdpi.com/1099-4300/19/3/99

[25] A. Bhattacharyya, R. B. Pachori, A. Upadhyay, U. R. Acharya, Tunable-q wavelet transform based multiscale entropy measure for automated classification of epileptic eeg signals, Applied Sciences 7 (4). doi:10.3390/app7040385.
URL http://www.mdpi.com/2076-3417/7/4/385

[26] R. Mahajan, B. I. Morshed, Unsupervised eye blink artifact denoising of eeg data with modified multiscale sample entropy, kurtosis, and wavelet-ica, IEEE Journal of Biomedical and Health Informatics 19 (1) (2015) 158–165. doi:10.1109/JBHI.2014.2333010.

[27] C. Varon, A. Caicedo, D. Testelmans, B. Buyse, S. V. Huffel, A novel algorithm for the automatic detection of sleep apnea from single-lead ecg, IEEE Transactions on Biomedical Engineering 62 (9) (2015) 2269–2278. doi:10.1109/TBME.2015.2422378.

[28] S. L. Oh, Y. Hagiwara, U. Raghavendra, R. Yuvaraj, N. Arunkumar, M. Murugappan, U. R. Acharya, A deep learning approach for parkinson's disease diagnosis from eeg signals, Neural Computing and Applicationsdoi:10.1007/s00521-018-3689-5.
URL https://doi.org/10.1007/s00521-018-3689-5

[29] W. H. Press, S. A. Teukolsky, W. T. Vetterling, B. P. Flannery, Numerical Recipes in C (2Nd Ed.): The Art of Scientific Computing, Cambridge University Press, New York, NY, USA, 1992.

[30] E. Rdel, Fisher, r. a.: Statistical methods for research workers, 14. aufl., oliver & boyd, edinburgh, london 1970. XIII, 362 s., 12 abb., 74 tab., 40 s, Biometrische Zeitschrift 13 (6) (1971) 429–430. doi:10.1002/bimj.19710130623.
URL https://doi.org/10.1002%2Fbimj.19710130623

[31] M. G. Kendall, A. Stuart, J. K. Ord (Eds.), Kendall's Advanced Theory of Statistics, Oxford University Press, Inc., New York, NY, USA, 1987.

[32] M. Sharma, D. Goyal, A. Pv, U. R. Acharya, An accurate sleep stages classification system using a new class of optimally time-frequency localized three-band wavelet filter bank, Computers in Biology and Medicine 98. doi:10.1016/j.compbiomed.2018.04.025.

[33] M. Sharma, S. Agarwal, U. R. Acharya, Application of new class of antisymmetric wavelet filter banks for obstructive sleep apnea diagnosis using ecg sig-

nals, Computers in Biology and Medicine 100. doi:10.1016/j.compbiomed.2018.06.011.

[34] M. Sharma, A. Bhurane, U. R. Acharya, Mmsfl-owfb: A novel class of orthogonal wavelet filters for epileptic seizure detection, Knowledge-Based Systemsdoi:10.1016/j.knosys.2018.07.019.

[35] M. Sharma, A. Pv, D. Deb, S. D. Puthankattil, U. R. Acharya, An automated diagnosis of depression using three-channel bandwidth-duration localized wavelet filter bank with eeg signals, Cognitive Systems Research 52. doi:10.1016/j.cogsys.2018.07.010.

[36] A. Nishad, A. Upadhyay, R. Pachori, U. R. Acharya, Automated classification of hand movements using tunable-q wavelet transform based filter-bank with surface electromyogram signals, Future Generation Computer Systems 93. doi:10.1016/j.future.2018.10.005.

[37] A. Bhurane, M. Sharma, R. S. Tan, U. R. Acharya, An efficient detection of congestive heart failure using frequency localized filter banks for the diagnosis with ecg signals, Cognitive Systems Researchdoi:10.1016/j.cogsys.2018.12.017.

[38] M. Sharma, R. S. Tan, U. R. Acharya, A novel automated diagnostic system for classification of myocardial infarction ecg signals using an optimal biorthogonal filter bank, Computers in Biology and Medicine 102. doi:10.1016/j.compbiomed.2018.07.005.

[39] A. Rnyi, On measures of entropy and information, in: Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics, University of California Press, Berkeley, Calif., 1961, pp. 547–561.
URL https://projecteuclid.org/euclid.bsmsp/1200512181

[40] A. Kraskov, H. Stögbauer, P. Grassberger, Estimating mutual information, Phys.

Rev. E 69 (2004) 066138. doi:10.1103/PhysRevE.69.066138.
URL https://link.aps.org/doi/10.1103/PhysRevE.69.066138

[41] R. R. Coifman, M. V. Wickerhauser, Entropy-based algorithms for best basis selection, IEEE Trans. Inf. Theor. 38 (2) (2006) 713–718. doi:10.1109/18.119732.
URL http://dx.doi.org/10.1109/18.119732

[42] V. Kecman, Learning and Soft Computing: Support Vector Machines, Neural Networks, and Fuzzy Logic Models, MIT Press, Cambridge, MA, USA, 2001.

[43] J. A. K. Suykens, J. Vandewalle, Least squares support vector machine classifiers, Neural Process. Lett. 9 (3) (1999) 293–300. doi:10.1023/A:1018628609742.
URL https://doi.org/10.1023/A:1018628609742

[44] N. Cristianini, J. Shawe-Taylor, An Introduction to Support Vector Machines: And Other Kernel-based Learning Methods, Cambridge University Press, New York, NY, USA, 2000.

[45] R. Kohavi, A study of cross-validation and bootstrap for accuracy estimation and model selection, Morgan Kaufmann, 1995, pp. 1137–1143.

[46] T. Fushiki, Estimation of prediction error by using k-fold cross-validation, Statistics and Computing 21 (2) (2011) 137–146. doi:10.1007/s11222-009-9153-8.
URL https://doi.org/10.1007/s11222-009-9153-8

[47] T. Saito, M. Rehmsmeier, The precision-recall plot is more informative than the roc plot when evaluating binary classifiers on imbalanced datasets, PLOS ONE 10 (3) (2015) 1–21. doi:10.1371/journal.pone.0118432.
URL https://doi.org/10.1371/journal.pone.0118432

[48] H. Liu, H. Motoda, Feature Selection for Knowledge Discovery and Data Mining, Kluwer Academic Publishers, Norwell, MA, USA, 1998.

[49] D. Ross, U. Scherf, M. B. Eisen, C. M. Perou, C. Rees, P. Spellman, V. Iyer, S. S. Jeffrey, M. Van de Rijn, M. Waltham, A. Pergamenschikov, J. C.F. Lee, D. Lashkari, D. Shalon, T. Myers, J. N. Weinstein, D. Botstein, P. Brown, Systematic variation in gene expression patterns in human cancer cell lines, Nature genetics 24 (2000) 227–35. doi:10.1038/73432.

[50] S. Theodoridis, Pattern recognition, Academic Press, Burlington, MA London, 2009.