Article

A Sample-Encoding Generalization of the Kohonen Associative Memory and Application to Knee Kinematic Data Representation and Pathology Classification

Badreddine Ben Nouma [1], Amar Mitiche [1] and Neila Mezghani [2,3,*]

[1] INRS-Énergie Matériaux et TéLécommunications, Montreal, QC J3X 1S2, Canada; Bennouma@emt.inrs.ca (B.B.N.); mitiche@emt.inrs.ca (A.M.)

[2] Centre de Recherche LICEF, TELUQ University, Montreal, QC H2T 3E4, Canada

[3] Laboratoire de Recherche en Imagerie et Orthopédie, Montreal, QC H2X 0A9, Canada

[*] Correspondence: neila.mezghani@teluq.ca

Abstract: Knee kinematic data consist of a small sample of high-dimensional vectors recording repeated measurements of the temporal variation of each of the three fundamental angles of knee three-dimensional rotation during a walking cycle. In applications such as knee pathology classification, the notorious problems of high-dimensionality (the curse of dimensionality), high intra-class variability, and inter-class similarity make this data generally difficult to interpret. In the face of these difficulties, the purpose of this study is to investigate knee kinematic data classification by a Kohonen neural network generalized to encode samples of multidimensional data vectors rather than single such vectors as in the standard network. The network training algorithm and its ensuing classification function both use the Hotelling T^2 statistic to evaluate the underlying sample similarity, thus affording efficient use of training data for network development and robust classification of observed data. Applied to knee osteoarthritis pathology discrimination, namely the femoro-rotulian (FR) and femoro-tibial (FT) categories, the scheme improves on the state-of-the-art methods.

1. Introduction

Frequent knee pain affects approximately one adult in four, limiting function and diminishing quality of life. Knee pain in people 50 years or older is predominantly caused by osteoarthritis (OA) and it is a major reason for knee replacements among knee osteoarthritis patients in general [1,2]. This severe impact on human health and the soaring financial cost justify the recent accrued research interest in computer-aided, objective knee disease diagnosis methods. Such methods would facilitate diagnosis and improve its accuracy so that the disease can be treated more effectively. Several studies have addressed the problem of distinguishing asymptomatic and OA groups [3–7] and assessing the severity of the OA disease according to the Kellgren Lawrence (KL) score [8]. However, none has considered distinguishing two classes of knee OA pathologies, namely femero-rotulian (FR) and femero-tibilal (FT), or further consider, in addition to FR and FT, category FR-FT representing the incidence of both diseases FR and FT in a same individual.

Currently, three-dimensional (3D) knee kinematic data, which can be easily acquired in clinical settings [9], is the foremost, most effective description of knee movement to develop a classification algorithm, an essential component in objective, computer-aided, knee pathology diagnosis [3,10].

Knee kinematic data measurements consist of three high-dimensional vectors that describe the temporal variation during a full gait cycle of locomotion of the three fundamental angles of knee rotation, namely the knee angles with respect to the sagittal, frontal, and transverse planes (Figure 1). The curse of dimensionality [11], the high intra-class variability and inter-class similarity in applications such as osteoarthritis pathology classification, make knee kinematic data categorization difficult [12].

These are the angles in three-dimensional (3D) space between the tibia and femur, corresponding to flexion/extension in the sagittal plane, abduction/adduction in the frontal plane, and internal-external rotation in the transverse plane. To measure and record these angles, the participant walks at a self-paced, comfortable speed on a conventional treadmill with the non-invasive knee attachment of the KneeKG system [13]. The setup is illustrated in Figure 2. The device is first calibrated to define the origin and axes of the 3D Cartesian reference system of the knee angle coordinates. The measurements produce three discrete kinematic curves, one for each angle. Curves are normalized by resampling to some fixed number of equally spaced points [9], one hundred in this study.
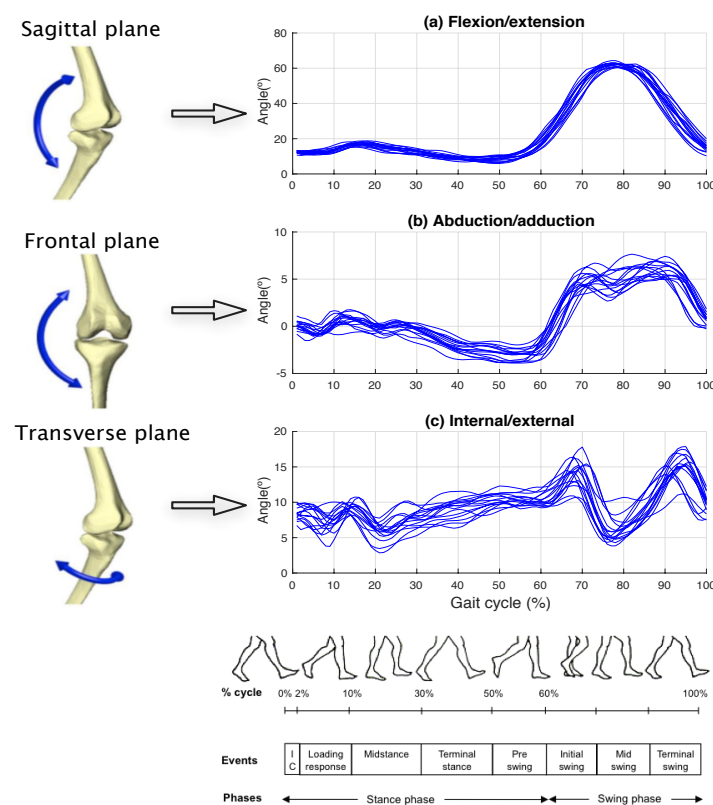

Figure 1. A family of twenty knee kinematic data curves measured for a particular participant: (**a**) flexion/extension, (**b**) abduction/adduction, and (**c**) internal/external rotation. Each curve was interpolated and re-sampled from 1% to 100% (100 points) of the gait cycle. Moreover, 1% corresponds to the initial contact (IC) and 100% to the end of the swing phase.

Figure 2. The KneeKG system.

Although the KneeKG system is accurate, a participant knee angle variation pattern varies in general, sometimes significantly, from one cycle to another during locomotion, due to the inherently uneven nature of a walker's cadence. However, current studies implicitly attribute these variations to noise, of the same structure across knee pathologies and individuals, and which does not, therefore, inform gait classification. As a result, an individual's measurements are repeated a few times, typically for ten to fifteen times, and the average of these is taken to be the individual's knee movement representative data for subsequent classification. However, summarizing a population by its average removes information about spread that might be essential to classification. Moreover, the availability of a sample of several pattern measurements for each individual opens up the unique opportunity to use statistical inference and apply potent statistical tests of hypothesis to measure similarity between class data and, consequently, determine the class of membership of a measurement or observation [14]. Such effective tests would not be applicable otherwise.

In this study, we investigate a Kohonen neural network generalized to encode data in the form of samples, which we apply to knee OA pathology categorization using knee kinematic data samples. Datasets of knee kinematic measurements are generally small, containing samples from typically fewer than one hundred subjects, each sample composed of about a dozen kinematic curves as described earlier. Kohonen neural networks, which are potent classifiers [15–20], are particularly apposite for such small-sized training data sets because they are associative memories that represent classes each by several patterns which training determines, and do so such that neighboring representations correspond to neighboring classes. Sample similarity evaluation, which underlies both training, to determine this spatially organized layout of representation patterns, and classification, to assign a category to observations, is done using the two-sample Hotelling T^2 statistic. This will be explained further when we the describe the sample-encoding generalization of the Kohonen network in greater detail subsequently. This sample-based Kohonen neural network outperformed other classifiers in knee OA experimentation to distinguish between two types of knee osteoarthritis pathologies, namely femero-rotulian (FR) and femero-tibial (FT).

The remainder of this paper is organized as follows: Section 2.1 describes the knee kinematic data and its collection. Section 2.2 explains the sample-encoding Kohonen associative memory, expounding its structure, the Hotelling T^2 statistic and its use to measure similarity between pattern samples, training, and classification role. Section 3 describes the experimental results to classify FR knee pathology versus FT, as well as three-class problems involving classes FR and FT as well as class FR-FT representing the joint occurrence of pathologies FR and FT in the same individual. Finally, Section 4 contains a discussion of the results and the last Section contains a conclusion.

2. Materials and Methods

2.1. Knee Kinematic Data Collection

Knee kinematic data describe the temporal variations of the knee movement three rotation angles during a walking cycle. Participants walk on a conventional treadmill at a self-paced, comfortable speed and the three angles of knee rotation are recorded by the KneeKG system using a non-invasive knee attachment [13]. The device is first calibrated with respect to the reference points and axes which serve to measure the three angles. The three angles of knee rotation are then recorded as the walk progresses for a full cycle. Each resulting discrete curve is normalized by a smooth fit of its points followed by resampling to some number of equally spaced gait cycle percentage points [9]. As illustrated in Figure 1, 1% corresponds to the initial contact and 100% to the end of the swing phase.

Because a person's gait varies from one cycle to another, albeit slightly, the kinematic curves are produced several times, typically about fifteen times, and then averaged under the informal assumption that unwanted outlying measurements are present which must be removed because they adversely affect classification. As a result, current methods take the average curve to be the participant's *representative curve* in subsequent analysis and classification. In this study, all of a participant's curves

are retained and used together as a *sample* rather than reducing them to a single curve of representation (Figure 1a–c) because such a reduction suppresses information that might be relevant to classification.

The dataset contains data from 21 patients of each class, FR, FT, and FR-FT. The demographic characteristics of the data in the three classes are shown in Table 1.

Table 1. Demographic characteristics of the data in the three classes (columns FR, FT, and FR-FT)

Characteristics	C_1:FR	C_2:FT	C_3:FR-FT
Age (years)	46.1 * $\pm$ 11.7	59.5 * $\pm$ 10.1	59.6 $\pm$ 11.4
Height (m)	1.71 $\pm$ 0.07	1.66 $\pm$ 0.09	1.66 $\pm$ 0.11
Weight (kg)	82.9 $\pm$ 20.7	76.2 $\pm$ 11.2	84.3 $\pm$ 15.9
BMI (kg/m^2)	28.3 $\pm$ 7.1	27.4 $\pm$ 3.9	30.3 $\pm$ 5.5
Men%	45	38	33.3

2.2. Classification by a Sample-Encoding Generalization of the Kohonen Neural Network

The Kohonen neural network is organized into an array of nodes, generally two-dimensional as illustrated in Figure 3. The purpose of the original network conception [21] was to materialize the development and function of an associative memory which runs an unsupervised algorithm to encode its input *sequentially* in the form of weight vectors, of the same data type as the input, which it stores in the nodes.
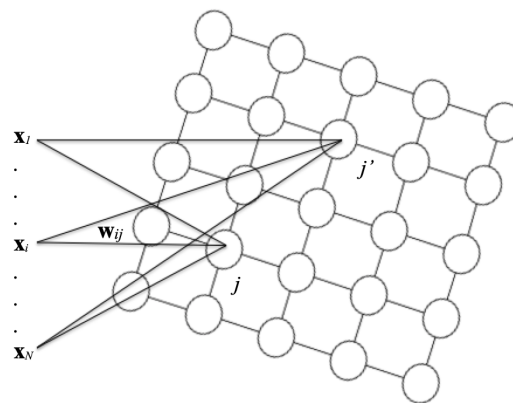

Figure 3. A 5 $\times$ 5 Kohonen neural network.

The network is said to be *topologically ordered* because the encoding is realized in such a way that neighboring nodes have neighboring values. The network is also called a *self organizing map*, abbreviated SOM, because of this topologically-ordered encoding capacity. In the practice of pattern classification, mapping labeled data into a Kohonen neural network, once it is topologically ordered and its weights settle, affords a class label to each node and, therefore, provides the network with a classification function as an associative memory: to an input pattern, it associates the class label of the node with the closest weight. The Kohonen neural network can be looked at as a vector quantizer [22,23] for its ability to reduce a data set to a group of representation prototypes. Several variants of a standard Kohonen network training algorithm have been investigated, such as the the Gibbs density modelling network [24], the probabilistic self-organizing map (PRSOM) [25], and the soft topographic mapping with kernels (STMK) [26].

In this study, we investigate a generalization of the standard Kohonen neural network algorithm that encodes input in the form of a sample of pattern characteristic vectors rather than a single such vector as with the standard algorithm. Pattern similarity, which underlies both network training and the network subsequent classification function, is defined in this generalization by the two-sample Hotelling T^2 statistic as presented next.

2.2.1. Sample Similarity and the Two-Sample Hotelling T^2 Statistic

Let $\mathbf{X} = \{\mathbf{x}_1, ..., \mathbf{x}_N\}$ and $\mathbf{W} = \{\mathbf{w}_1, ..., \mathbf{w}_M\}$ be samples of independent realizations of two D-variate multinomial random variables of equal covariance matrices and with means μ_X and μ_W. Let $\bar{\mathbf{X}}$ and $\bar{\mathbf{W}}$ be the sample means of $\mathbf{X}$ and $\mathbf{W}$, respectively:

$$\bar{\mathbf{X}} = \frac{1}{N}\sum_{i=1}^{n}\mathbf{x}_i; \quad \bar{\mathbf{W}} = \frac{1}{M}\sum_{i=1}^{m}\mathbf{w}_i, \tag{1}$$

and $\mathbf{C}_X, \mathbf{C}_W$ the sample covariances:

$$\mathbf{C_X} = \frac{1}{N-1}\sum_{i=1}^{N}(\mathbf{x}_i - \bar{\mathbf{X}})(\mathbf{x}_i - \bar{\mathbf{X}})^T,$$
$$\mathbf{C_W} = \frac{1}{M-1}\sum_{i=1}^{M}(\mathbf{w}_i - \bar{\mathbf{W}})(\mathbf{w}_i - \bar{\mathbf{W}})^T. \tag{2}$$

Finally, let $\mathbf{C}$ be the pooled (combined) covariance estimate of $\mathbf{X}, \mathbf{W}$ given by:

$$\mathbf{C} = \frac{(N-1)\mathbf{C_X} + (M-1)\mathbf{C_W}}{N+M-2}. \tag{3}$$

The two-sample Hotelling T^2 statistic is defined by [27]:

$$T^2 = \frac{NM}{N+M}(\bar{\mathbf{X}} - \bar{\mathbf{W}})^T\mathbf{C}^{-1}(\bar{\mathbf{X}} - \bar{\mathbf{W}}). \tag{4}$$

This statistic is ordinarily used in statistical hypothesis testing to test the null hypothesis $H_0 : \mu_X = \mu_W$ against the hypothesis $H_1 : \mu_X \neq \mu_W$ [14]. For large samples, the distribution under the null hypothesis of the T^2 statistic is approximately the χ^2 (Chi-squared) distribution with D degrees of freedom. For small sample sizes, as in our case of knee kinematic data, it is better approximated, under the null hypothesis, by the F distribution with D degrees of freedom for the numerator, and $N + M - 1 - D$ degrees of freedom for the denominator:

$$\frac{N+M-D-1}{(N+M-2)D}T^2 \sim F(D, N+M-1-D). \tag{5}$$

The F distribution in Label (5) can be a good approximation of the T^2 statistic distribution when the dimension of the data is less than the size of the samples [28]. For high-dimensional vectors, like knee kinematic data vectors, this study is dealing with dimensionality reduction, for instance by principal component analysis (PCA) or wavelet representation, affords a means to satisfy this condition. The F distribution with 11 degrees of freedom for the numerator and denominator, close to what we have in the knee data classification application of this study.

The two-samples T^2 statistic is in a fixed positive proportion to the squared Mahalanobis distance between the two samples means [29], as evident in Equation (4). Therefore, it is a legitimate measure of similarity of two samples, particularly when it is used to determine among a set of samples the closest to a given reference sample, as it is used in the sample-encoding generalization of the Kohonen memory which we describe next.

2.2.2. Sample-Encoding Kohonen Network Algorithm

The output of the network algorithm are samples of size N, $\mathbf{W}_j = \{\mathbf{w}_{1j}, \dots \mathbf{w}_{Nj}\}$, of D-dimensional weight vectors $\mathbf{w}_{ij} = (w_{ij}^1, \dots, w_{ij}^D)$, $i = 1, \dots N$, stored at nodes $j = 1, \dots, J$. In our application, each vector $\mathbf{w}_{i,j}$, $i = 1, \dots N$, at node j, encodes a kinematic data curve and we use the network as a knee pathology classifier. The network runs an algorithm which updates its weights iteratively as

inputs, in the form of samples of multi-dimensional vectors, are sequentially presented. This algorithm can be summarized as follows:

Algorithm 1 Network algorithm

Input $\mathbf{X} = \{\mathbf{x}_1, ..., \mathbf{x}_N\}$, $i = 1, \ldots, N$
Output $\mathbf{W}_j = \{\mathbf{w}_{1j}, \ldots, \mathbf{w}_{Nj}\}$, $j = 1, \ldots, J$

- Initialize samples $\mathbf{W}_j$ stored at nodes j, $j = 1, \ldots, J$.
- Get input sample $\mathbf{X}$ and compute the similarities (the Hotelling T^2 statistics) $s_j(\mathbf{X}, \mathbf{W}_j)$ between X and samples $\mathbf{W}_j$ stored at nodes j, $j = 1, \ldots, J$.
- Determine the node with weight vector closest to input:

$$j^* = \arg\min_j s_j, \tag{6}$$

- Update samples $\mathbf{W}_j = \{\mathbf{w}_{1j}, \ldots \mathbf{w}_{Nj}\}$ stored at nodes j, $j = 1, \ldots, J$.

$$\mathbf{w}_{1j}(t+1) = \mathbf{w}_{1j}(t) + \epsilon(t)h(t)^{j,j^*}(\bar{\mathbf{X}} - \mathbf{w}_{1j}(t))$$
$$\ldots \tag{7}$$
$$\mathbf{w}_{Nj}(t+1) = \mathbf{w}_{nj}(t) + \epsilon(t)h(t)^{j,j^*}(\bar{\mathbf{X}} - \mathbf{w}_{Nj}(t)),$$

$$\epsilon(t) = \epsilon_i \left(\frac{\epsilon_f}{\epsilon_i}\right)^{\frac{t}{tmax}}, \quad \sigma(t) = \sigma_i \left(\frac{\sigma_f}{\sigma_i}\right)^{\frac{t}{tmax}}, \tag{8}$$

$$h(t)^{j,j^*} = exp - \frac{||j - j^*||^2}{2\sigma(n)^2}. \tag{9}$$

The samples of weight vectors at the network nodes are initialized randomly. They are then modified iteratively, each modification triggered by an input sample, $\mathbf{X}$. The update consists of finding the node j^* with weight closest, most similar, to the current input and modifying the weight vectors at each node j according to its grid distance from j^*. Closeness is in terms of the two-sample Hotelling T^2 statistic, as explained earlier. For multivariate data, the two-sample Hotelling T^2 statistic is proportional to the squared Mahalanobis distance between the means of the two samples. The update equations, for multivariate data, are given by Equation (7), where t designates the iteration index. Function h^{j,j^*}, given by Equation (9), defines the influence of "winning" node j^* on node j: Every vector of the sample stored at each network node j is corrected by "pulling" it toward the current input sample by an amount decreasing with increasing grid distance from node j^*. This correction also lessens in time as a function of parameter σ which decreases between initial and final values σ_i and σ_f. This is shown in Equation (1). Finally, the correction is modulated by multiplicative parameter ϵ which also decreases in time, between initial and final values ϵ_i and ϵ_f as shown in Equation (8). Parameters ϵ, σ must be set so as to obtain ordering of the weights, in the sense described earlier, and convergence to their final values. These parameters are set experimentally.

2.3. Dimensionality Reduction

As presented earlier, the two-sample Hotelling T^2 statistic defines sample similarity used by the Kohonen neural network algorithm. However, for this statistic to be applicable, the dimension of the data must be less than the size of the samples [14]. Therefore, dimensionality reduction to satisfy this requirement must precede usage of the statistic. We performed a wavelet transform [30–32], which is often used for dimensionality reduction in pattern analysis and classification [33]. A wavelet representation retains of the data wavelet decomposition coefficients only those which correspond to a predetermined energy of the transformed signal [34–36]. A significant advantage of the wavelet representation is that a decomposition depends on the data item to describe, not on other data,

in contrast to other common feature selection methods such as principal component analysis (PCA) or singular value decomposition (SVD) [37].

2.4. Evaluation of the Sample-Encoding Kohonen Network Results

In order to evaluate the performance and generalization power of the sample-encoding Kohonen network in this application, we used the leave one out cross validation (LOOCV), a scheme that is proven to be much more accurate for small size samples than split-sample validation [38]. Classification performance was evaluated in terms of the accuracy (Acc) over all test data, as well as per class. Performance is presented in the form of a confusion matrix, where each row represents the instances in a predicted class and each column represents the instances in an actual class (ground truth).

3. Results

In the following, we apply the sample-encoding Kohonen associative memory to encode knee kinematic data samples and classify knee osteoarthritis pathologies. In a first experiment, we classify femero-rotulian (FR) vs. femero-tibial (FT), in a context where a single of the two pathologies occurs in any patient. In a second experiment, we extend the application to the three-class problem involving pathology categories FR and FT, as well as category FR-FT which represents patients having both diseases FR and FT. The dataset contains data from 21 patients of each of the three classes, FR, FT, and FR-FT.

Dimensionality Reduction

Dimensionality reduction is performed using a wavelet decomposition of the kinematic data in each plane separately, namely the flexion/extension angle, with respect to the sagittal plane (Figure 4a), the abduction/adduction angle, with respect to the frontal plane (Figure 4b), and the internal/external angle, with respect to the transverse plane (Figure 4c). The dimension of the data before feature extraction is 100, corresponding to the percentage of gait cycle (1% to 100%), for each of the three knee rotation angles (Figure 4, Line 1).
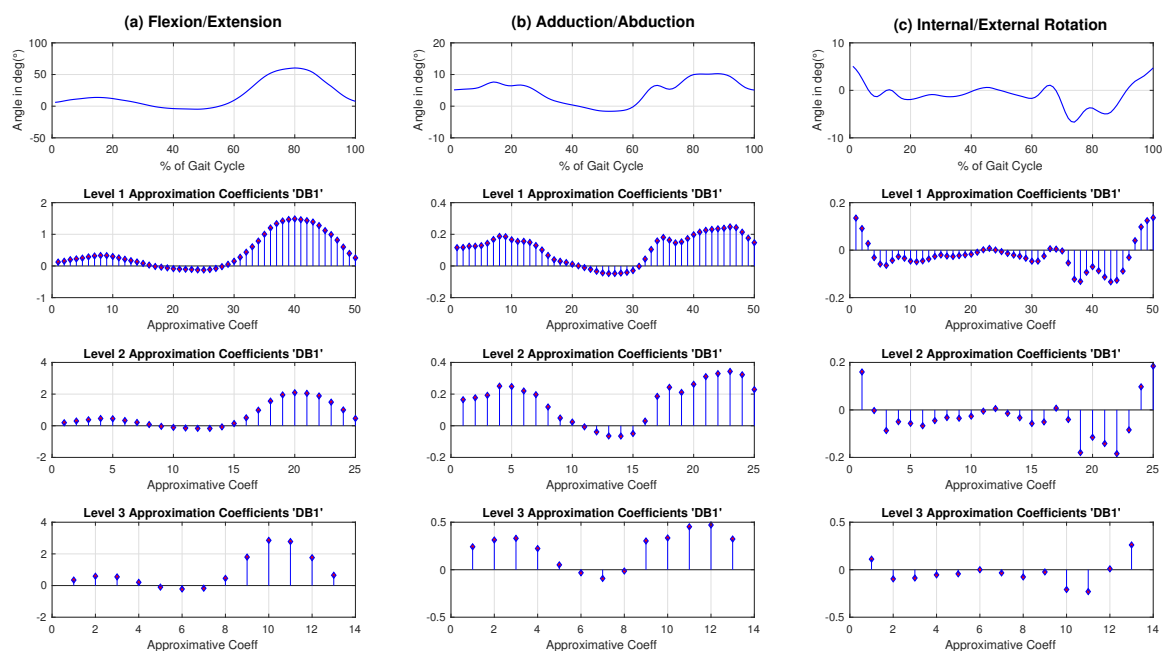

Figure 4. Wavelet decomposition using Daubechies db1 of the (**a**) the flexion/extension angle, with respect to the sagittal plane, (**b**) the abduction/adduction angle, with respect to the frontal plane, (**c**) and internal/external angle, with respect to the transverse plane. Each line corresponds to a decomposition level and each column to a kinematic plane.

Using the wavelet decomposition for dimensionality reduction, the dimension has been reduced to a fewer number of most relevant coefficients. We experimented with different wavelet families, namely Daubechies, Coiflet, and Symlet, and different levels of decomposition. The level of a wavelet representation, as well as the relevant planes of data, are chosen experimentally so as to maximize the recognition rate of the sample-encoding Kononen network. Following extensive testing, we were able to retain a subset of four coefficients of the Daubechies DB1 wavelet representation at level 3, which initially contained 13 coefficients (Figure 4, Line 4).

Sample-Encoding Kohonen Network

The Kohonen map is trained using a wavelet representation of kinematic data extracted in each plane separately (sagittal, frontal, and transverse planes). The wavelet family and the relevant planes have been evaluated by leave-one-out cross validation. This led to a data representation using the abduction/adduction and internal/external planes, and a level 6 Daubechies Db1 decomposition to four coefficients, i.e., the kinematic data is now represented by (feature) vectors of dimension 4.

The network parameters in the experiments are $\epsilon_i = 0.1$, $\epsilon_f = 0.01$, $\sigma_i = 3$, $\sigma_f = 1$. Figure 5 shows, for two pathologies classification (FR and FT), how the recognition rate varies with the number of network nodes, and with the number of the network training algorithm iterations. The best classification rate is 90.47%, obtained with an 8×8 network map (64 nodes) after 50 iterations. The corresponding confusion matrix, illustrated in Table 2, shows a balanced classification rate per class (20/21 in FR class and 18/21 in FT).

Table 2. The confusion matrix corresponding to the proposed Kohonen two class classification method. $\tau(\%)$ corresponds to the classification accuracy.

Predicted Real	C_1:FR	C_2:FT	$\tau(\%)$
C_1:FR	20	1	90.47
C_2:FT	3	18	

Recall that applicability of the Hotelling statistic requires that the dimension of the data vector space be less than the number of data vectors in the sample for which this statistic is written. In our case, a patient data sample contains between 0 to 15 vectors. Therefore, we must retain no more than nine coefficients of representation when we reduce dimensionality. The best performing set of coefficients in our dimensionality reduction experiments was of size 4. We could have safely retained up to nine coefficients. However, using more coefficients than we did does not necessarily translate to better classification. For instance, a nine-coefficient representation of the data gives a lower 88% classification accuracy.
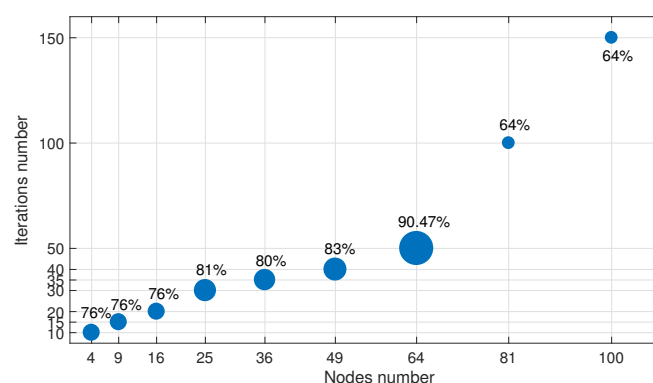

Figure 5. Variation of the recognition rate vs. the number of nodes and the number of iterations in the sample-encoding Kohonen network. The size of a circle is proportional to the classification rate it represents.

In Table 3, we present the results for the confusion matrix of the three-class classification problem. As illustrated in the table and as expected, the three-class classification problem is much more difficult than the main treated problematic. For this secondary experiment, the best achieved classification rate is 71.43 %.

Table 3. The confusion matrix corresponding to the proposed Kohonen three class classification method. $\tau(\%)$ corresponds to the classification accuracy.

Predicted Real	C_1:FR	C_2:FT	C_3:FR-FT	$\tau(\%)$
C_1:FR	18	1	2	
C_2:FT	3	14	4	71.43
C_3:FR-FT	4	4	13	

Component Planes

The sample-encoding Kohonen network training algorithm encodes nodes in such a way that neighboring nodes have neighboring weight values. In Figure 6, we visualize the weight planes, also called component planes, of each element of the input feature vector. A map node is represented by a hexagonal area. The label in each node designates the knee pathology class assigned to the node after network training (1 for FR and 2 for FT). Each sub-figure corresponds to one of the four components of the feature vector. The first and second components correspond to the wavelet representation of the abduction/adduction angles (respectively, Figure 6a,b). The third and fourth components correspond to the wavelet representation of the internal/external rotation angles (respectively, Figure 6c,d).
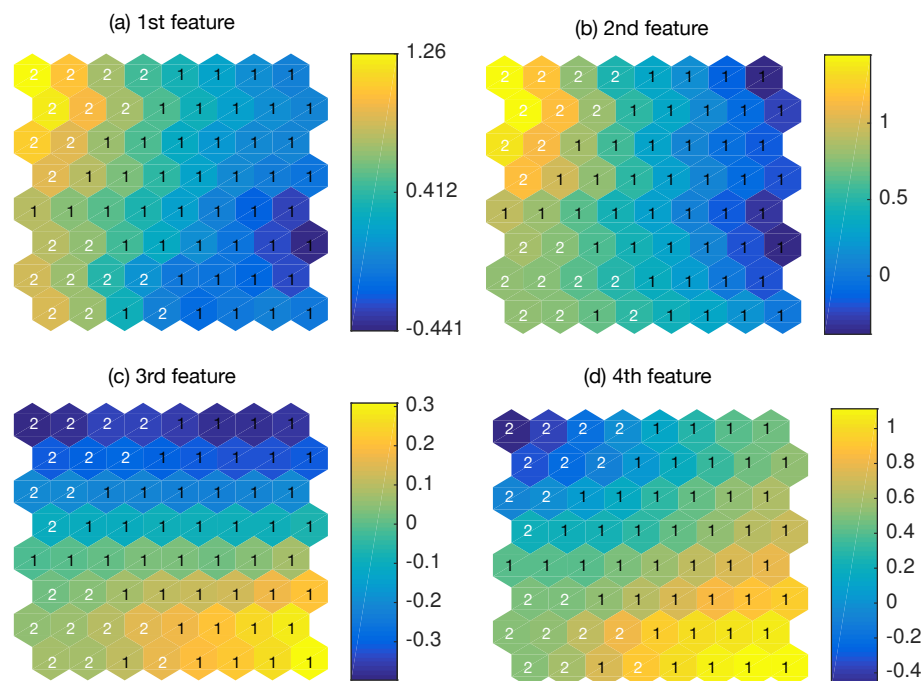

Figure 6. Feature weights visualization.

Execution Time

We measured the sample-encoding Kohonen network training and recognition times using a 2.3 GHz Intel core i7 processor with a RAM (random access memory) size of 16 Gigabytes. The network training took 25 min in a 8×8 map and 100 iterations. The classification time is negligible (0.01 s/sample).

Comparisons

Classification by the sample-encoding Kohonen network has been compared to reference classifiers used for this type of application, namely: K-nearest neighbors (KNN), support vector machine (SVM), linear discriminant analysis (LDA), Hotelling statistical hypothesis testing, and traditional Kohonen network.

Figure 7 shows the classification results of two experiments with different datasets, i.e., two classes and three classes classification.
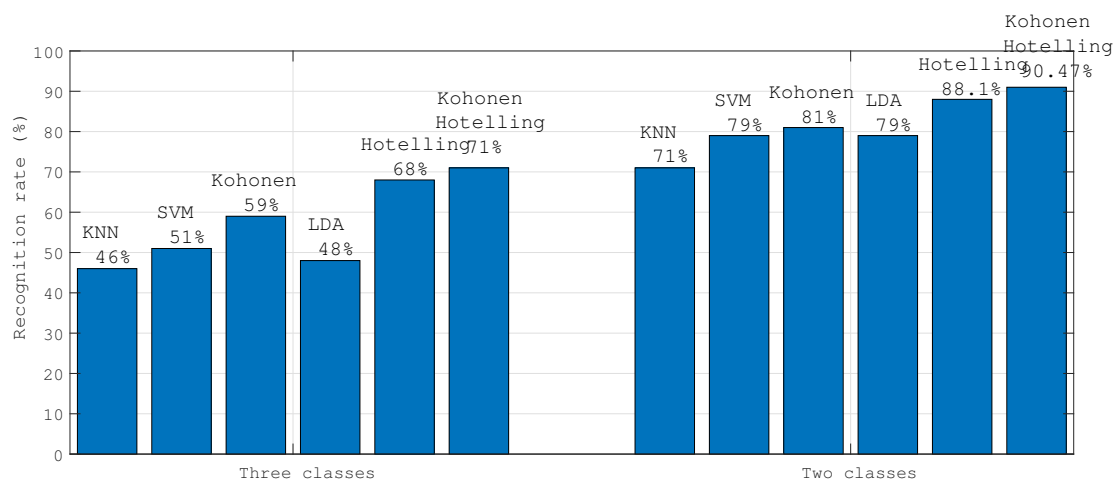

Figure 7. Comparison of the proposed sample-encoding Kohonen network method with other classifiers using a leave-one-out cross validation.

4. Discussion

The purpose of this study was to investigate a generalization of the Kohonen neural network that encodes samples of multidimensional data vectors rather than single such vectors, and apply it to knee kinematic data for osteoarthritis pathology classification. Knee kinematic data, which describe the temporal variation of each of the three fundamental angles of knee three-dimensional rotation (flexion/extension angle, with respect to the sagittal plane, abduction/adduction angle, with respect to the frontal plane, and internal/external angle, with respect to the transverse plane) during a walking cycle, are recorded in the form of a small sample of repeated measurements.

To confront the curse of dimensionality [11], the original high-dimensional kinematic data was mapped to a significantly lower dimensional space by Daubechies Db1 wavelet decomposition at level 6 to yield representation vectors of dimension 4. The training input of the sample-encoding Kohonen network consisted of this 4-dimensional representation applied to the abduction/adduction and internal/external original kinematic data. The selection of these two reference planes (discarding the third) has been determined by recognition rate maximization. This result is consistent with findings in previous studies on biomechanical data of knee pathologies. In these studies, several biomechanical parameters measured in the sagittal plane, related to the varus or valgus thrust during the loading phase, have been identified as the most useful parameters and serve diagnostic as biomarkers [8,39]. In addition, the range of motion of the abduction/adduction angle during loading phase has been identified as a component of burden of disease biomarkers to discriminate between moderate OA

grades and severe OA [8]. In addition, a study which compared a set of biomechanical parameters of patients categorized as sufferers of moderate to severe OA grades [39,40], reported that both the peak knee adduction moment and the knee adduction angular impulse increased with knee radiographic grade.

Network training and the ensuing classification function of the network both use the Hotelling T^2 statistic to evaluate the underlying similarity of pattern samples, affording robust class membership assignments to observed data. Applied to knee osteoarthritis pathology discrimination, the scheme improves on the state-of-the-art results by other methods. The classification rate reached 90.47% for the classification of FR and FT classes and 71.4% for the classification of FR, FT and FR-FT classes.

As Duda and Hart and others [11] have argued, the small size of this application dataset instructs us to use leave-one-out cross validation in the experimental evaluation of classification accuracy. There are two basic reasons for the choice of leave-one-out validation over k-fold cross validation with $k \neq n$, where n is the number of elements in the dataset (leave-one-out validation is n-fold cross validation). One obvious reason is that, while every data element serves testing once, training is done with as much data as possible, therefore using as much information about the underlying data classes as available to give a classifier more representative of the classes than it would otherwise be. This is so because when n is small and $k \neq n$, the smaller training set, due to the larger test set, is more likely to cause class information to be left out of classifier design. Another somewhat secondary reason to prefer leave-one-out validation is that proper random choice of folds for $k \neq n$ may take a great deal of computation and produce unbalanced test set sizes, causing some data elements to dominate testing and bias classification results. However, this is not a serious issue in practice because one may use pseudo-random routines, such as ones found in Matlab, that produce balanced test folds.

It may now be instructive to take a focused look at our data via an example of k-fold division and evaluation. Each item in the dataset is a sample of about a dozen (the number varies between 10 and 15) 4-dimensional vectors each containing four coefficients of a Daubechies wavelet representation of the original 300-dimensional knee rotation measurement vectors. There are 21 samples from each of two disease classes, and each sample obtained from a distinct patient. This is a small dataset. Let us use $k = 5$ folds, a sensible size which would give about 4–5 elements in each 5-fold. Following our discussion, the recognition rates should be lower than with leave-one-out validation if the folds are non redundant, i.e., if, in general, the samples left out from training to be in testing are "different" from the training fold data.

Figure 8 shows the results of the 21 5-fold cross validation experiments. Each 5-fold division was produced independently of the others. The horizontal axis lists the experiments from 1 to 21. The vertical axis unit is percentage correct classification, the star indicating the average performance in that experiment. The width of the vertical interval, centered about this mean, is twice the standard deviation of the cross validation recognition rate in the experiment. The overall average rate, computed over all the experiments, is 79%, with a standard deviation of 4.6 units of correct classification. These numbers are consistent with the expectations outlined in the discussion of k-fold validation above.
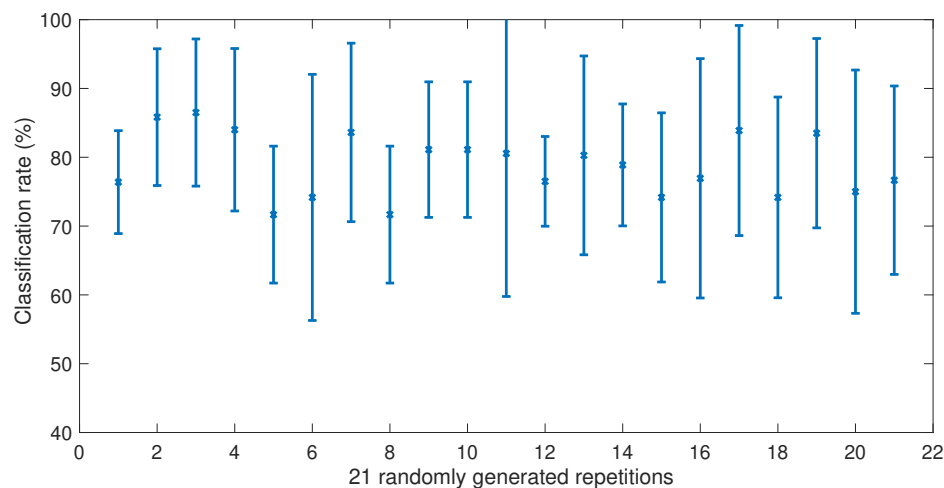
Figure 8. Five-fold cross validation experimentation. Horizontal axis: the experiments from 1 to 21. Vertical axis: the unit is percentage correct classification, the star indicating the average performance in that experiment. The width of the vertical interval about the mean is twice the standard deviation of the cross validation recognition rate in the experiment. The overall average rate, computed over all the experiments, is 79%, with a standard deviation of 4.6 units of correct classification.

We also ran a PCA plot of the original data to gain some insight into the dataset layout (Matlab *pca* and *scatter* routines). The first five coefficients of PCA account for 92% of the variance. Figure 9 shows the scatter plot for PCA coefficient pairs (1,2), (1,3), (1,4), and (1,5). These plots are sufficient to indicate that the data of the two classes (FR and FT) are neither redundant nor do they cluster away from each other, i.e., the classification problem in this application is not trivial. In addition, the spread of the data of each class that the plots show is consistent with the variations in the classification results of the 21 5-fold cross validation experiments shown in Figure 8, confirming that, in general, the test data used in an experiment contains information not present in the training data.
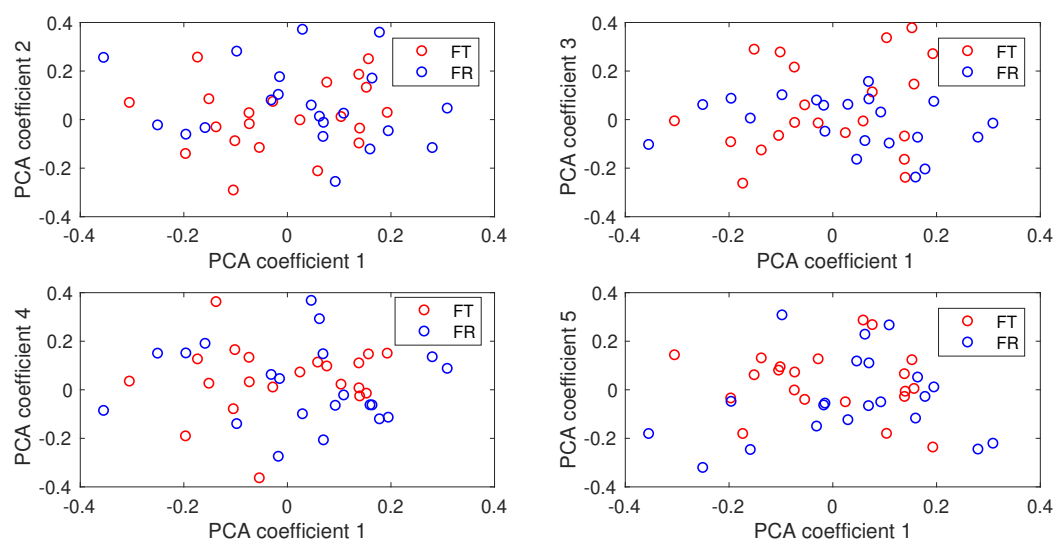

Figure 9. Scatter plots for PCA coefficient pairs (1,2), (1,3), (1,4), and (1,5) indicate that the data of the two classes (FR and FT) are neither redundant nor do they cluster away from each other.

The spatial ordering of the maps is evident in Figure 6. In this display, neighboring node values are assigned neighboring colors. In each sub-figure, neighboring node values describe the same pathology class.

We note that FR class labeled 2 is characterized by high values of the weight of the abduction/adduction angle (yellow color in Figure 6a,b) and low values of the internal/external rotation angle (blue color in Figure 6c,d). In contrast, that FT class labeled 1 is characterized by low values of the weight of the abduction/adduction angle (yellow color) and high value of the internal/external rotation angle (blue color). This is an interesting result because it correlates the kinematic data signals and their time-frequency transforms via wavelets to better understand the pathological knee behavior.

The classification time is marginal (0.01 s/sample). This can be important in computer-aided medical analysis. The training time is much longer, expectedly so, given that several parameters are to be determined experimentally. However, this is not an issue because training is done only once.

Applied to knee OA pathology classification (Formoro-tibial, Fomoro-rotulian, and co-incidence Fomoro-tibial/Fomoro-rotulian), the proposed sample-encoding Kohonen neural network outperforms K-nearest neighbor, the Support vector machine, linear discriminant analysis (LDA), and the traditional Kohonen neural network, all of which are reference classifiers usually instantiated for this type of application (Figure 7).

To conclude, the sample-encoding Kohonen neural network is efficient and can be used to represent and classify knee kinematic data, as well as other biomedical data with similar characteristics.

References

1. Nguyen, U.; Zhang, Y.; Zhu, Y.; Niu, J.; Zhang, B.; Felson, D. Increasing prevalence of knee pain and symptomatic knee osteoarthritis: Survey and cohort data. *Ann. Intern. Med.* **2011**, *155*, 725–732. [CrossRef]
2. Peat, G.; McCarney, R.; Croft, P. Knee pain and osteoarthritis in older adults: A review of community burden and current use of primary health care. *Ann. Rheum. Dis.* **2001**, *60*, 91–97. [CrossRef] [PubMed]
3. Mezghani, N.; Boivin, K.; Turcot, K.; Aissaoui, R.; Hagmeister, N.; De Guise, J. Hierarchical analysis and classification of asymptomatic and knee osteoarthritis gait patterns using a wavelet representation of kinetic data and the nearest neighbor classifier. *J. Mech. Med. Biol.* **2008**, *8*, 45–54. [CrossRef]
4. Astephen, J.L.; Deluzio, K.J.; Caldwell, G.E.; Dunbar, M.J.; Hubley-Kozey, C.L. Gait and neuromuscular pattern changes are associated with differences in knee osteoarthritis severity levels. *J. Biomech.* **2008**, *41*, 868–876. [CrossRef] [PubMed]
5. Elbaz, A.; Mor, A.; Segal, G.; Debi, R.; Shazar, N.; Herman, A. Classification of knee osteoarthritis according to spatio-temporal gait analysis. *Osteoarthr. Cartil.* **2013**, *21*, S63–S312. [CrossRef]
6. Astephen, J.L.; Deluzio, K.J.; Caldwell, G.E.; Dunbar, M.J. Biomechanical changes at the hip, knee, and ankle joints during gait are associated with knee osteoarthritis severity. *J. Orthop. Res.* **2008**, *26*, 332–341. [CrossRef]
7. Jones, L.; Holt, C.A.; Beynon, M.J. Reduction, classification and ranking of motion analysis data: An application to osteoarthritic and normal knee function data. *Comput. Methods Biomech. Biomed. Eng.* **2008**, *11*, 31–40. [CrossRef]
8. Mezghani, N.; Ouakrim, Y.; Fuentes, A.; Mitiche, A.; Hagemeister, N.; Vendittoli, P.A.; de Guise, J.A. Mechanical biomarkers of medial compartment knee osteoarthritis diagnosis and severity grading: Discovery phase. *J. Biomech.* **2017**, *52*, 106–112. [CrossRef]
9. Hagemeister, N.; Parent, G.; Van de Putte, M.; St-Onge, N.; Duval, N.; De Guise, J. A reproducible method for studying three-dimensional knee kinematics. *J. Biomech.* **2005**, *38*, 1926–1931. [CrossRef] [PubMed]
10. Ben Nouma, B.; Mezghani, N.; Mitiche, A.; Ouakrim, Y. A variational method to determine the most representative shape of a set of curves and its application to knee kinematic data for pathology classification. *MedPRAI* **2018**, 22–26. [CrossRef]

11. Duda, R.; Hart, P.; Stork, D. *Pattern Classification*; A Wiley-Interscience Publication; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2012.

12. Mezghani, N.; Ouakrim, Y.; Fuentes, A.; Mitiche, A.; Hagemeister, N.; Vendittoli, P.; De Guise, J. An Analysis of 3D Knee Kinematic Data Complexity in Knee Osteoarthritis and Asymptomatic Controls. *PLoS ONE* **2018**, *13*, e0202348. [CrossRef] [PubMed]

13. Lustig, S.; Magnussen, R.; Cheze, L.; Neyret, P. The KneeKG system: A review of the literature. *Knee Surg. Sport Traumatol. Arthrosc.* **2012**, *20*, 633–638. [CrossRef]

14. Härdle, W.; Simar, L. *Applied Multivariate Statistical Analysis*; Springer: Berlin/Heidelberg, Germany, 2007.

15. Kaski, S. Data exploration using self-organizing maps. In *Acta Polytechnica Scandinavica: Mathematics, Computing and Management in Engineering Series No. 82*; 1997.

16. Ritter, H.; Martinetz, T.; Schulten, K. *Neural Computation and Self-Organizing Maps: An Introduction*; Addison-Wesley Publishing Company: Boston, MA, USA, 1992.

17. Mitiche, A.; Aggarwal, J. Pattern category assignment by neural networks and nearest neighbors rule: A synopsis and a characterization. In *Studies In Pattern Recognition: A Memorial to the Late Professor King-Sun Fu*; World Scientific: Singapore, 1996; pp. 3–18.

18. Cottrell, M.; Rousset, P. The Kohonen algorithm: A powerful tool for analysing and representing multidimensional quantitative and qualitative data. In Proceedings of the International Work-Conference on Artificial Neural Networks, Lanzarote, Spain, 12 June 1997; pp. 861–871.

19. Deboeck, G.; Kohonen, T. (Eds.) *Visual Explorations in Finance: With Self-Organizing Maps*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2013.

20. Kaski, S.; Oja, E. *Kohonen Maps*; Elsevier Science Inc.: Amsterdam, The Netherlands, 1999.

21. Kohonen, T. *Self-Organizing Maps*, 3rd ed.; Springer: Berlin/Heidelberg, Germany, 2001.

22. Le Bail, E.; Mitiche, A. Quantification vectorielle d'images par le réseau neuronal de Kohonen. *Trait. Signal* **1989**, *6*, 529–539.

23. Nasrabadi, N.; Feng, Y. Vector quantization of images based upon the Kohonen self-organizing feature maps. In Proceedings of the IEEE 1988 International Conference on Neural Networks, San Diego, CA, USA, 24–27 July 1988; pp. 101–105.

24. Mezghani, N.; Mitiche, A.; Cheriet, M. Maximum Entropy Gibbs Density Modeling for Pattern Classification. *Entropy* **2012**, *14*, 2478–2491. [CrossRef]

25. Mezghani, N.; Mitiche, A.; Cheriet, M. Bayes Classification of Online Arabic Characters by Gibbs Modeling of Class Conditional Densities. *IEEE Trans. Pattern Anal. Mach. Intell.* **2008**, *30*, 1121–1131, doi:10.1109/TPAMI.2007.70753. [CrossRef] [PubMed]

26. González, I.; García, H. Fuzzy Labeled Self-organizing Map with Kernel-Based Topographic Map Formation. In Proceedings of the International Conference on Artificial Neural Networks, Porto, Portugal, 9–13 September 2007; pp. 341–348.

27. Hotelling, H. The generalization of Student's ratio. In *Breakthroughs in Statistics*; Springer: Berlin/Heidelberg, Germany, 1992; pp. 54–65.

28. Johnson, R.; Wichern, D. *Applied Multivariate Statistical Analysis*, 6th ed.; Prentice-Hall, Inc.: Upper Saddle River, NJ, USA, 2007.

29. McLachlan, G. Mahalanobis distance. *Resonance* **1999**, *4*, 20–26. [CrossRef]

30. Daubechies, I. *Ten Lectures on Wavelets*; CBMS-NSF Regional Conference Series in Applied Mathematics; Siam: Philadelphia, PA, USA, 1990.

31. Mallat, S. *A Wavelet Tour of Signal Processing: The Sparse Way*, 3rd ed.; Academic Press: Cambridge, MA, USA, 2008.

32. Cohen, A. Ten Lectures on Wavelets, CBMS-NSF Regional Conference Series in Applied Mathematics, Vol. 61, I. Daubechies, SIAM, 1992, Xix + 357 Pp. *J. Approx. Theory* **1994**, *78*, 460–461. [CrossRef]

33. Akanksha, N.; Mridu, S.; Shrish, V.; Raj, V. Dimensionality Reduction for Motor Imagery Signal Classification using Wavelet Analysis. *Int. J. Control Theory Appl.* **2017**, *10*, 65–76.

34. Thepade, S.; Erandole, S. Extended performance comparison of tiling based image compression using wavelet transforms amp; hybrid wavelet transforms. In Proceedings of the IEEE Conference on Information Communication Technologies, Singapore, 1–2 December 2013; pp. 1150–1155.

35. Taujuddin, N.; Ibrahim, R.; Sari, S. Wavelet Coefficients Reduction Method Based On Standard Deviation Concept For High Quality Compressed Image. *J. Theor. Appl. Inf. Technol.* **2015**, *79*, 380–388.

36. Boix, M.; Canto, B. Wavelet Transform application to the compression of images. *Math. Comput. Model.* **2010**, *52*, 1265–1270. [CrossRef]

37. Jolliffe, I.; Cadima, J. Principal component analysis: A review and recent developments. *Phil. Trans. R. Soc. A* **2016**, *374*, 20150202. [CrossRef]

38. Shao, J. Linear model selection by cross-validation. *J. Am. Stat. Assoc.* **1993**, *88*, 486–494. [CrossRef]

39. Sharma, L.; Song, J.; Felson, D.; Cahue, S.; Shamiyeh, E.; Dunlop, D. The role of knee alignment in disease progression and functional decline in knee osteoarthritis. *J. Am. Med. Assoc.* **2001**, *286*, 188–195. [CrossRef]

40. Thorp, L.; Sumner, D.; Block, J.; Moisio, K.; Shott, S.; Wimmer, M. Knee joint loading differs in individuals with mild compared with moderate medial knee osteoarthritis. *Arthritis Rheumatol.* **2006**, *54*, 3842–3849. [CrossRef] [PubMed]