UNIVERSIDAD NACIONAL DE COLOMBIA

Dynamic characterization of non-stationary signals using entropy-based relevance analysis

Juan Camilo López Montes

Universidad Nacional de Colombia
Departamento de Ingenieria Electrica y Electronica
Manizales, Colombia
Año 2019

Dynamic characterization of non-stationary signals using entropy-based relevance analysis

Juan Camilo López Montes

Tesis o trabajo de grado presentada(o) como requisito parcial para optar al título de:
Magister en Ingenieria en Automatización Industrial

Director(a):
Ph.D. German Castellanos Dominguez
Co-Director(a):
Ph.D. Oscar Cardona Morales

Línea de Investigación:
Signal Processing and Pattern Recognition
Grupo de Investigación:
Grupo de Control y Procesamiento Digital de Señales

Universidad Nacional de Colombia
Departamento de Ingenieria Electrica y Electronica
Manizales, Colombia
Año 2019

A mi familia por su apoyo incondicional y a mis parceros por ayudarme a transformar polas en algoritmos.

La preocupación por el hombre y su destino siempre debe ser el interés primordial de todo esfuerzo técnico. Nunca olvides esto entre tus diagramas y ecuaciones.

Albert Einstein

Agradecimientos

Agradezco a mi familia por su apoyo incondicional, a mi tutor Germán Castellanos Dominguez, co-tutor Oscar Cardona Morales y al Grupo de Control y Procesamiento Digital de Seãles por proporcionar siempre su conocimiento, experiencia, guía, paciencia y dedicación a esta tesis. También me gustaría hacer un especial agradecimiento a David cardenas, quien me ayudó a entender y discutir varios hallazgos de este trabajo.

This research is financial supported by research project Desarrollo de un sistema de monitoreo de condición y diagnóstico de fallas en línea de sistemas de generación de energía hidroeléctrica empleando una red de sensores inalámbricos de datos de alta resolucioón funded by COLCIENCIAS.

Resumen

Hoy en dia tratar con datos de alta dimensión, que contengan señales complejas cuya información puede estar contaminada de ruido, redundancia o información irrelevante puede ser un problema recurrente en procesos como las señales de EEG o señales de vibraciones producidas por máquinas rotatoria. Es necesario, proponer una etapa de análisis de relevancia que selecciona los componentes más importantes y reduzca el alto costo computacional.
Sin embargo, todavía no existe un método estándar para seleccionar componentes relevantes que puede variar de acuerdo con el objeto de estudio, por lo tanto, para desarrollar una metodología precisa, se debe incluir información inherente y tener un entendimiento del problema.
En este trabajo, se propone una metodología para revelar patrones relevantes en las dimensiones del espacio, tiempo y frecuencia considerando el escenario del experimento.
La propuesta consiste en dos etapas una etapa de extracción de características y una etapa donde se comparan las características guiados por las etiquetas, fue puesta a prueba en las señales de vibraciones y de encefalograma (EEG) enfocado en la tarea de motor imagery (MI).
En la tarea de MI la metodología se usa para mejorar el rendimiento de clasificación, el procedimiento consiste en filtrar cada canal en diferentes bandas de frecuencia previamente establecidas para extraer un conjunto de características basado en la entropía, para ser guiadas por una medida de similitud guiada por las etiquetas, finalmente se selecciona el conjunto de componentes que mejor discrimina las clases, las componentes significativas serán en conjunto más pequeño que logre el mejor acierto de clasificación.
La propuesta consiste en dos etapas de las cuales pueden ajustarse según la tarea, la etapa de extracción de características que debe ser seleccionado por el tipo de señal, codificando las dinámicas relevantes de la señal, la etapa medida de similitud se selecciona según el tipo de característica y su dimension.
En esta tesis probamos dos diferentes configuraciones de las etapas con el fin de mostrar la versatilidad de nuestro método y revelar que configuración codifica mejor las dinámicas relevantes. La primera configuración se basa en caracterizar las bandas filtradas por la entropia de Renyi y entropia Permutation para realizar un test estadístico guiado por las clases sobre estas, revelando el nivel de significancia de cada banda. La segunda configuración caracteriza mediante la construcción un kernel basado en la cross-correlation de cada banda, para posteriormente compararla contra el kernel de etiquetas por CKA obteniendo el valor de relevancia de cada banda y canal. Las metodologías fueron probadas en tres bases de datos de BCI y comparado contra distintos métodos del estado del arte donde nuestro análisis de relevancia mejora estadísticamente la clasificación de las tareas de MI con una menor cantidad de canales y bandas de frecuencia.
En las señales de vibraciones se propone un índice de salud basado en un análisis de relevancia que revela el estado actual de la máquina rotativa e identifica de manera temprana fallos en la misma, la propuesta codifica cada registro del sensor por Ordinal Symbolic Dynamical (OSD) y a partir de este se construye una función de densidad de probabilidad (fdp) para el registro actual y un conjunto de registros sanos, finalmente se obtiene un índice de salud comparando estos por la

divergencia de Jensen. El orden del OSD debe ser ajustado de manera óptima debido a que este controla la cantidad de símbolos que pueden existir en el diccionario y por lo tanto la cantidad de bins o eventos de la fdp. Este parámetro fue ajustado según el estado del arte y teniendo en cuanto que el orden que es directamente proporcional al costo computacional. En comparación con métodos del estado arte nuestra propuesta logra identificar el inicio de degradación del rodamiento más temprano.
En general, el análisis de relevancia permite una reducción considerable de características, esto facilita la interpretación fisiológica de los experimentos y puede mejorar el rendimiento y el costo computacional de los sistemas que utilizan estas características.

Palabras clave: (relevancia espacio-espectral, entropía, divergencia, Interfaz cerebro máquina, Permutation Entropy, Entropia de Renyi, Indice de Salud, Rodamientos, Divergencia de Jensen).

Abstract

Nowadays deal with data of high dimension, that contain complex signals whose information can be contaminated noise, redundancy or irrelevant information can be a recurring problem in processes such as EEG signals or signs of vibrations produced by machines rotating basis. In order to solve this problem, you can deploy an analysis phase of relevance to select the most important components and reduce the high computational cost.
However, there is still no standard method for selecting relevant components that may vary according to the object of study, therefore, in order to develop a precise methodology, inherent information should be included and an understanding of the problem.
This thesis proposes three methodologies to reveal relevant patterns in the dimensions of space, time and frequency considering the scenario of the experiment.
The proposal consists of two stages the first is the feature extraction stage and the second stage is where the characteristics guided by the labels are compared, the methodology which was tested in the vibration and EEG signals, focused on the Motor imagery task (MI).
In the MI task the methodology is used to improve the performance of classification in tasks of Brain Computer Interface (BCI), the procedure is to filter each channel in different bands to extract a set of characteristics based on the entropy, subsequently, the characteristics of class one against class two are compared by a measure of dissimilarity, in order to select the set of components that best discriminates the classes, finally, the significant components will be smaller together that achieves the best classification success.
The proposal consists of two stages of which can be adjusted according to the task, first the method of extraction of characteristics which must be selected by the type of signal, the type of extraction selected must be able to codify the relevant dynamics of the signal, second the measure of similarity is selected according to the type of features and its dimension.
In this document we tested two different stage configurations in order to show the versatility of our method and reveal which configuration better encodes relevant dynamics. The first configuration is

based on characterizing the bands filtered by the entropy of Renyi and Ordinal symbolic Dynamics to perform a statistical test guided by the classes on these revealing the level of significance of each band. The second configuration characterized by building a kernel based on the cross-correlation of each band, comparing the array built against the kernel of labels per CKA getting the value relevance of each band and channel.
For MI task our methodology was tested in three different BCI datasets and compared against different methods of the state of the art where our relevance analysis statistically improves the classification of MI tasks with a lesser amount of channels and frequency bands.
For vibration signals, the condition of the machine must be evaluated by a health index reveals the current state of the rotary machine and identifies early failures in it, first encodes each sensor record by Ordinal Symbolic Dynamical (OSD) and is built a probability density function (FDP) based on the coding, to be compared against sound records by the Jensen divergence.
The order of the OSD must be adjusted optimally because it controls the number of symbols that can exist in the dictionary and therefore the quantity of bins or events of the FDP. This parameter was adjusted according to the state of the art and taking in as far as the order that is directly proportional to the computational cost. In comparison with state methods art our proposed approach succeeds in identifying the start of degradation of bearing more early.
In general, the relevance analysis allows a considerable reduction of characteristics, this facilitates the physiological interpretation of the experiments and can improve the performance and the computational cost of the systems that use these characteristics.

Keywords: electroencephalogram, Brain-Computer Interface, Permutation Entropy, Renyi Entropy, Health index, Rolling element bearing, Jensen divergence

Contenido

Notation

Variables and Functions

Symbols	Description
$\boldsymbol{X}$	Set of signals
N	Number of acquired EEG recordings
T	Length of records
C	Channels
B	Bands
Bw	Bandwidth
dB	Overlapping
$\widetilde{\boldsymbol{X}}$	Set of signals selected from the proposed approach
$\boldsymbol{W}$	Spatial filter matrix
$\widetilde{\widetilde{\boldsymbol{X}}}$	Set of signals selected from the proposed approach mapped by CSP
$\boldsymbol{I}$	Identity matrix
$\boldsymbol{l}$	Matrix of the class
$\boldsymbol{\Sigma}$	Covariance Matrix
K	Number of eigenvalues selected
$\boldsymbol{\xi}$	Feature vector from CSP
α	Rényi Entropy order
$\boldsymbol{\rho}$	Relevance value
$\boldsymbol{\pi}$	particular pattern
m	Permutation Entropy order
τ	Time delay
ε	Significance threshold
λ	Lasso regularization factor
$\boldsymbol{\Sigma_{xy}}$	Cross-Correlation matrix
γ	Maximum value of the cross-correlation matrix
$\boldsymbol{\kappa}$	Tag Kernel
$\boldsymbol{S}$	Symmetric Positive-Definite matrix
d	Dimension matrix
$\boldsymbol{K}$	Stein kernel
$\boldsymbol{y}$	Envelope

Symbols	Description
D	order of OSD

Mathematical Operator

Symbols	Description
$\mathbb{E}\{\}$	Expectation Operator
$\text{var}\{\cdot\}$	Variance operator
$H_\alpha\{\cdot\}$	Renyi Entropy
$PeEn\{\cdot\}$	Permutation Entropy
$d\{\cdot\}$	Distance between class-related measures of entrop
$p(.)$	Discrete probability distribution
$\|.\|_{fro}$	Frobenius Norm
$\langle .,. \rangle$	Dot product
d^2_{SD}	Stein divergence
$\mathscr{H}\{.\}$	Hilbert transform

Lista de Figuras

Lista de Tablas

1 Preliminaries

1.1. Introduction

For long time neuroscience in the field of Motor imagery (MI) and vibration analysis has focused on methods to improve the tasks of classification and detection of faults respectively. However, the signals associated with these processes present a high degree of complexity for their analysis due to the enormous amount of data to be processed that also contain noise and redundant information, due to which, analyzing these type of signals can be computationally expensive and can be an exhausting task for the experts to analyze all the records in search of relevant patterns in addition to requiring a lot of time. Therefore, different techniques have been developed to extract the relevant patterns or dynamics caused by brain activity (EEG signal) and natural wear of bearings of a rotary machine (vibration signal).

In the field of MI, the most representative feature extraction method is CSP, several authors have studied different approaches to improving the performance of the classifier among this exciting one is the selection of bands or channels or both. Common spatial patterns (CSP) provide a feature extraction procedure that reduces the influence of the most common noise present along the recordings [30]. In state of the art, several methods are based on the extraction of features CSP because it reduces the influence of artifacts such as eye and muscle movements. However, CSP features are highly sensitive and yield the frequency band selection wrongly. Therefore, either unfiltered or inappropriately filtered EEG is performed [19].

Several approaches have been employed to address the problem of correctly select a specific frequency band for each subject performing either optimization of the spectral filter applied before CSP or selecting significant features from multiple frequency bands [7]. Sub-band Common Spatial Pattern (SBCSP) try to find the frequency band with the highest performance focusing on the relationship between CSP features and the frequency [13], Filter Bank Common Spatial Pattern (FBCSP) estimate the mutual information of the CSP features in each frequency bands highlighting the most discriminatory groups [2], Discriminative Filter Bank Common Spatial Pattern (DFBCSP) selects a specific frequency band for each subject comparing the Fisher ratio of filtered EEG signals [22], Sliding Window Discriminative CSP (SWD-CSP) uses affinity propagation to detect discriminative CSP features [19], and Sparse Filter Band Common Spatial Pattern (SFBCSP) perform a sparse regression to select the most discriminant bands [31].

Recently it has popularized the use of geometry Riemannian in the field of motor imagery (MI), because it allows to compare arrays of covariance, in addition to the fact that the geometry Riemannian use the covariance matrix directly as a feature without the need for it to be manipulated, as in

the case of the CSP. In state of the art, work has been done based on the direct use of covariance matrices or their manipulation. In [27] The authors propose to use different measurements between the covariance matrix to feed a Gaussian kernel and classify with an SVM in the BCI IV IIa dataset. In [11] tested different kernel families supplied by covariance matrices and their MKL-based combinations. Other authors estimate and manipulate covariance matrices as Sub-band Common Spatial Pattern (SBCSP) try to find the frequency band with the highest performance focusing on the relationship between CSP features and the frequency [13], and Sparse Filter Band Common Spatial Pattern (SFBCSP) perform a sparse regression to select the most discriminant bands [31].

However, it must be stated that Riemannian geometry has not solved every problem of BCI and encounters itself several limitations. For example, as noticed in [29], it seems that as the number of sensors rises (and so the bigger the dimension of the covariance matrix is), the worst the accuracy becomes, this may be due to the fact that, as the size rises, more samples are needed to build non-singular covariance matrices. When nearly singular covariance matrices are produced, they cannot be efficiently handled via Riemannian geometry. In such a case, the Euclidean geometry would outperform the Riemannian geometry [28].

The state of the art methods based on Common spatial patterns (CSP) and the geometry Riemannian are sensitive to redundant and noisy information due to a large amount of EEG data.

It is necessary to propose a methodology to attack this problematic, selected the most relevant information, reducing the size of the data without losing efficiency in the classifier, and gain interpretation in the task of BCI.

In the field of fault detection in rotary machines, Several authors had been proposed filter techniques based on the kurtosis by spectral bands such as kurtogram [24] and Spectral Kurtosis (SK) [17]. In the former case, a frequency band is selected according to the highest kurtosis, but in some cases, it does not incorporate the fault information due to the noise impulsivity. In the second case, SK is more robust to impulsive noise and non-stationary vibration signals since it utilizes a time-averaging estimator to estimate at which frequencies occur transients associated to bearing faults, nevertheless, the selected band could mask information that is not related with the failure. Other authors proposed linear filters to reduce the noise such as Wiener filter and Kalman filter [10], where it is assumed a stationary signal and additive noise, but the imposed model requires several parameters that not always satisfied the bearing fault conditions. Also, it is possible to find contributions using time-frequency representations like Short-Time Fourier Transform (STFT) [23] and Wavelet Transform [14], to describe the spectral information through the time, yet it implies to establish multiple parameters to improve the correct visualization of the signal information. Although, the provided information requires an additional stage to implement a trending analysis incorporating severity or wear indicators that will enhance the interpretation of the maintenance personnel. To assess the state machine, some indicators such as Root Mean Square (RMS) and kurtosis, provide information about its current status [5]. Nonetheless, there exist other characteristics that could describe the machine trend. Randall proposes a health index based on spectral coherence called IES, which better the envelope spectrum using spectral coherence [1].

Nonetheless, the signals acquired by the vibration sensors present difficulties to detect bearing

faults like the noise, modulated, and complex signals, this is because the vibration signals have some regularity when the bearing works under normal conditions, when the bearing deteriorates due to the progression of structural defects begin to look like modulations in time, and the number of components in the spectrum increases all this makes the signal complexity increase [26].
Accordingly, this work presents three methodologies that execute the analysis of relevance, the three methodologies use the same structure a stage of feature extraction and a stage where it is relevance based on a measure of similarity guided by the classes, however the measure of feature extraction and measure of similarity changes considering information inherent to each specific problem and the structure of the method.
For all the proposed methodologies the proposed approach improves the classification task in the case of BCI and early detection of the failure in case of vibrations and improves the the computational cost of the systems using the most relevant components of the original data.
The present research work is summarized below: In the following pages of this chapter, the general and specific objectives of this research are presented. In the Chapter 2 a novel method to select the significant components is described. The proposal allows obtaining a set of spatio-spectral components with a feature selection approach based on entropy and a statistical test, where each spatio-spectral component contribution are ranked in two-class motor imagery (MI) classification task. Here, the significant components will be the smaller set that achieves the best classification accuracy. In the Chapter 3 a method for selection based on relations inter-channel EEG (cross-correlation matrix) and a metric similarity between two kernels (Centered Kernel Alignment), also as a feature we use the Stein kernel on the covariance matrices associated with the selected components to feed the classifier for the MI task. Finally in the Chapter 4 a method for the detection of failing in the rotary machines using analysis of relevance based in permutation Entropy and the Jensen-divergence, the method proposes a Health Index (HI) of the machine, that encodes current dynamics over time by revealing the most relevant dynamics for fault detection.

1.2. Objectives

1.2.1. General objective

To develop a framework for extracting a set of significant spatio-spectral components in the signals EEG and components in time for vibration signals that allow better the particular task using the inherent information of the experiment.

1.2.2. Specific objectives

- To propose a methodology that allows selecting discriminative spatio-spectral components using the information provided by the EEG recordings and their class labels, within a kernel and entropy-based feature selection approach.

- To develop a methodology for extracting spatio-spectral components that improve the classification performance reducing the dimension of the data, for the MI task.
- To propose a relevance analysis methodology that decodes distribution changes on time series based on divergence measures for improving fault diagnosis in vibration recordings.

1.3. Academic Discussion

- López-Montes, C., Cárdenas-Peña, D., and Castellanos-Dominguez, G. (2018, September). Sub Band CSP Using Spatial Entropy-Based Relevance in MI Tasks. In International Workshop on Artificial Intelligence and Pattern Recognition (pp. 334-341). Springer, Cham.
- López-Montes, C., Cárdenas-Peña, D., and Castellanos-Dominguez, G.(2018, September). Relevance analysis in spatio-spectral components based on Permutation Entropy supporting MI discrimination ITISE 2018.

2 EEG Spatio-Spectral analysis relevance using a measure based in entropy

2.1. Introduction

We propose a spatio-spectral relevance analysis, simultaneously selecting channels and bands in a supervised scheme based on information measures (entropy). The approach, termed P_λFB and H_λFB used as a measure of complexity Permutation Entropy and Rényi entropy respectively, characterizes each channel within the set of bandpass-filtered EEG recordings trough one of the measures of entropy. Then, we test whether trials from two classes have the same mean entropy, so that resulting p-values rank the spatio-spectral. According to a subject-dependent significance level, spatio-spectral components (channels in a frequency band) are selected to feed the CSP feature extraction and LDA classifier. The obtained results, it is proven that the introduced relevance analysis improves the performance and enhances the representation of the motor imagery paradigm while reducing the space of the spatio-spectral components.

2.2. Methods

2.2.1. Sub-band Common Spatial Patterns

Let $\{\boldsymbol{x}_n^c \subset \boldsymbol{X} \in \mathbb{R}^T : n \in N, c \in C\}$ be a set of N acquired EEG recordings of length T and C the number of channels to be further bandpass filtered, adjusting two main parameters for each subject: elemental bandwidth $Bw \subset F$ and their band overlapping $\delta_B \subset B$. Therefore, the following set of bandpass-filtered EEG data is obtained: $\{\tilde{\boldsymbol{x}}_{n,b}^{\boldsymbol{c}} \subset \widetilde{\boldsymbol{X}} : b \in B\}$. For the MI dataset worked on this document, as carried out in [21], we use 17 band-pass filters with a bandwidth of $Bw{=}4\,Hz$ as to cover the whole frequency EEG band $F \in \mathbb{R}^+$, ranging from 4 to 40 Hz and fixing the overlap between each other at $\delta_B{=}2\,Hz$.

In binary classification tasks, conventional CSP finds a spatial filter matrix $\boldsymbol{W}_b \in \mathbb{R}^{Cx2K}$ to linearly map the bandpass-filtered EEG data $\widetilde{\boldsymbol{X}}_b \in \mathbb{R}^{NxC}$ onto a space $\widetilde{\widetilde{\boldsymbol{X}}}_{\boldsymbol{b}} = \boldsymbol{W}_{\boldsymbol{b}}\, \widetilde{\boldsymbol{X}}_{\boldsymbol{b}}$, so that the variance of the mapped signal is maximized for one class while the variance of another class is minimized. The spatial filters $\boldsymbol{w}_b^* \in \mathbb{R}^C$ are the solution of maximizing the Rayleigh quotient:

$$\boldsymbol{w}_b^* = \max_{\boldsymbol{w}_b} \frac{\boldsymbol{w}_b^\top \boldsymbol{\Sigma}_b^- \boldsymbol{w}_b}{\boldsymbol{w}_b^\top \boldsymbol{\Sigma}_b^+ \boldsymbol{w}_b}, \text{ s.t.: } \|\boldsymbol{w}_b\|_2 = \boldsymbol{I}_C \tag{2-1}$$

where $\boldsymbol{I}_C$ is the identity matrix size CxC, and the spatial covariance matrix of the class $\boldsymbol{l}\in\{-,+\}$ is estimated as $\widehat{\boldsymbol{\Sigma}}_b=\mathbb{E}\left\{\widetilde{\boldsymbol{X}}_b^r\,\widetilde{\boldsymbol{X}}_b^{r\top}:\forall r\in N_l\right\}$, being N_l the number of trials in class $\boldsymbol{l}$. Notations $\|\cdot\|_2$ and $\mathbb{E}\{\cdot\}$ stand for ℓ_2-norm and expectation operator, respectively.
In practice, the optimization framework in Eq. (2-1) is equivalently transformed into the generalized eigenvalue problem $\widehat{\boldsymbol{\Sigma}}_b^{-}\boldsymbol{w}_b^*=\lambda\widehat{\boldsymbol{\Sigma}}_b^{+}\boldsymbol{w}_b^*$ with $\lambda\in\mathbb{R}^+$. Thus, a set of spatial filters $\boldsymbol{W}_b^*=[\boldsymbol{w}_{b,1}^*\dots\boldsymbol{w}_{b,2K}^*]$ are obtained by collecting eigenvectors that correspond to the K largest and smallest eigenvalues of the generalized eigenvalue problem. Therefore, the CSP feature vector that accounts for the bandpass filtered components (termed sub-band CSP – SBCSP) is formed as $\boldsymbol{\xi}_b=[\xi_{b,k}:k\in 2K]$ with $\boldsymbol{\xi}_b\in\mathbb{R}^{2K}$, where $\boldsymbol{\xi}_b$ across all trials is defined as follows:

$$\boldsymbol{\xi}_b(\boldsymbol{X})=\ln(\mathrm{var}\{\boldsymbol{W_b}^{*\top}\,\widetilde{\boldsymbol{X}}_b\}) \tag{2-2}$$

where $\mathrm{var}\{\cdot\}$ stands for the variance operator.

2.2.2. Improved Data-driven Sub-band Spatial Relevance

Mainly, the plain SBCSP method in Eq. (2-2) does not reflect the contribution of $2K$ selected eigenvalues to increase the discriminating power of the spatial filter designing, encouraging the use of data-driven approaches to enhance the relevance estimation of each bandwidth. Besides, the CSP-based feature extraction methods poorly behave in the presence of nonstationarity, without mentioning its degradation due to outlier and artifact trials. To overcome such an issue, we propose to select the most discriminating spatio-spectral components based on the theory of Information this chapter uses the Rényi Entropy $H_\alpha(\cdot)\in\mathbb{R}^+$ and Permutation Entropy $PeEn(\cdot)\in\mathbb{R}^+$ due to its efficiency to characterize non-stationary signals and with complex dynamic [4], instead of further using variance-based methods as a measure of uncertainty. Therefore, we introduce an entropy-based relevance criterion for enhancing the discriminative power of CSP features, extracted from b-th temporal-frequency component, termed PFB in the case of *PeEn* y HFB for H_α:

$$\boldsymbol{\rho}_{\boldsymbol{b}}^{\boldsymbol{c}}(\tilde{\boldsymbol{x}}_{n,b}^{\boldsymbol{c}}|l)=\mathrm{d}\{\mathbb{E}\left\{Feat(\tilde{\boldsymbol{x}}_{r,b}^{\boldsymbol{c}}):\forall r\in N_-\right\}-\mathbb{E}\left\{Feat(\tilde{\boldsymbol{x}}_{r,b}^{\boldsymbol{c}}):\forall r\in N_+\right\}\} \tag{2-3a}$$

$$\text{s.t.: }\boldsymbol{\rho}_{\boldsymbol{b}}^{\boldsymbol{c}}(\tilde{\boldsymbol{x}}_{n,b}^{\boldsymbol{c}}|l)\leq\varepsilon\in\mathbb{R}^+;\forall b\in B,c\in C \tag{2-3b}$$

where $\mathrm{d}\{\cdot\}\in\mathbb{R}^+$ denotes the distance between class-related measures of entropy and *Feat* can be any of the proposed uncertainty measures. Note that the smaller the value $\boldsymbol{\rho}_{\boldsymbol{b}}^{\boldsymbol{c}}(\cdot)$, the more significance the component holds. For purposes of implementation, the supervised distance $\mathrm{d}\{\cdot\}$ is estimated as the significance of statistical t-test value, being ε a minimal threshold of accepted relevance computed for any bandpass filtered component that is fixed empirically. As regards computation of the measure of uncertainty in Eq. (2-3a), we can use the Permutation Entropy or use Rényi Entropy.

2.2.3. Permutation Entropy (PeEn)

PeEn is a linear complexity measure for time series. The relationship between present values and a fixed number of equidistant values at a given past time is captured through a symbolic mapping of the continuous time series [12]. The Permutation Entropy is estimated as below:

$$PeEn(\tilde{\boldsymbol{x}}_b^{\boldsymbol{c}}) = \sum_{i=1}^{m!} p(\boldsymbol{\pi_i}) \log \frac{1}{p(\boldsymbol{\pi_i})} \tag{2-4}$$

where $\boldsymbol{p}(\boldsymbol{\pi_i})$ is the discrete probability distribution of $m!$ ordinal patterns, which is built by mapping the time series $(\tilde{x}(t)_b^c, \tilde{x}(t+1)_b^c, ..., \tilde{x}(m-1)_b^c, \tilde{x}(m)_b^c)$ into a sequence of ordinal patterns. The particular pattern $\boldsymbol{\pi_t}$ at time t is obtained form the m-tuple of values:

$$(\tilde{x}(t)_b^c, \tilde{x}(t+\tau)_b^c, \tilde{x}(t+2\tau)_b^c ..., \tilde{x}(t+(m-1)\tau)_b^c) \rightarrow \boldsymbol{\pi_t} \in \Omega_m \tag{2-5}$$

being $m \in \mathbb{N}$ the order of the PeEn and $\tau \in \mathbb{R}^+$ the time delay, and the probability distribution is computed by counting the pattern occurrences. As a result, by measuring the inter-class distance from the trial-wise Permutation entropy, we handle a better accounting for time-variant dynamics of each channel and sub-band thanks to the construction of distributions based on ordinal pattern encodings. *PeEn* measures the departure of a time series from a completely random one: the smaller the value of the PE, the more regular the time series and a high value of *PeEn* measure is obtained when the $\boldsymbol{\pi_i}$ patterns have the same probability, this usually happens when high frequencies dominate the signal, this implies that *PeEn* increases with the irregularity of the time series at slower frequencies the permutations corresponding to peaks and troughs are observed less frequently, i.e. the EEG is more regular, hence the permutations appear with different probabilities, which decreases the *PeEn* of the signal. For EEG signals $m = 3, ..., 7$ values are recommended, for the time lag, it is adequate to use a value of $\tau = 1$ to extract most of the information in the EEG [12] hence this value is commonly chosen for EEG analysis.

2.2.4. Rényi Entropy

The generalized function of Rényi entropy estimated below:

$$H_\alpha(\tilde{\boldsymbol{x}}_b^{\boldsymbol{c}}) = \frac{1}{1-\alpha} \log \sum_{\forall k} p^\alpha(\tilde{\boldsymbol{x}}_b^{\boldsymbol{c}}(t)) \tag{2-6}$$

where $p(.) \in \mathbb{R}[0,1]$ is the probability of $\tilde{\boldsymbol{x}}_b^{\boldsymbol{c}}(t)$, The density estimation used in this study is the histogram-based density estimation . For this density estimation method, the variable $\tilde{\boldsymbol{x}}_b^{\boldsymbol{c}}(t)$ is partition into $k \in \mathbb{N}^+$ bins of equal width, $\alpha \in \mathbb{R}^+$ is the entropy order ,the number of bins,k and α order a are parameter which were fixed as 128 and 10 respectively, experimentally for the three databases.

The feature set in Equation (2-2) results from the subset of channels and bands that agree the relevance criterion, $\boldsymbol{X}_\varepsilon = \left\{ \tilde{\boldsymbol{x}}_{n,b}^{\boldsymbol{c}} : \boldsymbol{\rho_b^c}(\tilde{\boldsymbol{x}}_{n,b}^{\boldsymbol{c}} | l) \leq \varepsilon \right\}$, as $\boldsymbol{\xi}_b(\boldsymbol{X}_\varepsilon)$, enhancing the class discrimination in the new feature space.

2.3. Experimental Set-Up

2.3.1. Description of Datasets and processing pipeline

The proposed approach of relevance-based spatio-spectral selection is evaluated on the three online available datasets for motor imagery brain-computer interface described below.

The BCI Competition III dataset IVa

The BCI Competition III dataset IVa[1] contains the EEG signals recorded from five subjects performing two different (right hand and left hand) MI tasks. 118 channels at the positions of the extended international $10/20$ system were used for measuring the EEG signal, sampled at a rate of 1000Hz. The dataset contains a total of 280 trials for each subject with equal number of trials for each task. The data provided were interval filtered with a passband of $0{,}05 - 200$Hz the down-sampled data at 100Hz is used. Visual cues, lasting 3,5s indicated which of the following three motor imageries the subject should perform: (L) left hand, (R) right hand, (F) right foot. The presentation of target cues were intermitted by periods of random length, 1,75 to 2,25s, in which the subject could relax.

The BCI competition IV dataset IIa

The BCI competition IV dataset IIa[2] contain a collection of EEG signals recorded using a 22-electrode montage from nine subjects that compose the dataset. Each subject performs two sessions on different days with four motor imagery tasks, namely left hand, right hand, both feet, and tongue on two sessions. The session includes six runs with twelve trials per task, obtaining 144 trials for each class. The signals were sampled with 250Hz and bandpass-filtered between 0,5Hz and 100Hz. The subjects were sitting in a comfortable armchair in front of a computer screen. At the beginning of a trial ($t = 0$s), a fixation cross appeared on the black screen. In addition, a short acoustic warning tone was presented. After two seconds (t=2s), a cue in the form of an arrow pointing either to the left, right, down or up (corresponding to one of the four classes left hand, right hand, foot or tongue) appeared and stayed on the screen for 1,25s. This prompted the subjects to perform the desired motor imagery task. No feedback was provided. The subjects were asked to carry out the motor imagery task until the fixation cross disappeared from the screen at $t = 6$s.

GIGASCIENCE

The GIGASCIENCE[3] dataset contains an EEG signal recorded conducting a BCI experiment for the motor imagery movement of the left and right hands with 52 subjects [6] of which we selected 44 subject for the current experiment because the remainder subject hold a large amount of trials

[1] http:// www.bbci.de/competition/iii/
[2] http://www.bbci.de/competition/iv/
[3] http://gigadb.org/dataset/100295

with the artifact. EEG data were collected using 64 Ag/AgCl active electrodes 64−channel montage based on the international $10-10$ system of sampling rate 512Hz. The dataset contains 100 or 120 trials for each MI class. Each trial starts with the monitor showed a black screen with a fixation cross for $t = 2$s then one of two instructions (left hand or right hand) appeared randomly on the screen for $t = 3$s, and subjects were asked to move the appropriate hand depending on the instruction given. After the movement, when the blank screen reappeared, the subject was given a break for a random 4,1 to 4,8 seconds. These processes were repeated 20 times for one class (one run), and one run was performed.

Processing pipeline

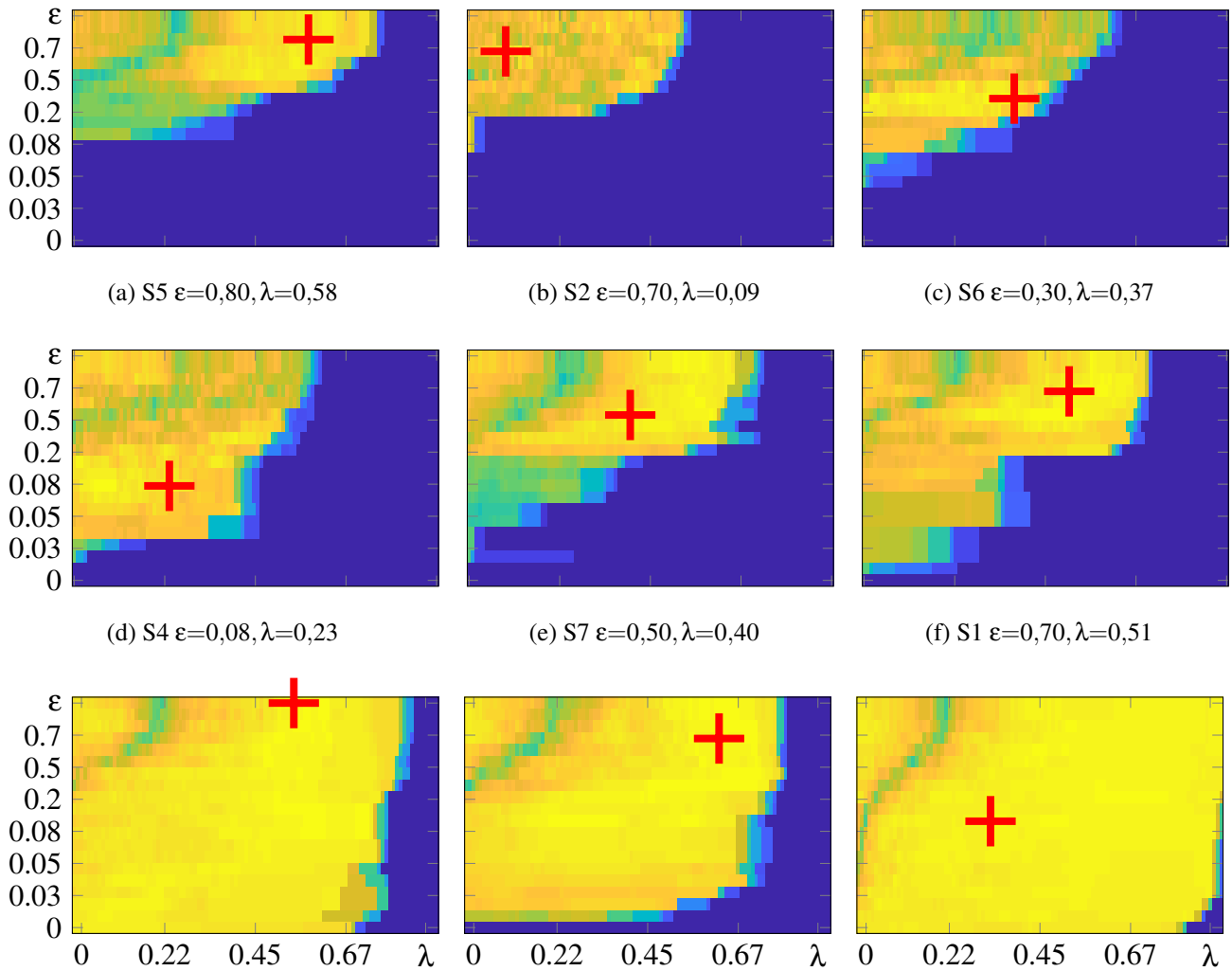
Being the most used MI paradigm in the literature, this work considers the binary classification of left and right movement for the three datasets. The trials from each subject are entered into a methodology subject-dependent BCI with four stages, channel-band selection, feature extraction, and classification. In the preprocessing stage, all EEG recordings are band-pass filtered between $[8-30]$Hz and downsampled at 100Hz using a fifth-order Butterworth filter to remove noise and slow baseline signal. We further extract the period of 2s after the cue to focus on the learning part of the MI instruction. Then, each trial is decomposed according to Section 2.2.2 and reconstructed using the resulting components from Section 2.2.4 Further, Common Spatial Patterns, one of the most efficient algorithms for extracting discriminating patterns from MI, maps each EEG covariance into a feature vector. Resulting trial-wise feature vector lastly feeds an LDA-based classifier.

2.3.2. Parameter tunning

The introduced approach for sub-band CSP using relevance based on the Rényi and Permutation Entropy (noted as HFB and PFb respectively) is validated in an example of MI tasks, employing the LDA classifier accuracy *Acc* as a measure of the provided system performance. For the sake of comparison, HFB and PFB are contrasted towards similar approaches: Filter-bank (FB) CSP [2] and Sparse Filter-bank (SFB) CSP [31] that accomplish the CSP feature selection stage using the mutual information and the Lasso regression, respectively. However, the latter approach includes a dimension reduction procedure, additionally through a Lasso regularization algorithm. Likewise, an expanded version of HFB and PFB that incorporates the Lasso regression is also considered (noted as H_λFB and P_λFB respectively).
The proposed methods H_λFB and P_λFB requires two procedures of parameter tuning; one is the significance threshold for the t-test, $\varepsilon \in (0,1]$ and another parameter is the Lasso regularization factor, $\lambda \in \mathbb{R}^+$. The first controls the amount of spatio-spectral components selected, and the second controls the number of features extracted in the frequency bands that are used for the classification. The parameters ε, λ are set up based on a grid search for the best classification accuracy.

Fig. **2-1** illustrates the performance obtained by the pair of parameters (ε ,λ), in addition a Red Cross describes the best parameter configuration for each subject in the BCI competition IV dataset

for the proposed approach H_λFB, proving the need for tuning ε, λ for each subject to properly decode the different brain activities.



(a) S5 ε=0,80, λ=0,58 (b) S2 ε=0,70, λ=0,09 (c) S6 ε=0,30, λ=0,37

(d) S4 ε=0,08, λ=0,23 (e) S7 ε=0,50, λ=0,40 (f) S1 ε=0,70, λ=0,51

(g) S3 ε=1,00, λ=0,54 (h) S8 ε=0,70, λ=0,62 (i) S9 ε=0,09, λ=0,32

Figura 2-1: Tuning curves for ε and λ. Subjects are sorted from the worst to the best per- forming in terms of the maximum achieved accuracy, the Red Cross represents the configuration of ε, λ with better performance for each subject.

It is observed that the subjects with the best success (S9, S8, and S3) have excellent performance in much of the space of parameters with respect to the rest of the subjects, note that these subjects succeed in having a success with a small ε or with fewer components to differentiate of the subject S5, S2, and S6 that is demanding a more substantial number of elements to learn the common spatial patterns. Note that the parameter tuning allows selecting a small subset of channels and sub-bands that achieves optimal performance in each subject.

Although two explicit dynamics are observed one of the useful subjects (S9, S8, and S3 and (S5,

S2, and S6) bad the set of parameters λ and ε is different for each subject, so it is shown that it is necessary to make a subject dependent analysis to encode the dynamics of each subject.

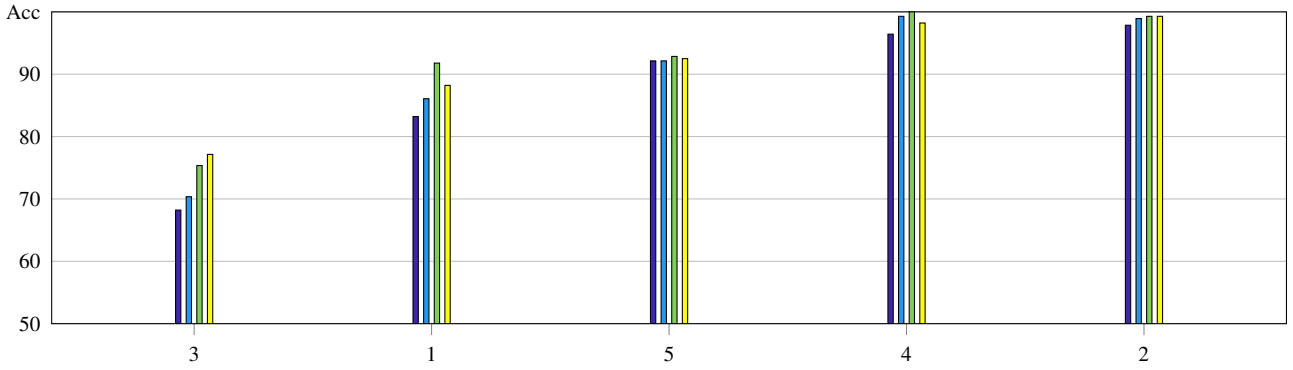
2.4. Performance results

We compare the P_λFB and H_λFB against the conventional CSP, Filter-bank (FB) CSP [2] and Sparse Filter-bank (SFB) CSP [31] methods, the Table **2-1** displays the overall five-fold cross-validation classification accuracy, component reduction rate (*crr*) and the *p*-value of the proposed approach against the competing methods for the three datasets (BCI Competition III dataset IVa, BCI Competition IV dataset IIa and GIGASCIENCE), in addition, as shown in the table add the results obtained by HFB and PFB methods, which are exactly equal to the proposed approach without including the stage of the Lasso.

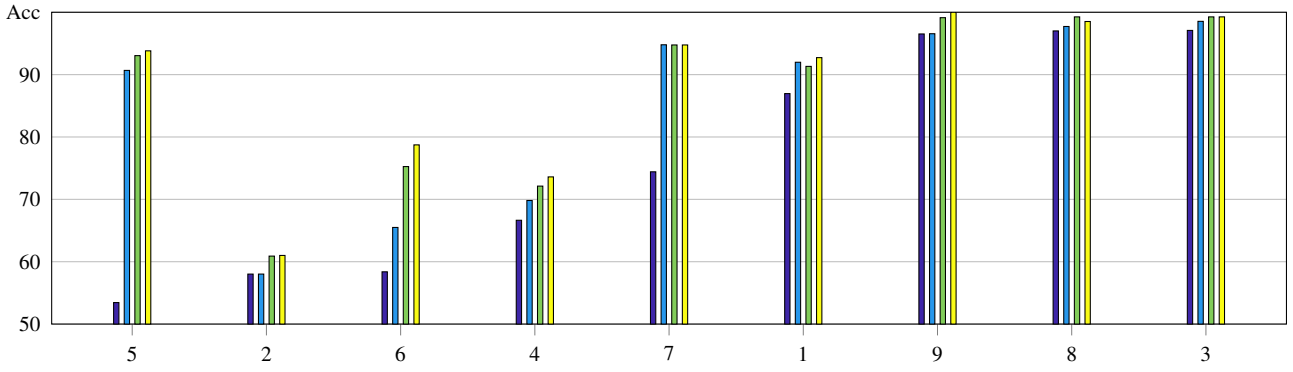
	CSP	**FBCSP**	**PFB**	**HFB**	**SFB**	**P_λFB**	**H_λFB**
				BCICIII IVa			
$\boldsymbol{\mu \pm \sigma}$	87.6 ± 4.2	80.4 ± 4.5	86.5 ± 4.9	86.1 ± 3.8	89.3 ± 2.8	$\mathbf{91.8 \pm 2.5}$	91.1 ± 3.3
p-value	$7.2e^{-4}$	$9.9e^{-7}$	$5.9e^{-4}$	$3.1e^{-5}$	$1.7e^{-2}$	$3.9e^{-1}$	-
crr	118/118	2006/2006	54/2006	352/2006	2006/2006	298/2006	881/2006
				BCICIV IIa			
$\boldsymbol{\mu \pm \sigma}$	76.5 ± 5.6	59.9 ± 9.3	81.3 ± 5.7	81.7 ± 6.4	84.8 ± 5.8	87.2 ± 4.3	$\mathbf{88.0 \pm 3.9}$
p-value	$4.6e^{-6}$	$1.3e^{-15}$	$1.0e^{-6}$	$1.0e^{-5}$	$2.8e^{-2}$	$4.1e^{-1}$	-
crr	22/22	374/374	248/374	213/374	374/374	244/374	229/374
				GIGASCIENCE			
$\boldsymbol{\mu \pm \sigma}$	68.9 ± 7.6	61.3 ± 7.3	71.0 ± 6.7	73.5 ± 6.7	72.6 ± 6.3	77.3 ± 6.5	$\mathbf{79.4 \pm 6.5}$
p-value	$1.2e^{-24}$	$2.6e^{-51}$	$7.9e^{-19}$	$1.5e^{-18}$	$2.4e^{-12}$	$1.9e^{-3}$	-
crr	64/64	1088/1088	382/1088	359/1088	1088/1088	353/1088	342/1088

Tabla 2-1: Classification accuracy for the considered approaches. Mean and standard deviation for five-fold cross-validation is computed, *p*-value is the statical test that measures significance level between H_λFB and the other methods and crr is the component rate selection.

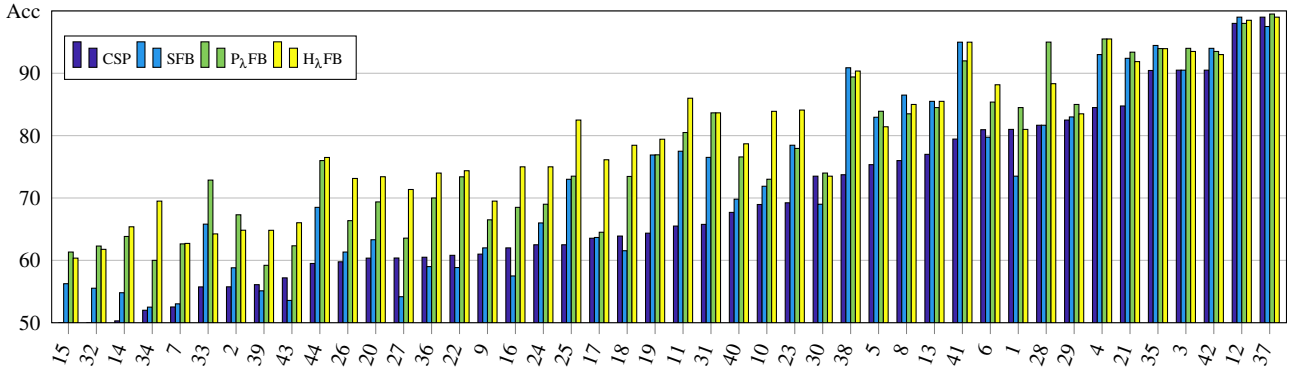
Fig. **2-2** presents the accuracy rate attained for each subject by the evaluated approaches. Note that subjects are sorted according to the CSP accuracy, aiming to highlight the performance improvement by each method. Subject-wise accuracies evidence that P_λFB and H_λFB improves the performance of the worst subjects while maintaining the accuracy of the best subjects in comparison with CSP and SFB. Regarding the other filter-banked approaches, our method outperforms by achieving the highest accuracy rates for most of the subjects in the three datasets.



(a) BCIIII IVa



(b) BCICIV IIa



(c) GIGASCIENCE

Figura 2-2: Performed accuracy by each tested approach on the three datasets. Average of a five-fold cross validation is depicted. Subjects are sorted according the performed accuracy, note that CSP is the baseline approach.

For BCI Competition III dataset IVa P$_\lambda$FB achieves the highest classification accuracy It achieves the highest classification accuracy with an average of 91,8 %, with respect to the state of the art methods such as CSP, PFB, and SFB an improvement of 4,2 %, 11,4 % and 2,5 %, respectively is

obtained, besides being more reliable due to being the method that presents less standard deviation, in addition, it is found that the proposed method is statistically significant from those of the state of the art as it is observed in the Table **2-1** in the parameter p-value, This is because the feature selection stage proposed for this task uses the most relevant and discriminating information which is only 14,8 % of the original data.

The BCI Competition IV dataset IIa, presents a bigger challenge with regard to BCI Competition III dataset IVa, because for CSP shows subjects with performance, low as the subject $S5$, $S2$ and $S6$ which are located below the 60 % Fig. **2-2**b, when applying H_λFB has obtained a success of 88,0 % that is statistically superior to CSP, PFB and SFB with an improvement of 11,5 %, 28,1 % and 3,2 % respectively, this achieves due it at the stage of feature selection reduces the components of space and frequency from 374 to 229 components, in addition, the statistical test finds that the proposed methodology (H_λFB) is statistically significant with a level of significance less than 3 % Table **2-1** (p-value).

Gigascience is the third database in which the proposed approach is tested, this database account with 44 subjects were in each one was measured their brain activity by 64 electrodes, where each channel was analyzed in frequency by a bank of 17 filters, for a total of 1088 components, which were selected 342, which is equivalent to only use a 31,4 % of the original data, the selection of these components better the success 10,5 %, 19,1 % and 6,8 % compared against CSP, PFB and SFB respectively. This database has a superior performance by subject; this can be seen in the Fig. **2-2**c, where it is observed as H_λFB considerably, improves each subject.

2.5. Discussion

This work proposes a spatio-spectral relevance analysis to select the most discriminative spatial bandpass filtered components of EEG recordings, termed HλFB and PλFB. Finding spatio-spectral elements that represent better-stimulated areas of the brain is very important to develop a highly performing BCI system. Since it is known that different parts of the brain are involved in different MI, actions, and feelings, nonetheless, addressing this problem can be very hard due to inter-subject variability. In this study, an automatic method based on spatio-spectral relevance analysis supporting EEG discrimination is presented to find the bands and channels that contain the most discriminating and relevant information for MI task.

The proposed approach holds two free parameters, namely, the significance level of the statistical test and the λ regularization that trades-off the label regression and the component sparsity. The significance level sets a threshold to reject the null hypothesis that the difference of the mean entropies is zero. Hence, the smaller the ε, the more restrictive the spatio-spectral component selection. Subsequently, λ controls the number of selected CSP features, so that large regularization values exclude useful features, while no regularization yields overfitted high dimensional spaces. Results in Fig. **2-1** reveal that each subject demands its parameter to decode the inherent subject variability. Therefore, the parameter tuning allows extracting only the most discriminative components

for each subject according to its training difficulties.
The performance of PFB and HFB is compared against the baseline CSP, FB, and SFB. The Fig. **2-2** presents the accuracy by subject and Table **2-1** summarizes the global performance achieved by each method where it is observed the statistical superiority of PFB over CSP and FB, while using less information. However SFB turns out to be superior at BCI Competition III DataSet IVa and BCI Competition IV DataSet IIa to compare with SFB, a feature reduction based on Lasso regression was included after the feature selection, termed PλFB and HλFB. Therefore, the proposed component selection in both space and frequency removes irrelevant and redundant information under the t-test criterion applied over entropy-based measures. These results reveal the importance of eliminating information that does not contribute to the task or that can be redundant and the effectiveness of entropy as a mean for the codification of the dynamics in the MI task, so validating the proposed methodology.

3 EEG Spatio-Spectral analysis relevance: A Kernel-Based supervised approach

3.1. Introduction

We propose a spatio-spectral relevance analysis simultaneously selecting channels and bands in a supervised scheme. The approach termed MFS reduces the spatial and frequency dimension by decreasing the probability of a single covariance matrix, then constructing a Stein kernel on the covariance matrices of each band, so there is no need for an extraction stage of features. Finally, you get a single kernel using MKL and is classified with SVM. Obtained results, it is proven that the introduced relevance analysis improves the performance and enhances the representation of the motor imagery paradigm while reducing the space of the spatio-spectral components.

3.2. Methods

3.2.1. Filter bank description

Let $\{\boldsymbol{x}_n^c \subset \boldsymbol{X} \in \mathbb{R}^T : n \in N, c \in C\}$ be a set of N acquired EEG recordings of length T and C the number of channels to be further bandpass filtered, adjusting two main parameters for each subject: elemental bandwidth $Bw \subset F$ and their band overlapping $\delta_B \subset B$. Therefore, the following set of bandpass-filtered EEG data is obtained: $\{\tilde{\boldsymbol{x}}_{n,b}^{\boldsymbol{c}} \subset \widetilde{\boldsymbol{X}} : b \in B\}$. For the above-described MI dataset, as carried out in [21], we use 17 band-pass filters with a bandwidth of $Bw{=}4\,Hz$ as to cover the whole frequency EEG band $F \in \mathbb{R}^+$, ranging from 4 to 40 Hz and fixing the overlap between each other at $\delta_B{=}2\,Hz$.

3.2.2. Relevance analysis based on centered kernel alignment

To observe the relevance of each spatio-spectral, a relevance index is constructed based on cross-correlation $\boldsymbol{\Sigma_{xy}}$ and Centered Kernel Alignment (CKA). First the maximum value of the cross-correlation is calculated between all the possible pairs of the records for each band and channel:

$$\boldsymbol{\gamma_{b,c}} = \max_{\tau} \left\{ \boldsymbol{\Sigma_{xy}^{\tau}} \{x_{n_i,b}^c, x_{n_j,b}^c\} \right\} \subset \boldsymbol{\Gamma} \in \mathbb{R}^{NxN} \tag{3-1}$$

Where $\boldsymbol{\Sigma_{xy}}$ is defined as:

$$\boldsymbol{\Sigma}_{\boldsymbol{xy}}^{\boldsymbol{\tau}}\{x_{n_i,b}^c, x_{n_j,b}^c\} = \frac{\mathbb{E}\left\{x_{n_i,b}^c \quad x_{n_j,b}^{c,\tau}\right\}}{\sqrt{\mathbb{E}\left\{x_{n_i,b}^c \quad x_{n_i,b}^c\right\}\mathbb{E}\left\{x_{n_j,b}^c \quad x_{n_j,b}^c\right\}}}; \forall b \in B, c \in C \tag{3-2}$$

Where $i, j \in (0,..,N)$ are subindex that index the records and $\mathbb{E}\{\cdot\}$ is the expectation operator. The maximum value of the cross-correlation $\boldsymbol{\gamma}$ is compared against the tag kernel ($\boldsymbol{\kappa} \in \mathbb{R}^{NxN}$) is taken for each band and channel.

$$\rho = CKA\{\boldsymbol{\gamma}_{\boldsymbol{b},\boldsymbol{c}}, \boldsymbol{\kappa}\} \in \mathbb{R}^{BxC} \tag{3-3}$$

Where *CKA* compare the relevance matrix of each channel in its 17 bands against the label kernel and is defined as:

$$CKA\{\boldsymbol{\gamma}_{\boldsymbol{b},\boldsymbol{c}}, \boldsymbol{\kappa}\} = \frac{\left\|\left\langle \boldsymbol{\gamma}_{\boldsymbol{b},\boldsymbol{c}}, \boldsymbol{\kappa} \right\rangle\right\|_{fro}}{\left\|\boldsymbol{\gamma}_{\boldsymbol{b},\boldsymbol{c}}\right\|_{fro}\left\|\boldsymbol{\kappa}\right\|_{fro}}; \forall b \in B, c \in C \tag{3-4}$$

The notations $\|.\|_{fro}$ and $\langle .,. \rangle$ is the Frobenius Norm and dot product respectively.

3.2.3. Clasification stage

The channels of each band are ranked according to the score given by the ρ, in this way the most relevant channels are selected for each band and with these the covariance matrix is built for each band and is estimated at $\boldsymbol{\Sigma}_{b'}^r = \mathbb{E}\left\{X_{b'}^r X_{b'}^{r\top} : \forall r \in N, b' \in B'\right\}$, being N the number of trials and B' the bands selected by the relevance analysis.

From a BCI point of view, spatial covariance matrices carry useful information. Indeed, on the diagonal of such an array, we can retrieve the variance of the signal measured by every, and the off-diagonal entries give extra-information (i.e.the correlations) on the sensors. Covariance matrices belong to the space of Symmetric Positive-Definite (SPD) matrices. As a subset of the Euclidean space of real symmetric matrices, this space can be equipped with the Euclidean scalar product and then a structure of convex cone arises. The SPD matrix belongs to a special Riemannian Manifold; this space is provided with several dissimilarity metrics are proposed to estimate the distance in this space as follow [11] [27].

- **Riemannian distance:** defines the geodesic distance between two SPD S_i and S_j as

$$d_R(S_i, S_j) = \left\|\log S_i^{-1} S_j\right\| = \sqrt{\sum_i^n \log \lambda_i{}^2} \tag{3-5}$$

  Where $\lambda_i = eig_i(S_i^{-1} S_j)$, This metric is invariant to an affine transformation and inversion. However, solving the generalized eigenvector is very computationally expensive for practical application.

- **Tangent Space distance:** metric can be approximated by the distance between tangent vectors through a common reference point S. The tangent vector $\widetilde{S}_i$ of a point S_i at the reference point S_{ref} is defined as

$$\widetilde{S}_i = \log_{S_{ref}} S_i = \log(S_{ref}^{-\frac{1}{2}} S_i S_{ref}^{-\frac{1}{2}}) \tag{3-6}$$

and the distance between S_i and S_j are derived as

$$d_{TS}^2(S_1, S_2) = \left\| \widetilde{S}_i - \widetilde{S}_j \right\|_F^2 = tr((\widetilde{S}_i - \widetilde{S}_j)(\widetilde{S}_i - \widetilde{S}_j)^T) \tag{3-7}$$

To obtain a good approximation with Riemannian geodesics, the reference point S_{ref} needs to be close to the two points. Hence, S_{ref} is heuristically selected as the geometric mean of the point set S_i.

- **Log-Euclidean distance:** selects the reference point S_{ref} at the identity matrix I. hence the distance is simplified as

$$d_{LE}^2 = \left\| \log S_i - \log S_j \right\|_F^2 \tag{3-8}$$

- **Kullback-Leiber (KL) divergence:** is not a geodesics but instead based on informative geometry

$$d_{KL}^2(S_i, S_j) = \frac{1}{2} tr\left(S_i^{-1} S_j + S_j^{-1} S_i \right) - d \tag{3-9}$$

where d is the dimension of the covariance matrix.

- **Stein divergence or LogDet Divergence:** defines the not geodesic distance between two SPD S_i and S_j as

$$d_{SD}^2(S_i, S_j) = \log\det\left(\frac{S_i + S_j}{2} \right) - \frac{\log\det(S_i S_j)}{2} \tag{3-10}$$

- **Von Neuman divergence:** defines quantum relative entropy between two SPD covariance matrices as

$$d_{VN}^2(S_i, S_j) = tr((S_i - S_j)(\log S_i - \log S_j)) \tag{3-11}$$

In the proposed methodology, we have used Stein's Kernel on covariance matrices as an elegant way to simplify BCI's system while having good performances. We use a Gaussian Kernel on the Riemannian geometry as well:

$$\boldsymbol{K_{b'}} = e^{-\alpha d^2_{SD}(\Sigma_{n_i b'}, \Sigma_{n_j b'})} \forall b' \in B' \tag{3-12}$$

Where $\boldsymbol{K_{b'}} \in \mathbb{R}^{NxN}$ is the Stein Kernel for each band selected , α is the bandwidth of the kernel and $i, j \in (0, .., N)$ are subindex that index the records. To apply the Stein kernels obtained in each band on BCI, we focus on a single kernel based on band weighting using Metric kernel Learning (MKL).

$$\boldsymbol{K_{MKL}} = \sum_{\forall b'} w_{b'} K_{b'} \tag{3-13}$$

We use a Support Vector Machine (SVM) equipped with a kernel on covariance matrices; it is possible to simplify and classify BCI systems. The proposed approach reduces the dimension spatio spectral based on the analysis of relevance offered in addition to simplifying the BCI system eliminating the stage of feature extraction achieving a better performance compared to methods in state of the art.

3.3. Experimental Set-Up

3.3.1. Description of Datasets and processing pipeline

The proposed approach of relevance-based spatio-spectral selection is evaluated on the three online available datasets (The BCI Competition III dataset IVa[1], BCI competition IV dataset IIa[2] and GIGASCIENCE[3]), for motor imagery brain-computer interface described in Section 2.3.1, a summary with the parameters as number of subjects and number of channels is shown in Table **3-1**.

3.3.2. Parameter tuning

The proposed method MFS requires two procedures of parameter tuning; one is the kernel bandwidth $\sigma \in \mathbb{R}^+$, and another parameter is the threshold $\lambda \in \mathbb{N}^+$ which controls the amount of spatio-spectral components selected. The parameter σ and λ is set up based on a grid search for the best classification accuracy.
The spatio-spectral components used for the classification are selected according to a ranking from relevance analysis, the Fig. **3-1** illustrates the learning curves for in each subject in the BCICIV

[1] http:// www.bbci.de/competition/iii/
[2] http://www.bbci.de/competition/iv/
[3] http://gigadb.org/dataset/100295

Sample rate	Subjects	Channel	trial/class	time MI
BCICIII IVa				
100Hz	5	118	140	2.5-4.5s
BCICIV IIa				
250Hz	9	22	69	2.5-4.5s
GIGASCIENCE				
512Hz	44	64	100	2.5-4.5s

Tabla 3-1: Dastandet description

IIa dataset, the component best pair of parameters found in each subject in the three evaluated datasets. Note that the maximum success is with a number of different components for each subject, evidencing the need for tuning λ for each subject to coding properly the different activities of the brain.

Observing the learning curves Fig. **3-1** we identify two main dynamics one for the good subjects (S1, S3, S5, S7, S8, S9) and the other for the subjects with lower performance (S2, S4, S6).

The first dynamic presents in principle a monotonous trend growing, but its main feature is that achieved stabilized with a few components and maintain your success revealing the goodness of our approach for the selection of relevant elements.

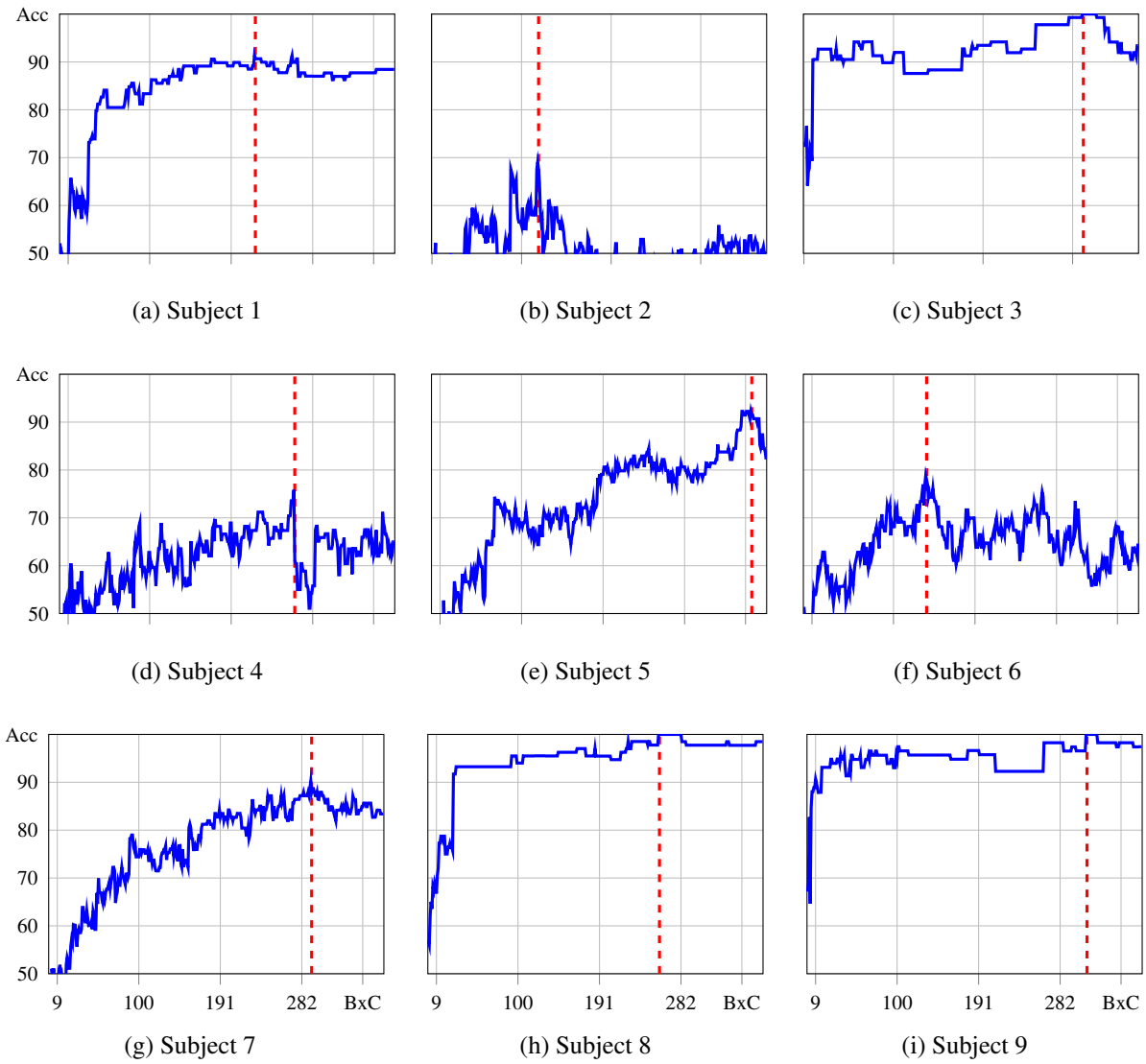
The dynamics associated with subjects S2, S4, and S6, are more sensitive as components are added because of this there are abrupt changes induced by noisy components that disturb the classifier, however the analysis of relevance proposed ranked than these components obtaining the best performance with few elements and a superior success compared to the success with all the components.

3.4. Performance results

We compare the MFS against the conventional CSP and Sparse Filter-bank (SFB) CSP [31] and Kernel Stein without information of frequency and dimension reduction [11] methods for the considered datasets. Fivefold cross-validation is used to evaluate the performance of all experiments conducted using BCICIII4a, BCICIV2a, and GIGASCIENCE datasets.

Table **3-2** summarizes the accuracy performed by each different approach, the table also shows the standard deviation represented by $\pm$, the component reduction rate (*crr*) and the *p*-value of the proposed approach against with other competing methods in the literature.

As can be seen from the results in Table **3-2** the use the relevance analysis proposed for band and channel selection (MFS) shows an improved performance of 8,1 %, y 8,8 % and 7,8 % and a reduction in 56,7 %, 32,6 % and 59,0 % in the number of spatio-spectral components (for BCICIII4a,



(a) Subject 1 (b) Subject 2 (c) Subject 3

(d) Subject 4 (e) Subject 5 (f) Subject 6

(g) Subject 7 (h) Subject 8 (i) Subject 9

Figura 3-1: BCICIV IIa learning curves, the best accuracy is depicting by a dashed line the best for each subject and BxC is the number of spatio-spectral components (Band X Channel)

BCICIV2a and GIGASCIENCE datasets, respectively), this data show that much of the information in the original data hinders the classification process, In Table **3-2** parameter Crr (component rate selection) shows that the quantity of spatio-spectral components used in the methodology which reveals a considerable reduction, which is necessary for the method to have stability, to eliminate redundant and noisy information, the need for the feature selection stage is validated by the success obtained in comparison with the flat Stein Kernel method that does not use features selection. On the other hand, the p-value indicates that our approach is statistically significant against the methods in state of the art with a value of less than 1 %.

Fig. **3-2** presents the accuracy rate attained for each subject by the evaluated approaches. Note that

	CSP	Kernel Stein	SFB	MFS
		BCICIII IVa		
$\mu \pm \sigma$	87.6 ± 4.2	84.3 ± 4.4	87.7 ± 2.4	92.4 ± 3.1
p-value	$2.0e^{-3}$	$8.8e^{-6}$	$5.1e^{-3}$	-
crr	118/118	2006/2006	2006/2006	867/2006
		BCICIV IIa		
$\mu \pm \sigma$	76.5 ± 5.6	79.8 ± 7.0	84.5 ± 5.4	88.6 ± 5.1
p-value	$4.8e^{-7}$	$2.1e^{-4}$	$6.6e^{-3}$	-
crr	22/22	374/374	374/374	252/374
		GIGASCIENCE		
$\mu \pm \sigma$	68.9 ± 7.6	70.8 ± 6.6	72.2 ± 6.3	78.1 ± 6.5
p-value	$4.0e^{-7}$	$7.9e^{-28}$	$4.1e^{-13}$	-
crr	64/64	1088/1088	1088/1088	446/1088

Tabla 3-2: Classification accuracy for the considered approaches. Mean and standard deviation for five-fold cross-validation is computed, p-value is from the statical is the significance level between MFS and the other methods and crr is the component rate selection

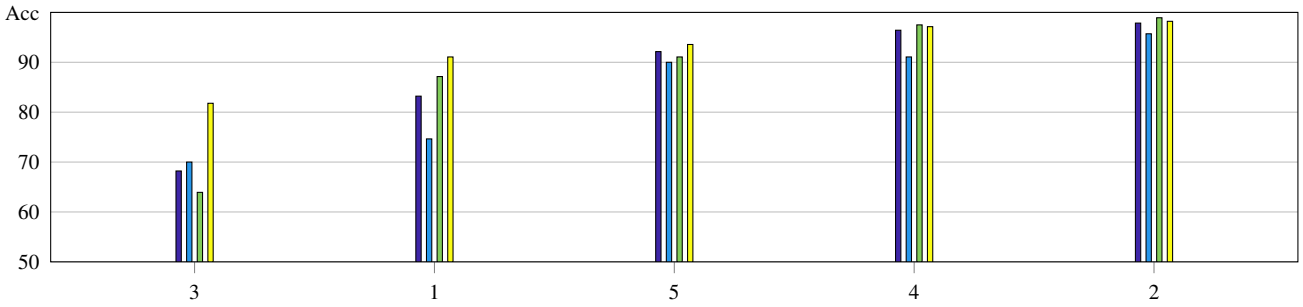
subjects are sorted according to the CSP accuracy, achieving show that MFS improves the performance of the worst subjects while maintaining the accuracy of the best subjects in comparison with CSP, Kernel Stein, and SFB, our method outperforms by achieving the highest accuracy rates for most of the subjects in the three datasets.

3.5. Discussion

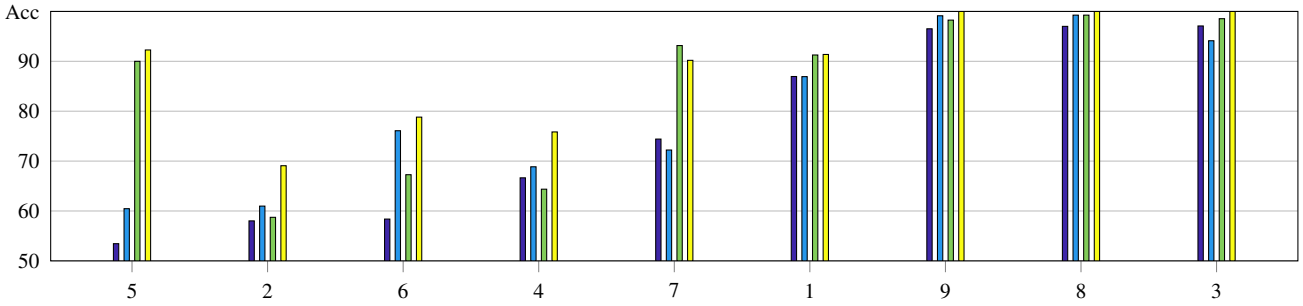
This work proposes a spatio-spectral relevance analysis to select the most discriminative spatial bandpass filtered components of EEG recordings, termed MFS. As the accuracy results prove supervised information to the relevance criterion enhances the stability and the discriminative power of the extracted Kernel Stein method features.

The Stein Kernel method and the Riemannian geometry suffer from stability problems this is the case because of the high dimensionality of the SPD matrix (covariance matrix) product of the configurations with a large number of sensors, the proposed approach solves this problem decreasing the probability of a single array, thanks to the reduction of dimension and redundant information.

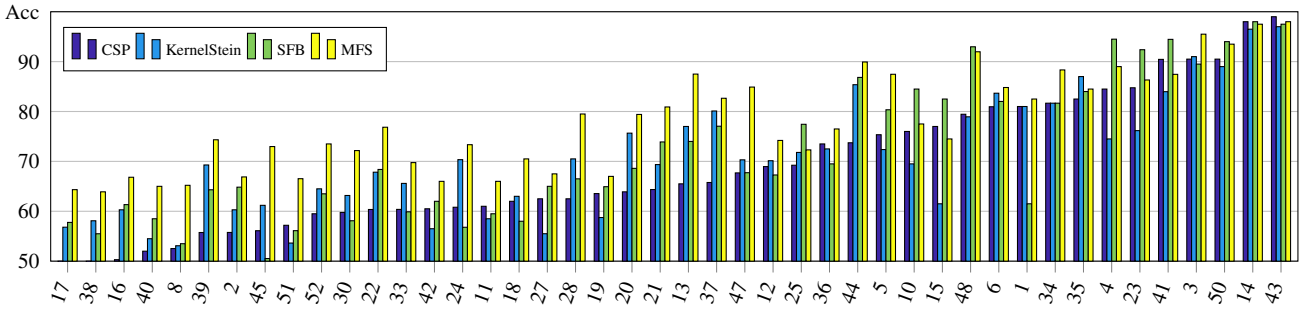
MFS incorporates frequency information this is a significant improvement in success because specific bands encode the information related to brain activity, this can be seen in Fig. **3-3** shows the



(a) BCIIII IVa



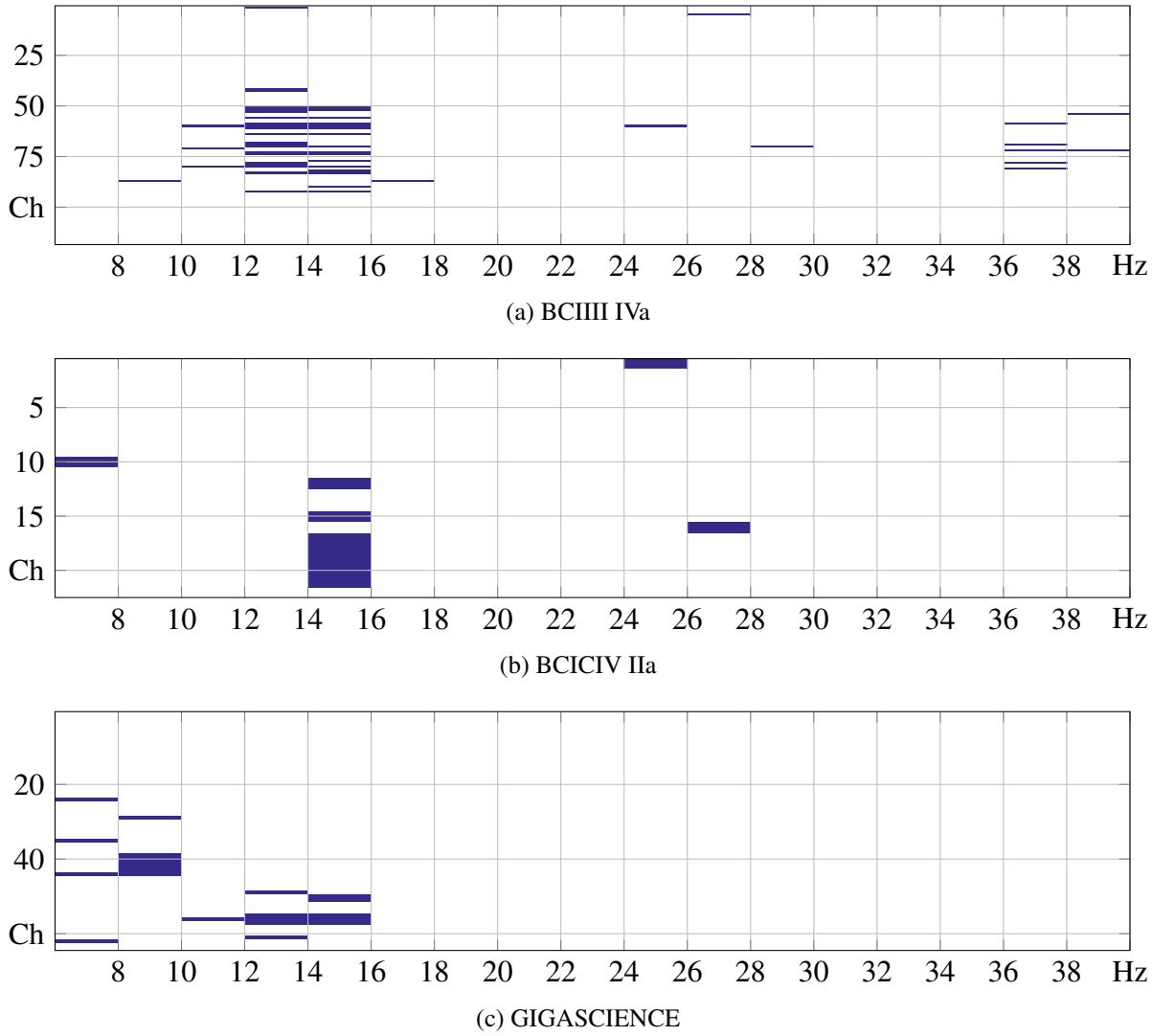
(b) BCICIV IIa



(c) GIGASCIENCE

Figura 3-2: Performed accuracy by each tested approach on the three datasets. Average of a five-fold cross validation is depicted. Subjects are sorted according the performed accuracy, note that CSP is the baseline approach.

bands most used by the channels in each of the databases, wherein the BCI database Competition III DataSet Vat is clearly seen as the relevant information was found in bands 12 to 16 Hz, in the database BCI Competition IV DataSet IIa the band of $14-16\,Hz$ is repeated and in GIGASCIENCE the band of $8-10\,Hz$ appears and the band of $12-16\,Hz$ is maintained, these bands are strongly related to the MI task because finding that imagination of movement leads to short-lasting and circumscribed attenuation (or accentuation) in mu ($8-12\,Hz$) and beta ($13-28\,Hz$) rhythmic activities [25]. Entering frequency information in the stage feature selection improves the success this can be seen reflected in subjects such as S5, S2 of BCI Competition IV DataSet IIa Fig. **3-2**b



(a) BCIIII IVa

(b) BCICIV IIa

(c) GIGASCIENCE

Figura 3-3: The most selected bands and channels for each dataset, the y-axis represents the channels and the x-axis the frequency bands.

that when entering frequency and component selection information to the Stein Kernel method It is improved by approximately 30%, this can also be seen in the first subjects of the BCI Competition III DataSet V Fig. **3-2**a and in GIGASCIENCE Fig. **3-2**c.

Our proposed approach obtains a better accuracy compared with the other methods evaluated for the three datasets, using a lesser amount of information eliminating components that could be redundant or noisy for the MI task and, in addition to obtaining a lower standard deviation, so statistically, our method turns out to be superior to those of the state of the art.

4 Fault trend of rotatory machines using health indicators based on analysis of relevance in time.

4.1. Introduction

In this chapter, we propose a methodology to describe the machine trend based on Ordinal Symbolic Dynamics called (OSDT); this proposed approach can code these complex changes of the signal vibrations. The methodology is compared against state-of-the-art methods, and the response of different indicators is also considered. A public dataset of bearing run-to-failure is utilized to test the proposed approach.

4.2. Divergence-based relevance analysis

Let a time serie $\{\boldsymbol{x}^n \in \mathbb{R}^T : n \in N\}$ be a set of N acquired recordings of length T, the envelope of $\boldsymbol{x}$ is calculated as:

$$\boldsymbol{y}^n = |\boldsymbol{x}^n + j\mathscr{H}\{\boldsymbol{x}^n\}|^2 \tag{4-1}$$

Where $|.|$ is the absolute value and $\mathscr{H}\{.\}$ is the Hilbert transform operator, the envelope $\boldsymbol{y}^n \in \mathbb{R}^T$ is partitioned into a D-dimensional space that will hopefully reveal “relevant "details of the ordinal structure of a given one-dimensional time series [9]. A series of ordinals patterns is generated for the signal with embedded dimension $D < 1$ and delay time τ, the ordinal patterns of order D are generated by:

$$y_i^n \rightarrow \left(y_{i-(D-1)}, \quad y_{s-(D-2)}, \quad ..., \quad y_{i-1} \quad y_i\right); \forall n \in \mathbb{R}^N \tag{4-2}$$

Which assigns to each time i the $D-$dimensional vector of values at times $i, i-1, ..., i-(D-1)$. Clearly, the greater the $D-$value, the more information on the past is incorporated into our vectors. By “ordinal pattern related"to the time (i), we mean the permutation $\pi = (r_0, r_1, ..., r_{D-1})$ of $[0, 1..., D-1]$ is defined by:

$$x^n_{i-r_{D-1}} \leq x^n_{i-r_{D-2}} \leq ... \leq x^n_{i-r_1} \leq x^n_{i-r_0}; \forall n \in \mathbb{R}^N \tag{4-3}$$

In order to get a unique result we set $r_i < r_{i-1}$, if $x_{s-r_i} = x_{s-r_{i-1}}$. Thus, for the $D!$ possible permutations π of order D, the associated relative frequencies can be naturally computed by the number of times this particular order sequence is found in the time series divided by the total number of sequences. The probability distribution for each record $p^n(\pi)$ is defined as in [3]:

$$p^n(\pi) = \frac{\theta\{s|s \leq T-D+1;(s), \quad has \quad type \quad \pi\}}{T-D+1} \tag{4-4}$$

In this expression, the symbol θ stands for “number". With the target to demonstrate the change of dynamics of the machine, there is quantified the difference of a current state (CS) and the healthy state (HS), based on the codification of ordinal symbolic dynamics (OSD). It is proposed to use Jensen’s divergence as a measure of dissimilarity between states as well:

$$p_i = P^n(\pi_i); \forall n \in \mathbb{R}^N \quad q_i = Q(\pi_i|HS) \tag{4-5}$$

$$OSDT(p_i, q_i) = \frac{1}{2}\sum_{\forall i} p_i \ln p_i - (p_i + q_i)\ln\frac{p_i+q_i}{2} + q_i \ln q_i \tag{4-6}$$

Where $P^n(\pi_i)$ and $Q((\pi_i)|y_{n^+}) : \forall n^+ \in 1 < n < 200$ is the Probability distribution function (PDF) for currently record and the healthy records y_{n^+} respectively, the Eq. (4-6) describes the trend of degradation of the machine and is the health index proposed. The proposed approach Evaluates each of the records and gets a value indicating that it changes its dynamics to healthy records, as shown below.

$$OSDT^n(p^{n^+}, q^n) \geq \varepsilon : \forall n \in N \tag{4-7}$$

Where p^{n^+} and q^n It is the discrete probability density function of the machine in its healthy state and in its current state respectively, ε is risk threshold equals the mean plus six times the standard deviation (i.e. $thres = \mu + 6\sigma$), when the Eq. (4-7) exceeds the threshold then it detects a potential failure in the bearing.

4.3. Experimental Set-Up

4.3.1. Description of Dataset and processing pipeline

IMS bearing data base

The bearing data are obtained from Prognostics Center Excellence through the prognostic data repository contributed by Intelligent Maintenance System - *IMS*, University of Cincinnati [15]. The data are collected in an experimental setup showed in Fig. **4-1**, which holds test-to-failure bearing rig, including four Rexnord ZA-2115 double row bearings on one shaft that is driven by an AC motor and coupled by rub belts, under a constant rotation speed adjusted to 2000 `r/min`.

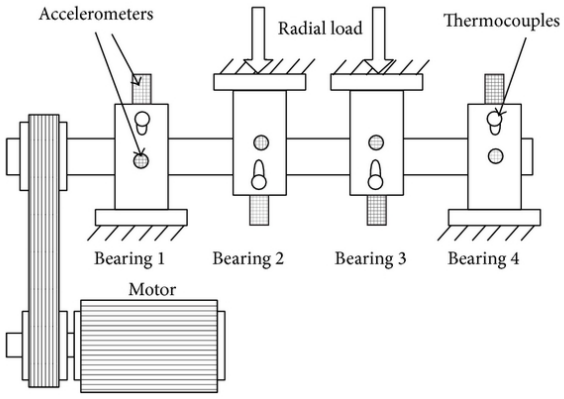


Figura 4-1: Bearing test rig and sensor placement [15].

Each bearing has 16 rollers on each row, a diameter of 2,815 in, roller diameter of 0,331 in, and a tapered contact angle of 15,17°. A PCB 353B33 High Sensitivity Quartz ICP Accelerometer has been installed on each bearing housing. The data sampling rate is 20kHz, and the data length is 20480 points. Vibration data had been collected every 20min. In this case, we use the test #1 built to examine two different faulty degradation process: *i)*a ball defect in the bearing #4, and *ii)* an inner race bearing fault in the bearing #3 and as detailed in [16].
In the preprocessing stage, the bearing fault frequency is approximated by the geometrical properties of the rolling element bearings, and the shaft rotational speed f_r of the system (see Table **4-1**), where p is the number of rolling elements, D and d are respectively the bearing pitch and ball diameters, θ is the contact angle of the load from the radial plane.

Ball pass frequency, outer race:	$BPFO = \frac{pf_r}{2}\left(1 - \frac{d}{D}\cos\theta\right)$
Ball pass frequency, inner race:	$BPFI = \frac{pf_r}{2}\left(1 + \frac{d}{D}\cos\theta\right)$
Fundamental train frequency (cage speed):	$FTF = \frac{f_r}{2}\left(1 - \frac{d}{D}\cos\theta\right)$
Ball spin frequency:	$BSF = \frac{Df_r}{2d}\left(1 - \left(\frac{d}{D}\cos\theta\right)^2\right)$

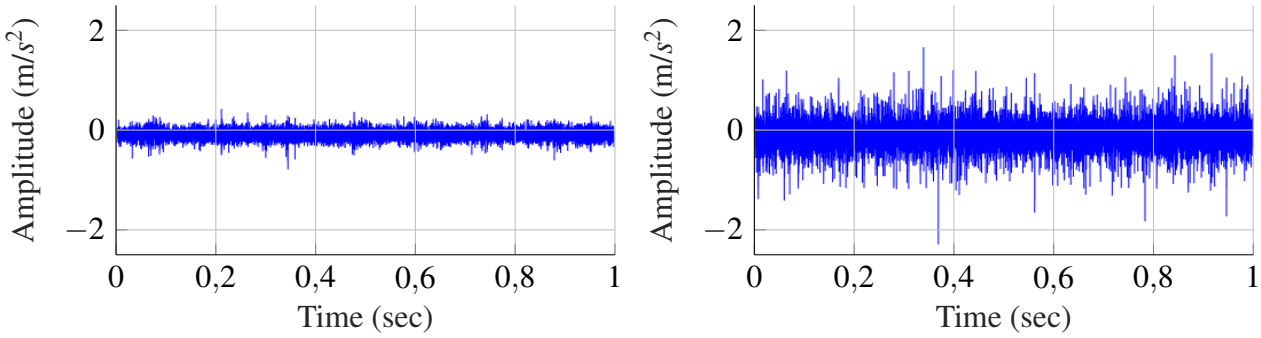
Tabla 4-1: Bearing fault frequencies in rotating machines.

The above fault frequencies are based on kinematic relationships, assuming that there is not frequency slip. It is always present having a variation of the calculated frequency with values up to $1-2\,\%$ [18]. The main spectral characteristics, which may be identified in the envelope spectrum, are as follows for the outer race, inner race, and rolling elements [20]: *i)* $BPFO$ and harmonics, sidebands spaced at f_r; *ii)* $BPFI$ and harmonics, sidebands spaced at f_r, harmonics of f_r; And *iii)* BSF and harmonics (even harmonics often dominant), sidebands spaced at FTF and harmonics of

FTF.

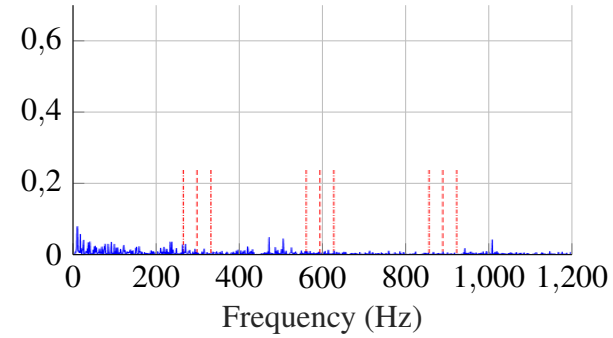
Processing pipeline

This section investigates and compares the proposed approach, RMS, and CS-based methodology on vibration signals acquired from the IMS database.

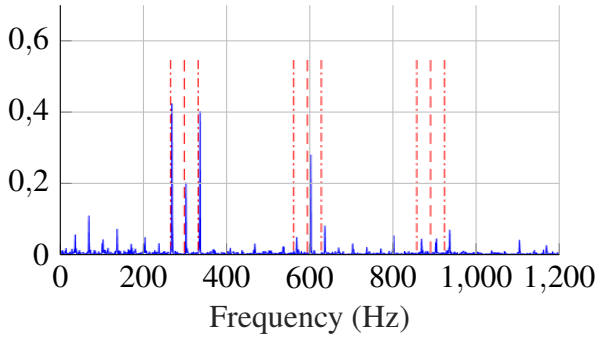


(a) Record 1 undamaged

(b) Record 2150 damaged



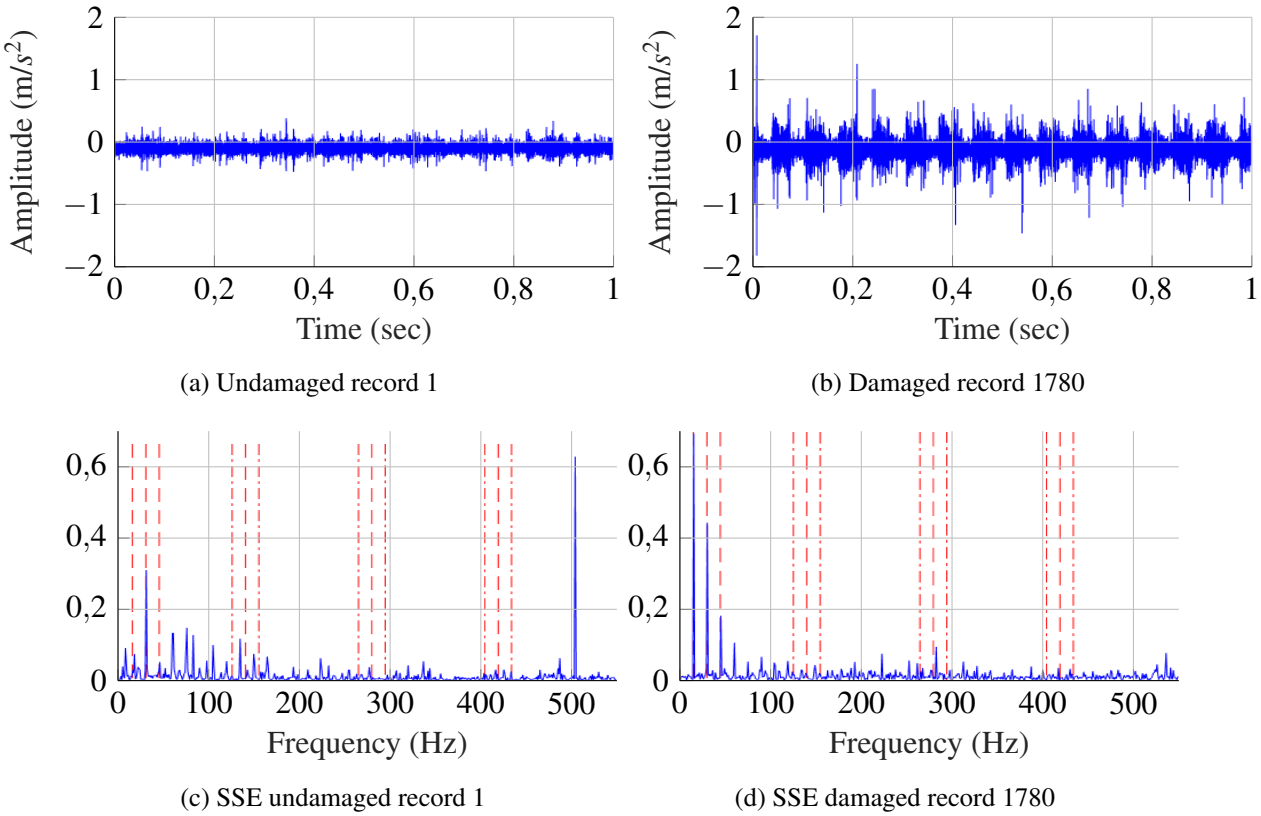
(c) SSE Record 1 undamaged



(d) SSE record 2150 damaged

Figura 4-2: Exemplary of the vibration signal in bearing #3: raw signal (top) and the spectrum of its envelope (bottom) under undamaged and damage condition. $BPFI$ (297 Hz) and its sidebands in $BPFI \pm f_r$ are displayed in red color.

The dataset consists of a laboratory benchmark comprising a Rolling element bearing (REB) subjected to natural degradation; our goal is the early identification of a potential failure. The proposed methodology in Initially at the start of the acquisition process, the squared envelope Eq. (4-1) will be applied on the raw time domain vibration data; the ordinal symbolic dynamic will be used to build the ordinal patterns which will be used to construct the probability distribution. For any new acquired signal, Jensen divergence evaluated by calculating the distance between the probability distribution of healthy and the new signal, the result of this dissimilarity is the health index of the machine, named Ordinal dynamic symbolic trending (OSDT). In practical situations, while acquiring data over the entire lifespan of the bearing, sharp rises (outliers) and random fluctuations can be seen in the values of the condition monitoring parameter.



(a) Undamaged record 1

(b) Damaged record 1780

(c) SSE undamaged record 1

(d) SSE damaged record 1780

Figura 4-3: Exemplary of the vibration signal in bearing #4: raw signal (top) and the spectrum of its envelope (bottom) under undamaged and damage condition. Harmonic cursors multiples of FTF (14,5 Hz) are displayed in red color.

This poses a challenge in the identification of the change in the health state of a bearing. To assess the current status, it is essential to benchmark the healthy working state of the bearing. Therefore a reference data set is obtained from the initial operation of the bearing after commissioning of the fresh bearing, for this dataset the first 200 records are used to build the reference of the healthy state, in addition to constructing a risk threshold equals the mean plus six times the standard deviation (i.e. $thres = \mu + 6\sigma$) for the bearing as shown in the Fig. **4-4**a. If the health index is more than the predefined threshold and its next 20 records also recognize a signal dynamics change which may imply a bearing failure in another case it is considered that the register is an outlier, this is done to minimize the issue of false alarm for bearing degradation.

To provide a better interpretation of the analyzed signals, undamaged and damaged state of bearings #4 and #3 are shown in Fig. **4-3** and Fig. **4-2**, respectively.

Where it is displayed the vibration in the time domain (top row) and the square spectrum of the signal envelope (bottom row), it is possible to highlight that the undamaged condition neither present modulation processes nor impulsive behavior since the highest peak on envelope analysis matches with the rotational frequency at f_r=33,33 Hz, and there is no information related to bearing

fault frequencies.

In contrast, the vibration signal associated to the damaged machine evolution through the time, notably the recording 1780, shows several impulses with a period of 0,0687 sec, which is equivalent to FTF (14,5 Hz). In the signal, spectrum envelope is worth noting that the dominant frequency corresponds to FTF and its harmonics (dashed lines in Fig. **4-3**d-(bottom)), and hence, it is possible to say that the bearing has a ball defect.

4.4. Performance results

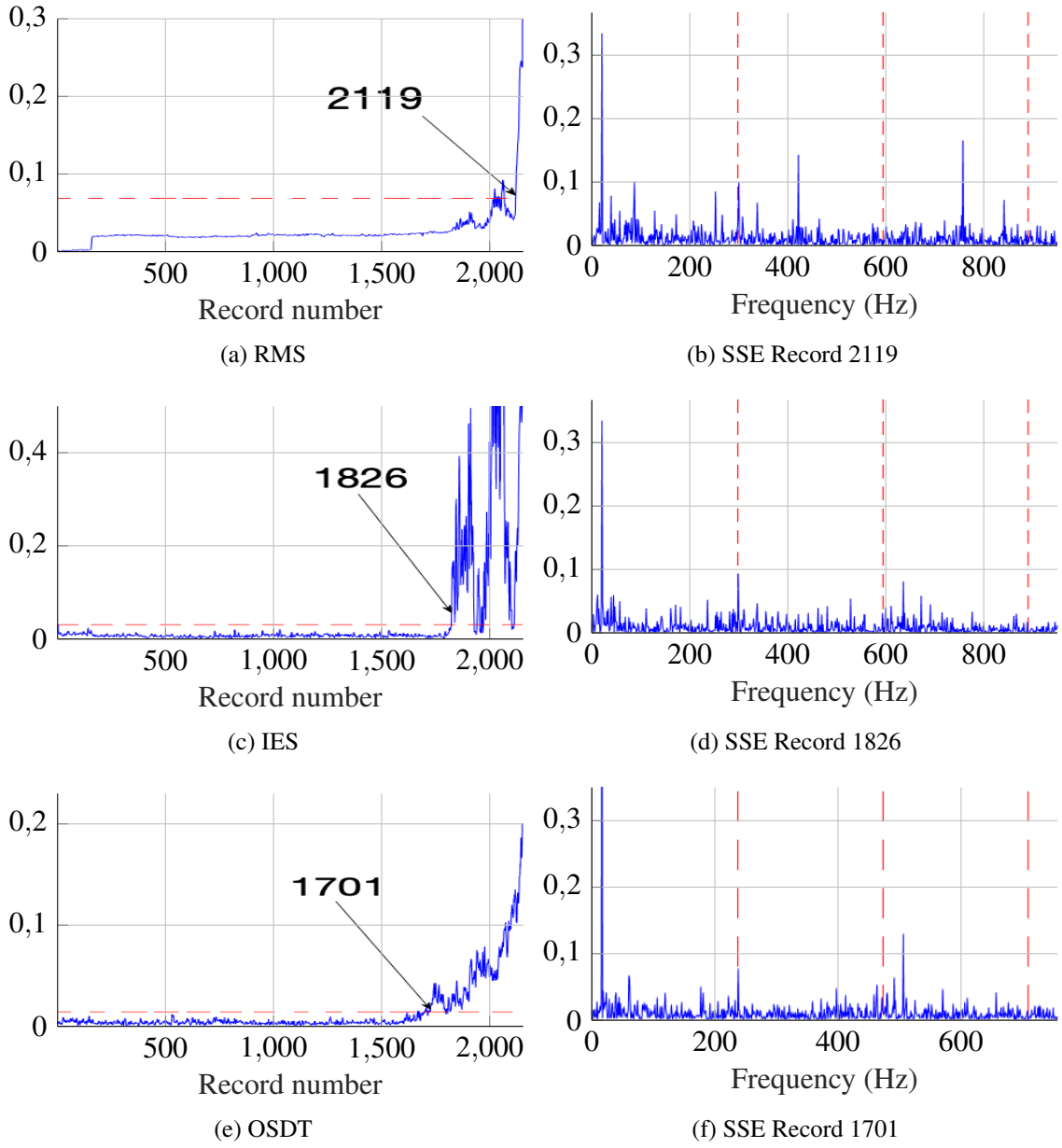
This subsection is dedicated to comparing the methods on a run-to-failure experiment in which the defect has been naturally developed. Comparison is performed in three distinct accelerometers. The aim of this study is to verify the performance of Jensen divergnece on OSD measure in distinctly identifying defect stages and show significant changes in the Jensen divergnece values as the stage changes after every 200 data points.

This subsection investigates the fault detectability of the bearing through the signals acquired by the accelerometer three mounted on the faulty bearing (i.e. bearing 3) which contains inner race defect. The importance to investigate such a simple case is to compare the detection precocity of these approaches in a low noise environment. The signal RMS value and CS-based indicators (IES) [1] as well the proposed approach OSDT are computed on the envelope of the acquired signals and exposed in Figs. **4-4**a, **4-4**c and **4-4**e respectively. The RMS value is a general indicator of the system; a change in its value may indicate a change in the system operation which may be related to a defect.

Fig. **4-4**a shows an increase of this indicator, which exceeds the threshold at a record 2219, indicating the presence of the defect. The value of the RMS has then decreased under the limit around the register 2000, before exceeding it again soon afterward. The reason may be that the fault surface has extended during the degradation process, which has led to smoothening of the impulsive excitation [1], this phenomenon is also observed in the other indicators. When it comes to the CS-based indicator Fig. **4-4**c has increased and exceeded the threshold at a record 1826 . Interestingly, the OSDT indicator exceeds the limit at register 1701, allowing the identification of the fault type earlier than the RMS and CS-based indicator.

To make this clear, we analyze the records where the fault was identified for each method. The first record given by RMS is the 2119 Fig. **4-4**b shows the squared envelope spectrum (SEE), which presents a peak in the first harmonic of 236Hz this is the frequency related to the failure of external Psita Ball pass frequency, outer race (BPFO) which is dominant in the spectrum there is already a change in the dynamics of the bearing due to a physical degradation in its structure.

Note that although the three methods detected a failure in different early records when reviewing their spectrum, there is a clear dominant peak in the frequency in BPFO is already a change in the dynamics of the bearing due to a physical degradation in its structure. The diagnosis in the spectrum of the BPFI fault cannot be made because the record is very early and the fault has not



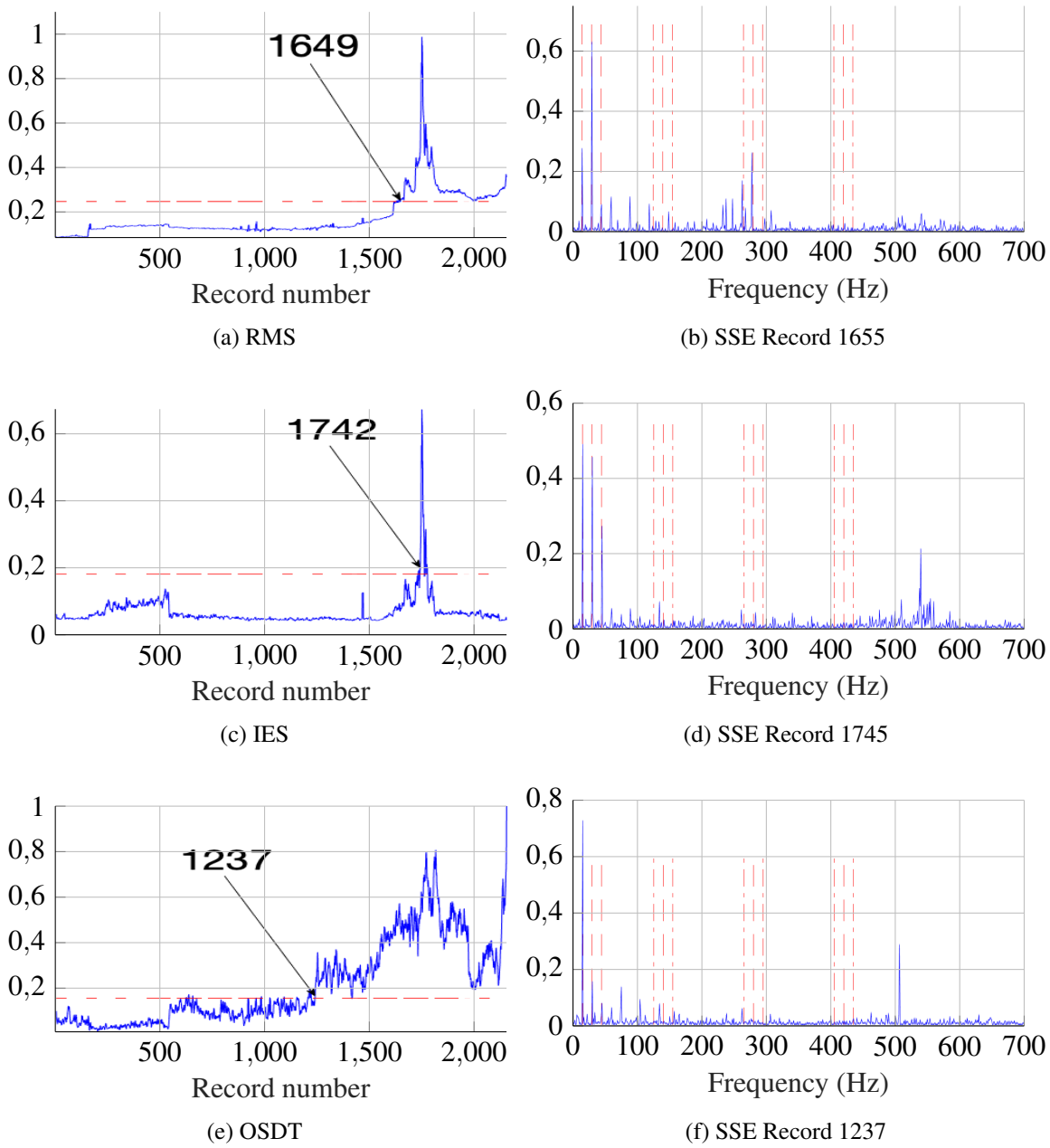
(a) RMS (b) SSE Record 2119

(c) IES (d) SSE Record 1826

(e) OSDT (f) SSE Record 1701

Figura 4-4: Health index in the 1st test bearing#3: a) The RMS value, c) the CS-based indicators, e) the health index proposed OSDT.

been developed to show all its pattern, the failure of BPFI is observed in record 2150 in Fig. **4-2**d where the useful life of the bearing an advanced for That his pattern is observed clear Figs. **4-4**d and **4-4**f shows the SEE of IES and OSDT respectively, both show clear components in the first harmonic 236Hz of BPFO, their spectra validate the three methodologies to detect the early failure of the bearing. Interestingly OSDT detects a change in the dynamics of the PDF of the bearing, 125 records before IES and 418 records before RMS which translates to approximately 3 and 7 days, for the industry this implies more time to do the maintenance or change of the bearing damaged,



(a) RMS (b) SSE Record 1655

(c) IES (d) SSE Record 1745

(e) OSDT (f) SSE Record 1237

Figura 4-5: Health index in the 1st test bearing#4: a) The RMS value, c) the CS-based indicators, e) the health index proposed OSDT.

which translates into savings for the company, even avoid one stopped in production.

The bearing 4 presents a failure related roller element, indicators RMS Fig. **4-5**a, IES [1] Fig. **4-5**c, and OSDT Fig. **4-5**e show a failure alarm in the records 1649, 1742 and 1237 respectively. OSDT succeeds in identifying the presence of a potential failure 505 records before that IES and 412 records before that RMS or 9 and 7,5 days.

The Figs. **4-5**b, **4-5**d and **4-5**f shows the SEE of RMS,IES, and OSDT for the record where the

respective method indicates the start of the fails, the spectrum of the three methods reveals key components in FTF and its harmonics which is a clear pattern in the ball failure [18], indicating the presence of a fault on rolling elements (BSF), in registry 1649 Fig. **4-5**b where RMS alert of a fault start, is recognized the presence of a component in the fundamental of BSF, this is because it is an advanced record where the fault begins to make more evident and to be dominant on the noise and other components of the signal.

Investigates the fault detectability of the bearing through the signals acquired by the accelerometer one mounted on the faulty bearing (i.e. bearing 1), in test 2, this is an easy case as the defect (i.e. the source) is close to the sensor, and the fault component is expected to dominate the response. The signal RMS value, IES, and OSDT indicators are computed on the acquired signals and exposed in Fig. **4-6**. The RMS value Fig. **4-6**a is a general indicator of the system; a change in its value may indicate a difference in the system operation which may be related to a defect Figura shows an increase of this indicator which exceeds the threshold at a record 535, indicating the presence of the defect. The OSDT Fig. **4-6**e related to the BPFO exceeds the limit at record 532, allowing the identification of the fault type earlier than the RMS and IES Fig. **4-6**c indicator.

Note that as the spectra Figs. **4-6**b, **4-6**d and **4-6**f evolve, the peak related to the fault becomes more dominate and its pattern more evident, for these records the three methods are very close in the detection of change of dynamic, it is presumed that the bearing had a crack considerable in the outer race of the bearing, as it is shown mainly in IES and OSDT the upward trend of indicator has a pronounced slope, then all methods come to a tipping point in about the record 1550, and its tendency begins to decrease, this is because probably the fissure has been filled with particles product of the continuous rubbing of the balls with the surface or the reason may be that the fault surface has extended during the degradation process which has led to smoothening of the impulsive excitation [1].

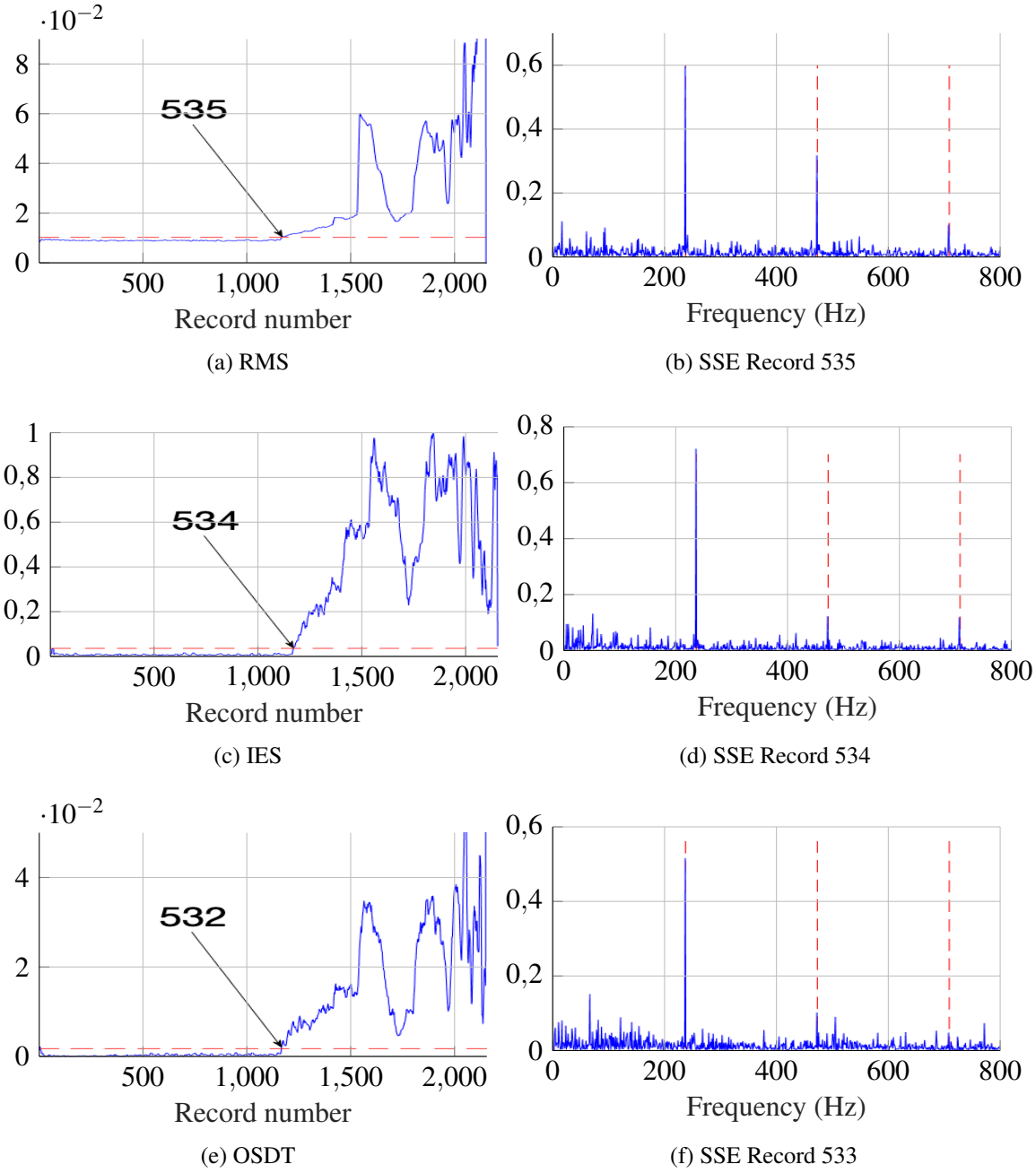
4.5. Discussion

This chapter proposes an index of the health status of the rotary machine used signal processing techniques for vibration-based fault detection of rolling element bearings and a relevance analysis based on entropy (OSDT) compares it against the methods of state of the art RMS and IES.

In the case studies, the three methods were able to detect the bearing faults. However, results have asserted the superiority of the OSDT in terms of detection precocity, robustness, and immunity against nonstationary noise interference.

To show the benefits of OSD, it is shown in Section 4.5 the PDFs built from the ordinal patterns. When shows the PDF of the healthy state and in a state of degradation. The probability of each event encodes the dynamics of the signals, note that the likelihood of some patterns where the failure arise differ from the patterns of a healthy state, the Jensen divergence finds this differences and the quantified.

On the one hand, ordinal patterns are relatively robust concerning small noise, affecting only a few



(a) RMS (b) SSE Record 535

(c) IES (d) SSE Record 534

(e) OSDT (f) SSE Record 533

Figura 4-6: Health index in the 2nd test bearing#1: a) The RMS value, b) the CS-based indicators, c) the health index proposed OSDT.

of the order relations between values, and on the other hand, they do not change under monotone transformations which means that they are invariant in particular with respect to different scalings or offsets of the original signal [8]. The results obtained reveal the superiority of the proposed method to detect a change of dynamics related to the health of the machine, as future work is to adapt the methodology to focus on a trend for each bearing failure.

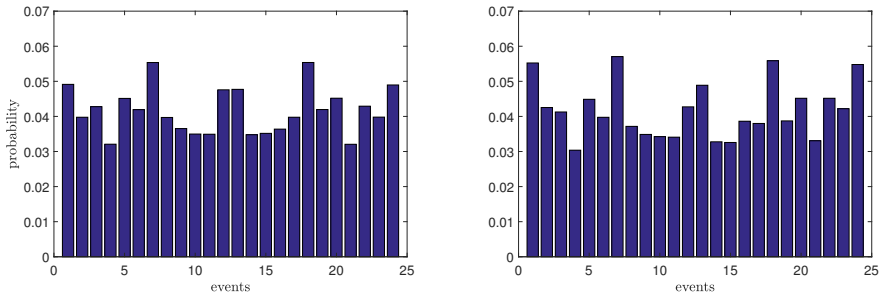


Figura 4-7: The probability density function obtained by the encoding ordinal symbolic dynamics in the 1st test bearing#3: The PDF in health state(left) and th FDP in the degradation state.

5 Conclusions and Future Work

5.1. General Conclusions and Main Contributions

Development of a supervised tool based on kernels and entropy to estimate significant spatio-spectral components in a classification task.

Finding stimulated areas of the brain is very important to develop a high performing BCI system. Since it is known that different parts of the brain are involved in various MI, actions, and feelings. Nonetheless, addressing this problem can be very hard due to inter-subject variability. In this study, an automatic method based on spatio-spectral relevance analysis supporting EEG discrimination is presented to find the bands and channels that contain the most discriminating and relevant information for MI task. The spatio-spectral component selection approaches using analysis based on different characteristics the Renyi entropy, Permutation entropy, and Kernel Stein; each spatio-spectral component is ranked by analyzing its contribution in classifying two MI classes. The purpose of this tool is to select a reduced set of components that improve the discrimination performance and reduce computational costs of classification; these connections were defined as significant components.

To develop a methodology for extracting saptio-spectral components that improve the success of the classifier reducing the dimension of the data, for the MI task.

The results showed that a set of few meaningful connections selected automatically using the information of the MI classes could achieve the maximum classification accuracy. Given that the analysis of relevance proposed kernel and entropy-based uses a smaller amount of components to 14 % in BCICIII IVa, 61 % in BCICIV IIa and 31 % GIGASCIENCE of the total elements in each one of the databases used and that the proposed methodology achieved better classification accuracy than state of the art methods.

To propose a methodology that evaluates the relevance of records over time using vibration signals, within an entropy-based relevance analysis.

We propose a methodology to describe the machine trend based on Ordinal Symbolic Dynamics called (OSDT); this proposed approach can code these complex changes of the signal vibrations over time. Results have asserted the superiority of the OSDT in terms of detection precocity, ro-

bustness, and immunity against nonstationary noise interference.

5.2. Future Work

For the proposed methodology of analysis of relevance using kernel and entropy-based feature selection approach for MI task, as future work, we plan to construct sub-bands that can be adapted to each subject aiming to tackle the computational burden. Also, optimization of free parameters has to be studied using more elaborated strategies like particle swarm optimization (PSO) to avoid the exhaustive search of the tuning parameters λ and ε to reduce the training. For the proposed methodology of analysis of relevance using entropy-based for the construction of a health index of the rotary machine, the author proposes as a future work the development of an index that can be focused on specific bearing failures such as BPFO, BPFI, and BSF.

Bibliografía

[1] ABBOUD, D. ; ELBADAOUI, M. ; SMITH, W.A. ; RANDALL, R.B.: Advanced bearing diagnostics: A comparative study of two powerful approaches. En: *Mechanical Systems and Signal Processing* 114 (2019), jan, p. 604 – 627. – ISSN 0888–3270

[2] ANG, Kai K. ; CHIN, Zheng Y. ; ZHANG, Haihong ; GUAN, Cuntai: Filter Bank Common Spatial Pattern (FBCSP) in Brain-Computer Interface. En: *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, 2008. – ISSN 2161–4393, p. 2390–2397

[3] BANDT, Christoph ; POMPE, Bernd: Permutation entropy: a natural complexity measure for time series. En: *Physical review letters* 88 (2002), Nr. 17, p. 174102

[4] CARRICARTE-NARANJO, Claudia ; CORNFORTH, David J. ; SANCHEZ-RODRIGUEZ, Lazaro M. ; BROWN, Marta ; ESTÉVEZ, Mario ; MACHADO, Andres ; JELINEK, Herbert F.: Rényi and permutation entropy analysis for assessment of cardiac autonomic neuropathy. En: ESKOLA, Hannu (Ed.) ; VÄISÄNEN, Outi (Ed.) ; VIIK, Jari (Ed.) ; HYTTINEN, Jari (Ed.): *EMBEC & NBC 2017*. Singapore : Springer Singapore, 2018. – ISBN 978–981–10–5122–7, p. 755–758

[5] CERRADA, Mariela ; SÃ¡NCHEZ, RenÃ©-Vinicio ; LI, Chuan ; PACHECO, Fannia ; CABRERA, Diego ; DE OLIVEIRA, JosÃ© V. ; VÃ¡SQUEZ, Rafael E.: A review on data-driven fault severity assessment in rolling bearings. En: *Mechanical Systems and Signal Processing* 99 (2018), p. 169 – 196. – ISSN 0888–3270

[6] CHO, Hohyun ; AHN, Minkyu ; AHN, Sangtae ; KWON, Moonyoung ; JUN, Sung C.: EEG datasets for motor imagery brain computer interface. En: *GigaScience* 6 (2017), Nr. 7, p. 1–8

[7] HIGASHI, H. ; TANAKA, T.: Simultaneous Design of FIR Filter Banks and Spatial Patterns for EEG Signal Classification. En: *IEEE Transactions on Biomedical Engineering* 60 (2013), April, Nr. 4, p. 1100–1110. – ISSN 0018–9294

[8] KELLER, Karsten ; SINN, Mathieu: Ordinal symbolic dynamics. En: *Preprint, Lübeck* (2005)

[9] KOWALSKI, Andres M. ; MARTIN, Maria T. ; PLASTINO, Angelo ; JUDGE, George: On Extracting Probability Distribution Information from Time Series. En: *Entropy* 14 (2012), Nr. 10, p. 1829–1841. – ISSN 1099–4300

[10] LI, Ke ; ZHANG, Yue L. ; LI, Zhi X.: Application research of Kalman filter and SVM applied to condition monitoring and fault diagnosis. En: *Applied Mechanics and Materials* Vol. 121 Trans Tech Publ, 2012, p. 268–272

[11] NGUYEN, Chuong H. ; ARTEMIADIS, Panagiotis: EEG feature descriptors and discriminant analysis under Riemannian Manifold perspective. En: *Neurocomputing* 275 (2018), p. 1871 – 1883. – ISSN 0925–2312

[12] NICOLAOU, Nicoletta ; HOURIS, Saverios ; ALEXANDROU, Pandelitsa ; GEORGIOU, Julius: Permutation Entropy for Discriminating 'Conscious' and 'Unconscious' State in General Anesthesia. En: ILIADIS, Lazaros (Ed.) ; JAYNE, Chrisina (Ed.): *Engineering Applications of Neural Networks*. Berlin, Heidelberg : Springer Berlin Heidelberg, 2011. – ISBN 978–3–642–23957–1, p. 280–288

[13] NOVI, Q. ; GUAN, C. ; DAT, T. H. ; XUE, P.: Sub-band Common Spatial Pattern (SBCSP) for Brain-Computer Interface. En: *2007 3rd International IEEE/EMBS Conference on Neural Engineering*, 2007. – ISSN 1948–3546, p. 204–207

[14] PENG, Z.K. ; CHU, F.L.: Application of the wavelet transform in machine condition monitoring and fault diagnostics: a review with bibliography. En: *Mechanical Systems and Signal Processing* 18 (2004), Nr. 2, p. 199 – 221. – ISSN 0888–3270

[15] QIU, Hai ; LEE, Jay ; LIN, Jing ; YU, Gang: Robust performance degradation assessment methods for enhanced rolling element bearing prognostics. En: *Advanced Engineering Informatics* 17 (2003), Nr. 3, p. 127 – 140. – Intelligent Maintenance Systems. – ISSN 1474–0346

[16] QIU, Hai ; LEE, Jay ; LIN, Jing ; YU, Gang: Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics. En: *Journal of Sound and Vibration* 289 (2006), Nr. 4, p. 1066 – 1090. – ISSN 0022–460X

[17] RANDALL, RB: Applications of spectral kurtosis in machine diagnostics and prognostics. En: *Key Engineering Materials* Vol. 293 Trans Tech Publ, 2005, p. 21–32

[18] SMITH, Wade A. ; RANDALL, Robert B.: Rolling element bearing diagnostics using the Case Western Reserve University data: A benchmark study. En: *Mechanical Systems and Signal Processing* 64-65 (2015), p. 100 – 131. – ISSN 0888–3270

[19] SUN, G. ; HU, J. ; WU, G.: A novel frequency band selection method for Common Spatial Pattern in Motor Imagery based Brain Computer Interface. En: *The 2010 International Joint Conference on Neural Networks (IJCNN)*, 2010. – ISSN 2161–4393, p. 1–6

[20] TAYLOR, J.I.: *The vibration analysis handbook*. Vibration Consultants, 1994. – ISBN 9780964051706

[21] THOMAS*, K. P. ; GUAN, C. ; LAU, C. T. ; VINOD, A. P. ; ANG, K. K.: A New Discriminative Common Spatial Pattern Method for Motor Imagery Brain x2013;Computer Interfaces. En: *IEEE Transactions on Biomedical Engineering* 56 (2009), Nov, Nr. 11, p. 2730–2733. – ISSN 0018–9294

[22] THOMAS, K. P. ; GUAN, C. ; LAU, C. T. ; VINOD, A. P. ; K., K.: A New Discriminative Common Spatial Pattern Method for Motor Imagery Brain-Computer Interfaces. En: *IEEE Transactions on Biomedical Engineering* 56 (2009), Nov, Nr. 11, p. 2730–2733. – ISSN 0018–9294

[23] WANG, Guofeng ; LUO, Zhigao ; QIN, Xuda ; LENG, Yonggang ; WANG, Taiyong: Fault identification and classification of rolling element bearing based on time-varying autoregressive spectrum. En: *Mechanical Systems and Signal Processing* 22 (2008), Nr. 4, p. 934 – 947. – Special Issue: Crack Effects in Rotordynamics. – ISSN 0888–3270

[24] WANG, Yanxue ; XIANG, Jiawei ; MARKERT, Richard ; LIANG, Ming: Spectral kurtosis for fault detection, diagnosis and prognostics of rotating machines: A review with applications. En: *Mechanical Systems and Signal Processing* 66-67 (2016), p. 679 – 698. – ISSN 0888–3270

[25] YAMAWAKI, Nobuyuki ; WILKE, Christopher ; LIU, Zhongming ; HE, Bin: An enhanced time-frequency-spatial approach for motor imagery classification. En: *IEEE transactions on neural systems and rehabilitation engineering* 14 (2006), Nr. 2, p. 250–254

[26] YAN, Ruqiang ; LIU, Yongbin ; GAO, Robert X.: Permutation entropy: A nonlinear statistical measure for status characterization of rotary machines. En: *Mechanical Systems and Signal Processing* 29 (2012), may, p. 474 – 484. – ISSN 0888–3270

[27] YGER, F.: A review of kernels on covariance matrices for BCI applications. En: *2013 IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*, 2013. – ISSN 1551–2541, p. 1–6

[28] YGER, F. ; BERAR, M. ; LOTTE, F.: Riemannian Approaches in Brain-Computer Interfaces: A Review. En: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 25 (2017), Oct, Nr. 10, p. 1753–1762. – ISSN 1534–4320

[29] YGER, F. ; LOTTE, F. ; SUGIYAMA, M.: Averaging covariance matrices for EEG signal classification based on the CSP: An empirical study. En: *2015 23rd European Signal Processing Conference (EUSIPCO)*, 2015. – ISSN 2076–1465, p. 2721–2725

[30] ZHANG, Yu ; WANG, Yu ; JIN, Jing ; WANG, Xingyu: Sparse Bayesian Learning for Obtaining Sparsity of EEG Frequency Bands Based Feature Vectors in Motor Imagery Classification. En: *International Journal of Neural Systems* 27 (2017), Nr. 02, p. 1650032. – PMID: 27377661

[31] ZHANG, Yu ; ZHOU, Guoxu ; JIN, Jing ; WANG, Xingyu ; CICHOCKI, Andrzej: Optimizing spatial patterns with sparse filter bands for motor-imagery based brain computer interface. En: *Journal of Neuroscience Methods* 255 (2015), p. 85 – 91. – ISSN 0165–0270