UNIVERSIDAD DE CÓRDOBA

Agrupamiento, predicción y clasificación ordinal para series temporales utilizando técnicas de machine learning: aplicaciones

Doctorado con Mención Internacional
Programa de Doctorado: Computación Avanzada, Energía y Plasmas

David Guijo Rubio

Directores
César Hervás Martínez
Pedro Antonio Gutiérrez Peña

Departamento de Informática y Análisis Numérico
Córdoba, abril de 2021

TITULO: *Clustering, prediction and ordinal classification of time series using machine learning techniques: applications*

AUTOR: *David Guijo Rubio*

© Edita: UCOPress. 2021
Campus de Rabanales
Ctra. Nacional IV, Km. 396 A
14071 Córdoba

https://www.uco.es/ucopress/index.php/es/
ucopress@uco.es

UNIVERSITY OF CÓRDOBA

Clustering, prediction and ordinal classification of time series using machine learning techniques: applications

International PhD Mention
PhD Program: Advanced Computing, Energy and Plasmas

David Guijo Rubio

Supervisors
César Hervás Martínez
Pedro Antonio Gutiérrez Peña

Department of Computer Science and Numerical Analysis
Córdoba, April 2021

La memoria titulada “*Clustering, prediction and ordinal classification of time series using machine learning techniques: Applications*”, que presenta D. David Guijo Rubio para optar al Título de Doctor, ha sido realizada dentro del programa de doctorado “Computación Avanzada, Energía y Plasmas” del Departamento de Informática y Análisis Numérico de la Universidad de Córdoba bajo la dirección del Doctor D. César Hervás Martínez y del Doctor D. Pedro Antonio Gutiérrez Peña.

El doctorando D. David Guijo Rubio y los directores de la Tesis D. César Hervás Martínez y D. Pedro Antonio Gutiérrez Peña garantizamos, al firmar esta Tesis Doctoral, que el trabajo ha sido realizado por el doctorando, bajo la dirección de los directores de la Tesis y que, hasta donde nuestro conocimiento alcanza, en la realización del trabajo, se han respetado los derechos de otros autores a ser citados, cuando se han utilizado sus resultados o publicaciones.

Córdoba, 22 de marzo de 2021

El doctorando

Firmado por GUIJO RUBIO DAVID - 45886352H el día 22/03/2021 con un certificado emitido por AC FNMT Usuarios

Fdo: David Guijo Rubio

Los directores

Firmado por HERVAS MARTINEZ CESAR - 04528484Z el día 22/03/2021 con un

Fdo: César Hervás Martínez

GUTIERREZ PEÑA PEDRO ANTONIO - 45735457A

Firmado digitalmente por GUTIERREZ PEÑA PEDRO ANTONIO - 45735457A
Nombre de reconocimiento (DN): c=ES, serialNumber=IDCES-45735457A, givenName=PEDRO ANTONIO, sn=GUTIERREZ PEÑA, cn=GUTIERREZ PEÑA PEDRO ANTONIO - 45735457A
Fecha: 2021.03.22 20:23:04 +01'00'

Fdo: Pedro Antonio Gutiérrez Peña

TÍTULO DE LA TESIS: Agrupamiento, predicción y clasificación ordinal para series temporales utilizando técnicas de *machine learning*: Aplicaciones.

DOCTORANDO: David Guijo Rubio

INFORME RAZONADO DE LOS DIRECTORES DE LA TESIS

(se hará mención a la evolución y desarrollo de la tesis, así como a trabajos y publicaciones derivados de la misma).

En su Tesis, D. David Guijo Rubio propone métodos novedosos asociados al análisis de series temporales con el objetivo de resolver diferentes problemas del mundo real. Concretamente, el tema general de la Tesis está centrado en metodologías del campo del aprendizaje automático (ML), que podría dividirse en varias subáreas en función de la tarea a abordar o la metodología a aplicar.

En este sentido, temas más específicos son la clasificación ordinal, las redes neuronales artificiales, los algoritmos evolutivos y muchas y variadas aplicaciones relacionadas con problemas del mundo real: predicción y detección de niebla y de situaciones convectivas, ambos problemas desarrollados en aeropuertos españoles, pronóstico y determinación de altura de olas, predicción de energía solar, y modelos de asignación donante-receptor en trasplante hepático utilizando la base de datos americana UNOS, entre otros problemas analizados.

Específicamente, los enfoques de ML se aplican a problemas relacionados las series temporales. Las series temporales son un tipo especial de datos en los que los datos muestrales se recopilan cronológicamente. Las series temporales están presentes en una amplia variedad de campos, que van desde problemas asociados al clima y a la economía y las finanzas, hasta aplicaciones en ingeniería. Existen diferentes tareas en la literatura aplicadas a las series temporales, algunas de ellas son aquellas en las que se centra principalmente esta Tesis, como son: el agrupamiento, la clasificación, la predicción y, en general, su análisis.

El trabajo desarrollado en esta Tesis está respaldado por 10 artículos indexados por JCR en revistas internacionales (7 Q1, 2 Q2, 1 Q3), 7 artículos en congresos internacionales y 5 artículos en congresos nacionales. Además, se está revisando un trabajo enviado a otra revista internacional. Queda patente de esta forma, la calidad científica de las contribuciones de la Tesis. Esto nos lleva a presentar la Tesis como compendio de artículos.

Por todo ello, se autoriza la presentación de la tesis doctoral.

Córdoba, 22 de marzo de 2021

Firma de los directores

Firmado por HERVAS MARTINEZ CESAR - 04528484Z el día 22/03/2021 con un

Fdo.: César Hervás Martínez

GUTIERREZ PEÑA PEDRO ANTONIO - 45735457A

Firmado digitalmente por GUTIERREZ PEÑA PEDRO ANTONIO - 45735457A
Nombre de reconocimiento (DN): c=ES, serialNumber=IDCES-45735457A, givenName=PEDRO ANTONIO, sn=GUTIERREZ PEÑA, cn=GUTIERREZ PEÑA PEDRO ANTONIO - 45735457A
Fecha: 2021.03.22 19:49:54 +01'00'

Fdo.: Pedro Antonio Gutiérrez Peña

Esta Tesis Doctoral ha sido financiada en parte con cargo a los Proyectos **TIN2014-54583-C2-1-R** y **TIN2017-85887-C2-1-P** del Ministerio de Ciencia, Innovación y Universidades (MICINN), al Proyecto **UCO-1261651** de la Consejería de Transformación Economía, Industria, Conocimiento y Universidad de la Junta de Andalucía, con fondos FEDER, y con el programa de Formación del Profesorado Universitario FPU, referencia **FPU16/02128**, del Ministerio de Educación, Cultura y Deporte / Ministerio de Ciencia, Innovación y Universidades. Las estancias de investigación para la obtención de la Mención Internacional han sido financiadas, por un lado, por el programa de Formación del Profesorado Universitario FPU, bajo ayuda con referencia **EST18/00280** y por otro lado, por el programa de Fomento de Tesis con Mención Internacional de la Universidad de Córdoba, bajo ayuda con referencia **2019/00213**.

This Doctoral Thesis has been partially subsidised by the **TIN2014-54583-C2-1-R** and **TIN2017-85887-C2-1-P** projects of the Spanish Ministry of Science, Innovation and Universities (MICINN), by the **UCO-1261651** project of the Department of Economic Transformation, Industry, Knowledge and University of the Andalucía's Government, FEDER funds, and by the FPU Predoctoral Program of the Spanish Ministry of Education, Culture and Sport / Ministry of Science, Innovation and Universities, grant reference **FPU16/02128**. The research stays have been subsidised, on the one hand, by the FPU Predoctoral Program of the Spanish Ministry of Education, Culture and Sport / Ministry of Science, Innovation and Universities, grant reference **EST18/00280**, and, on the other hand, by the Program for Promotion of International Doctorate of the University of Córdoba, grant reference **2019/00213**.

GOBIERNO DE ESPAÑA
MINISTERIO DE CIENCIA, INNOVACIÓN Y UNIVERSIDADES
GOBIERNO DE ESPAÑA
MINISTERIO DE EDUCACIÓN, CULTURA Y DEPORTE
FEDER
Fondo Europeo de Desarrollo Regional
Junta de Andalucía
Consejería de Transformación Económica, Industria, Conocimiento y Universidades
UNIVERSIDAD D CÓRDOBA

Mención de Doctorado Internacional

Esta Tesis cumple los criterios establecidos por la Universidad de Córdoba para la obtención del Título de Doctor con Mención Internacional. Para ello se presentan los siguientes requisitos:

1. Estancias predoctorales realizadas en otros países europeos:
    - **School of Computing Sciences, University of East Anglia, Norwich, Reino Unido.** Duración de tres meses desde el 1 de marzo hasta el 1 de junio de 2019. Tutor de la estancia: **Dr. Anthony Bagnall**, *Full professor*.
    - **School of Computing Sciences, University of East Anglia, Norwich, Reino Unido.** Duración de tres meses desde el 1 de septiembre hasta el 1 de diciembre de 2019. Tutor de la estancia: **Dr. Anthony Bagnall**, *Full professor*.
2. Esta Tesis está avalada por los siguientes informes de idoneidad realizados por doctores de otros centros de investigación internacionales:
    - **Dr. Romain Tavenard**. *Assistant Professor* of School of Computer Science and Statistics, Université de Rennes (France).
    - **Dr. Mario Gongora**. *Principal Lecturer* of School of Computer Science and Informatics, De Montfort University (Reino Unido).
3. La defensa de la Tesis y el texto se han realizado completamente en inglés. Entre los miembros del tribunal se encuentra un doctor procedente de un centro de educación superior europeo, tratándose del Dr. **David Elizondo**, *Professor* of School of Computer Science and Informatics, De Montfort University (Reino Unido).

La raíz de todo bien reposa en la tierra de la gratitud.

Dalai Lama

Agradecimientos

A mis directores, Pedro y César, por despertar en mí su pasión por la inteligencia artificial y la investigación, y por transmitirme su entusiasmo por la docencia.

A mis padres, Segundo e Isabel, por enseñarme a ser y a vivir.

A mi hermana, Isabel, por mostrarme de forma sincera el significado del cariño y el amor.

A mi familia, por su cercanía y afecto invaluable.

A mis amigos y amigas, por su entrega y por ser un apoyo incondicional.

A mis compañeros del Grupo AYRNA, por evidenciar que la cooperación es el motor de cualquier progreso.

A todos y todas los que aportásteis un granito de arena, gracias por haber participado en un maravilloso camino lleno de aprendizaje.

A todos y todas, gracias de corazón.

Contents

List of Figures

Acronyms

A

ACF autocorrelation function. 25, 26
AI artificial intelligence. 1, 2
AIC Akaike information criterion. 26
AMAE average mean absolute error. 61, 82, 84, 103
ANN artificial neural network. III, 2, 3, 8, 13, 16–20, 22, 39, 41–43, 45, 47, 67, 68, 70, 88, 91, 94, 107
AR autoregressive model. 24, 25, 28, 36, 47, 50, 59, 61, 102, 103
ARI adjusted rand index. 12
ARIMA autoregressive integrated moving average model. 24, 25, 59
ARMA autoregressive moving average model. 24, 25, 39, 59

B

BIC bayesian information criterion. 26
BOP bag of patterns. 28
BOSS bag of symbolic Fourier approximation symbols. 28
BP back-propagation. 18, 41

C

CAMS Copernicus atmosphere monitoring service. 39, 70, 104
CCR correct classification rate. 19, 77, 82, 84, 94, 104
CH Caliński-Harabasz index. 12, 57
CNN convolutional neural network. 33, 34, 49, 78, 79, 84, 102, 105, 109
COTE collective of transformation-based ensemble. 28

D

DB Davies Boulding index. 12

DL deep learning. 29, 33, 39, 46, 102, 109
DT decision tree. 3
DTW dynamic time warping. 27, 28, 33
DU dunn index. 12
DW desiccant wheel. 41, 48, 88, 107, 109

E

E cross-entropy. 19
EA evolutionary algorithm. 18–20, 22, 47, 67
EANN evolutionary artificial neural network. 19, 47, 50, 68, 70, 71, 73, 107
EE elastic ensemble. 28
ELM extreme learning machine. 39, 40, 71, 74, 104

F

FNN feed-forward neural network. III, 13, 14, 16, 17
FS fast shapelets. 28

G

GADF gramian angular difference field. 34
GASF gramian angular summation field. 34
GB gradient boosting. 98, 108
GP Gaussian process. 3

H

HCV hepatitis C virus. 36, 42, 43, 48, 93, 94, 107, 109
HIV human immunodeficiency virus. 36, 42, 43, 48, 93, 94, 107, 109
HIVE-COTE hierarchical vote system collective of transformation-based ensemble. 28, 29, 46, 77–79, 84, 105, 106
HMM hidden Markov models. 28, 33

I

IG information gain. 30, 82, 106

K

KDLOR kernel discriminant analysis for ordinal regression. 3, 9, 61
KM kernel model. 28

L

LR logistic regression. 78, 79, 98, 105
LS learned shapelets. 28, 78, 105
LSTM long-short-time memory. 39, 102
LT liver transplantation. 2, 42, 43, 48, 96, 98, 107–109

M

MA moving average model. 24, 25, 59, 102
MELD model for end-stage liver disease. 43, 98, 108
ML machine learning. III, 2–5, 13, 23, 38–40, 43, 45, 47, 48, 70, 71, 74, 87, 93, 96, 98, 108, 109
MLP multilayer perceptron. 16, 36, 71, 98, 104
MOEA multi-objective evolutionary algorithm. 20, 22
MOEANN multi-objective evolutionary artificial neural network. 20–23, 50, 68, 104
MOP multi-objective problem. 20, 21, 23, 47
MS minimum sensitivity. 19, 38, 61, 68, 94, 103, 104
MSE mean squared error. 19, 73, 88, 91
MTEANN multi-task evolutionary artificial neural network. 40–42, 73, 74, 87–89, 91, 104, 107
MTF Markov transition field. 34

N

NCAR national center for atmospheric research. 40, 73
NDBC national data buoy center. 73
NSGA non-dominated sorting genetic algorithm. 22
NSGA-II non-dominated sorting genetic algorithm II. 22
NWP numerical weather prediction. 38, 39

O

OC ordinal classification. III, 2, 5–7, 9, 26, 37, 38, 49, 50, 59, 61, 67, 103, 106
OR ordinal regression. 3, 5

P

PAA piecewise aggregation approximation. 34
PACF partial autocorrelation function. 25, 26
POM proportional odds model. 3, 9, 82, 84, 106
PU product unit. 15–17, 68, 70, 73, 88, 91, 94, 104, 107
PUNN product unit neural network. 16
PWID people who inject drugs. 94

R

RBF radial basis function. 14, 16, 17, 68, 70, 73, 94, 104, 108
RBFNN radial basis function neural network. 16
RedSVM reduction applied to support vector machine. 9
Resnet residual networks. 29, 46
RF rotation forest. 79, 82, 105
RI rand index. 12, 57
RP recurrence plot. 34
RPMCNN relative position matrix and convolutional neural network. 29
RVR runway visual range. III, 37, 61, 103

S

SA survival analysis. 96, 108
SEP standard error of prediction. 73, 89, 91
SGD stochastic gradient descent. 105
SI silhouette index. 12
SSE sum of squared error. 13
ST shapelet transform. 28–30, 47–49, 78, 79, 81, 82, 84, 105, 106, 108
SU sigmoidal unit. 15–17, 68, 70, 73, 88, 91, 94, 104, 107
SUNN sigmoidal unit neural network. 16
SVM support vector machine. 3, 9, 42, 96, 98
SVOREX support vector for ordinal regression with explicit constraints. 3, 9
SVORIM support vector for ordinal regression with implicit constraints. 3, 9, 82, 84, 106
SVR support vector regressor. 71, 74, 103, 104, 106

T

TAF terminal aerodrome forecast. 65, 68, 103, 104
TS-CHIEF time series combination of heterogeneous and integrated embedding forest. 28, 84, 106
TSBF time series bag of features. 28
TSC time series classification. III, 26–29, 34, 35, 45, 47–49, 56, 57, 77–79, 81, 82, 84, 102, 105, 106, 108, 109
TSF time series forest. 28
TSOC time series ordinal classification. 26, 29, 45, 47–49, 77, 81, 82, 84, 106, 108, 109

U

UEA/UCR University of East Anglia and University of California Riverside. 49, 56, 57, 79, 81, 82, 102, 105

UNOS united network for organ sharing. 96, 98, 108

W

WDTW weighted dynamic time warping. 27

I do not fear computers. I fear the lack of them.

Isaac Asimov

1

Introduction

In the last decade, artificial intelligence (AI) has been a hot-topic in the day-to-day life, significantly increasing its interest over time. This wide area consists in the study of those devices able to perceive their environment, collect data, and, in consequence, perform an action to maximise the chances of success [155]. Nowadays, it is difficult to think in a field of science or, in general, in daily life, in which data generated in the different processes is not collected. Moreover, the massive growth of this data makes the extraction of knowledge a process only feasible for automatic techniques, given that this extraction is not within the reach of human beings.

Commonly, AI is designed to improve the quality of our lives: not only does it make our lives better, but also it is able to make them more comfortable. There are countless applications of AI in the health field improving the associated processes. For instance, in comparison to humans, AI is able to analyse radiological scans up to $10,000$ times faster than a radiologist[1], or to provide advice on how to create a better workout by monitoring both the heartbeats and 3D movement[2].

In the end, AI is a novel science making computers, robots, and, in general, any device, to think as human beings. It tries to simulate the way human brains think, work, and make decisions in every situation. AI aims to solve a wide range of problems: knowledge

[1] `https://www.enlitic.com/`

[2] `https://welcome.moov.cc/`

representation, natural language processing or learning, among others.

This Thesis proposes novel methods for time series data mining aiming to solve different real-world problems. Specifically, this Thesis is framed in the machine learning (ML) field, which can be divided in several sub-areas depending on the task to be tackled or the methodology to be applied. In this sense, more specific subtopics of ML covered in this Thesis include ordinal classification (OC), artificial neural networks (ANNs), time series, and many and varied applications related with real-world problems: fog prediction and detection of convective situations, both in airports, forecasting for wave and solar energy, and donor-recipient matching in liver transplantation (LT), among others.

1.1 Machine learning

Artificial intelligence (AI) consists of several wide sub-areas, and, in this Thesis, we focus on the process of learning from experience, which is also known as machine learning (ML). More specifically, ML is the field of study dealing with automatic techniques or algorithms able to improve their performance automatically through experience (expressed in the form of existing examples of data or patterns) [26]. Besides, regarding the kind of data analysed in this Thesis, we focus on time series, which can be defined as temporal data in which data points are collected chronologically.

These automatic techniques (also known as ML techniques) can be classified following different criteria: according to the specific reasoning approach, according to the type of input/output data, or according to the task or problem to be solved. In this Thesis, the latter is the criterion considered. In this way, ML approaches can be divided into these learning problems: supervised, unsupervised, semi-supervised and reinforcement learning [26]. Figure 1.1.1 shows a taxonomy with these main four paradigms of ML techniques.

Specifically, these four types of learning can be described as follows:

- Supervised learning is the learning problem in which the examples of input data (commonly represented as a matrix of vectors of features or attributes $\mathbf{X}$) have a desired output value [129]. The main goal of this sort of learning is to accurately learn a mapping function from the set of input attributes to the output, using just the training set, which is obtained from historical data. Once the learning function has been trained, it will be able to predict the output value of the examples belonging to the test dataset (unseen data during the training step). This learning function is assessed iteratively during the training step, aiming to predict, as accurate as possible, new unknown examples.

  Regarding the output variable, the following subdivision can be distinguished:

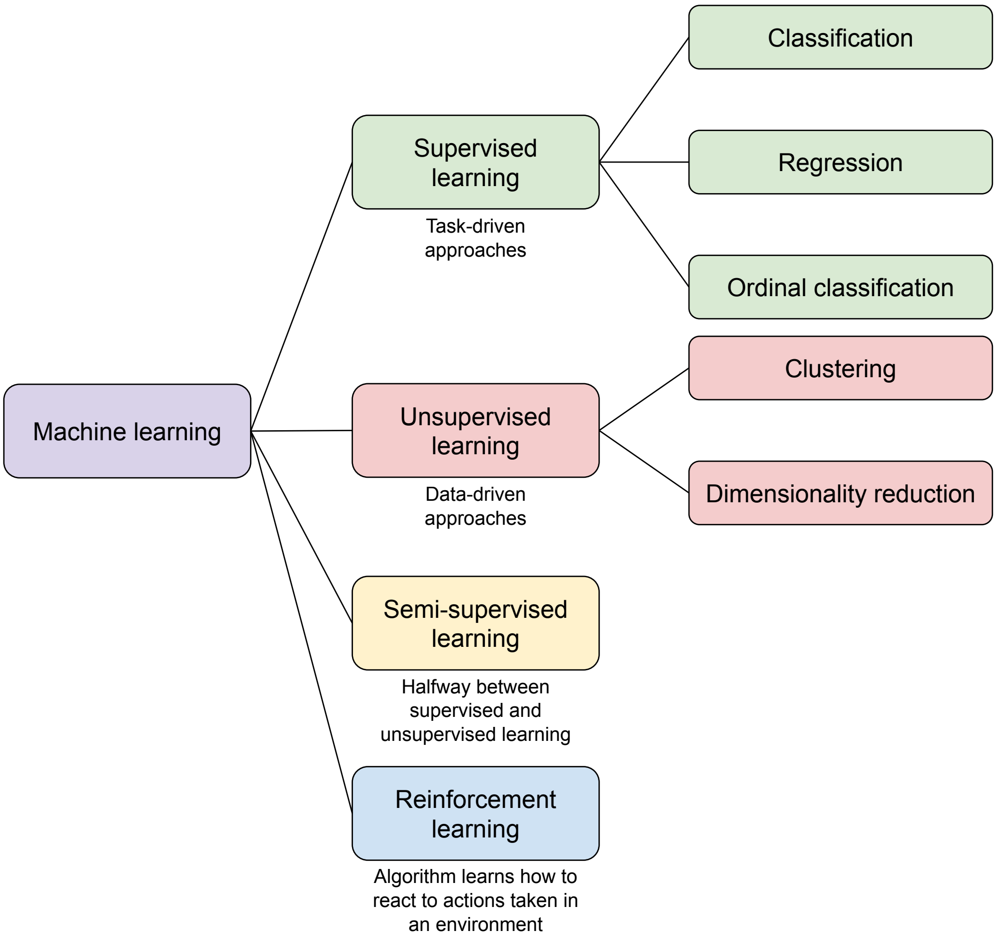

Figure 1.1.1: Paradigms of ML techniques.

- Regression is the task in which the output variable Y can only take real values. Linear [131], ridge [91] or lasso [173] regressions are some of the most popular techniques belonging to this group.
- Classification consists in predicting a discrete or nominal output variable $Y \in \mathcal{Y}$. Given that it is the most used technique in ML, there are a wide range of approaches included in this group: support vector machines (SVMs) [45], Gaussian processes (GPs) [147], decision trees (DTs) [142], or artificial neural networks (ANNs) [197], among others.
- Ordinal classification is an emerging field, belonging to the classification area, with the particularity that categories follow a natural order relationship among them. It is also known as ordinal regression (OR). The better known approaches of this group are: proportional odds model (POM) [125], kernel discriminant analysis for ordinal regression (KDLOR) [170], support vector for ordinal regression with implicit constraints (SVORIM), and support vector for ordinal

regression with explicit constraints (SVOREX) [39], to mention a few.

- Unsupervised learning is the learning problem applied to datasets whose examples only include input data (for notation simplicity, the input data is also represented as $\mathbf{X}$) [90]. In other words, the examples are not labelled according to ground-truth, therefore, there is not a division of the dataset into training and test sets. The main goal of unsupervised learning is twofold: to determine groups of data presenting a similar structure among them, and to reduce the dimensionality of the data.

  Depending on the goal, the following subdivision can be distinguished:

  - Clustering is the unsupervised task aiming to group the patterns according to the similarities of the input characteristics $\mathbf{X}$. Based on the way the clusters are formed, there are numerous clustering methods: partitional [33], hierarchical [104] or density-based algorithms [110], among others.
  - Dimensionality reduction is the task aiming to simplify the complexity of the model and avoid overfitting by reducing the number of variables, characteristics or features of the dataset. There are two main categories: feature selection, where the goal is to choose a subset of variables or features [82], and feature extraction, whose aim is to derive the features in order to build a new subspace [83].

- Semi-supervised learning is a mixed variant making use of both supervised and unsupervised learning approaches [204]. This paradigm is applied to those datasets combining, generally, a small amount of labelled (or supervised) patterns with a huge amount of unlabelled (or unsupervised) examples. Specifically, this sort of learning is applied when the acquisition of labelled data is an arduous task involving huge costs. Thus, having some supervised data is of great practical value.

- Reinforcement learning is based on how software agents maximise the cumulative rewards by taking actions in a concrete environment [171]. The goal is to learn which actions are the most worthy to increase the reward. It differs from supervised learning in that the patterns do not need to be labelled, and it differs from unsupervised learning in that it does not aim to find a hidden structure on the data.

To sum up, ML is a field of science that aims to extract relevant information and knowledge from processes present in almost all the fields. However, depending on the problem to be solved, and the amount and sort of available data, different algorithms can be applied. ML techniques can be employed in several applications, raising special attention to those problems infeasible to be solved by standard statistical techniques. ML has been used in a wide range of applications, from health (donor-recipient matching in order to increase the survival time) through technology-related applications (face recognition systems aiming to identify human faces from images or video frames) to weather forecasting (predicting whether an airport is able to properly operate under certain future conditions).

In this Thesis, several ML techniques have been applied to different problems, not only to solve them, but also as a baseline to check if the proposed methodologies achieve better results than previous approaches. Specifically, we have primarily focused on the first two types of learning described above (supervised –regression, classification, and ordinal classification– and unsupervised –clustering–).

1.2 Ordinal classification

Classification algorithms learn a mapping function from a set of features to a categorical target variable, also known as class attribute. Particularly, there are countless applications involving tasks for whose the order among the labels is of significant impact, such as people's age estimation [35, 134] or predicting the severity's degree of an illness [193]. Figure 1.2.1 shows examples from the convective clouds formation prediction problem considering 4 different ordinal classes. These problems could be solved using a standard regressor, where the goal is predicting the classes as integer values ("1" for the first class, "2" for the second class...). However, the main disadvantage of the standard regression paradigm is assuming that all the values in the class attribute are separated by the same distance [189].

In this way, it is interesting to consider this kind of learning problems from the ordinal point of view, which is commonly referred in the literature as ordinal classification (OC) or ordinal regression (OR) [81]. OC, as standard classification, tries to find a mapping function from the features to the class attribute $Y \in \mathcal{Y} = \{C_1, C_2, \ldots, C_J\}$, where $C_1, C_2, \ldots, C_J$ are the class labels. However, as previously stated, the main difference is that OC techniques take into account the order relationship between the labels, i.e. $C_1 \prec C_2 \prec \ldots \prec C_J$. In this way, for OC, the greater the distance is between the real and the predicted class values, the greater the misclassification cost is [13] (although a specific distance between labels is not assumed).

In order to get a deeper insight into these misclassification errors and their associated costs, we have the following example. Lets consider a disease with the following stages: $none \prec mild \prec moderate \prec severe \prec extreme$. A standard classifier will penalise equally all the misclassifications, this is, predicting a person whose disease's stage is $none$ as $mild$ or as $extreme$ is equally penalised, given that neither the order of the labels is taken into account nor the distance between them. However, obviously, there is a signifi-

Figure 1.2.1: Examples taken from a convective clouds formation prediction problem. Images are ordered by the class they belong, from left to right and top to bottom.

cant difference between both misclassification errors, and therefore, the use of an ordinal classifier is a much better approach. In this case, the ordinal classifier will penalise differently both errors, giving more importance to misclassifying a person whose disease's stage is $none$ as $extreme$ than misclassifying it as $mild$.

More formally, the OC paradigm could be defined as follows: let consider a set of N points, $D = \{\mathbf{X}, \mathbf{y}\}$, where $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$ and $\mathbf{y} = \{y_1, y_2, \ldots, y_N\}$, being $\mathbf{x}_i \in X \subseteq \mathbb{R}^K$ the input feature vectors and $y_i \in \mathcal{Y} = \{C_1, C_2, \ldots, C_J\}$ the label, i.e. $\mathbf{x}_i$ is the i-th pattern taking values in a K-dimensional input space and y_i takes values in a label space of J different labels. The goal of an ordinal classifier is to learn a classification function $f(\mathbf{x}) : X \to \mathcal{Y}$, where $\mathbf{x} \in X$, by using a training set D, able to predict accurately the categories of new patterns. Under the OC paradigm, a natural label ordering is assumed, i.e. $C_1 \prec C_2 \prec \ldots \prec C_J$, where the operator $\prec$ represents an order relationship. There are several OC techniques and metrics considering the rank of the label, in other words, the position of a given label in the ordinal scale $\mathcal{Y}$. This position can be expressed by the function $\mathcal{O}(\cdot)$, so that $\mathcal{O}(C_j) = j$, where $j \in 1, 2, \ldots, J$.

Previous to the use of specific OC techniques, these problems were solved by using both nominal classification and regression approaches. The former did not consider the order information between the labels, whereas the latter considered $Y \in \mathbb{R}$, and hence, the real values in $\mathbb{R}$ are ordered by the operator $<$, replacing the qualitative information of the labels Y by quantitative information. Therefore, the use of ordinal classifiers avoids both problems and leads to better performances [81]. There are several ordinal classifiers that could be organised according the taxonomy shown in Figure 1.2.2.

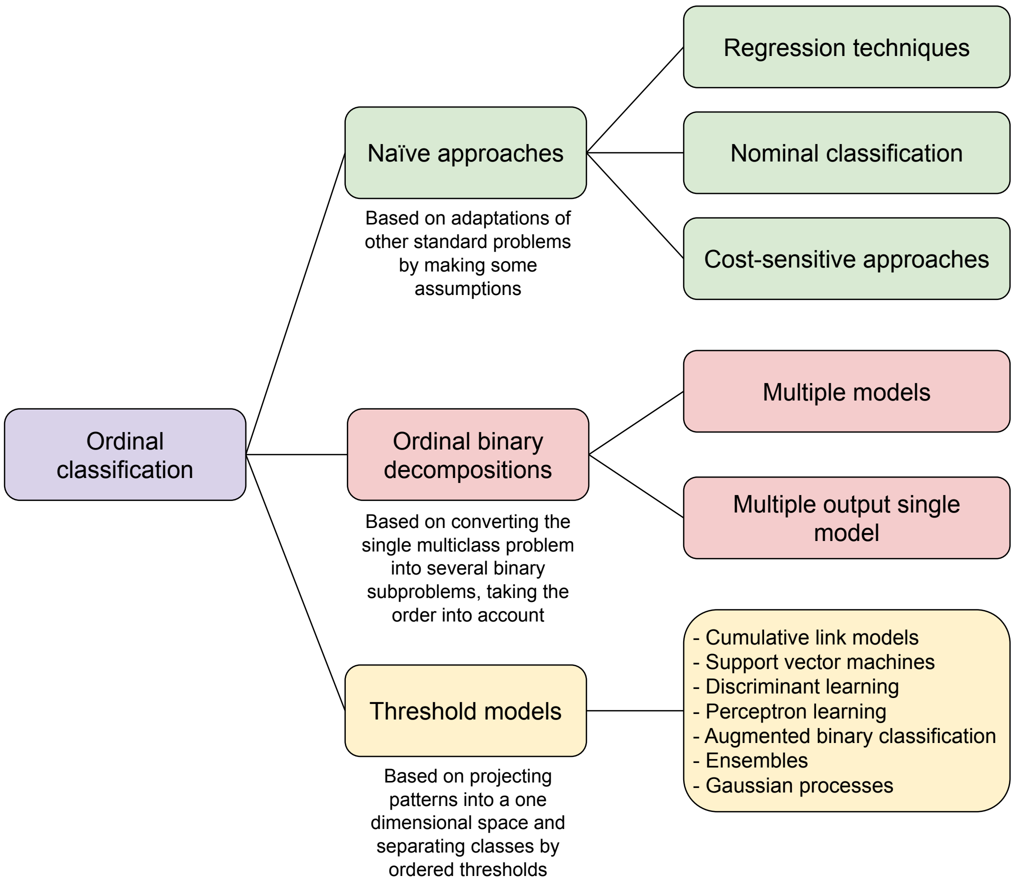

Figure 1.2.2: OC taxonomy.

According to this taxonomy, three different groups of methods can be established:

- Naïve approaches are simplifications or adaptations of well-known methods from the nominal classification or regression literature. The following subdivision can be made:
  - Regression techniques can be adapted to the ordinal paradigm by mapping all the different labels into real values, typically related with their position in the ordinal scale [109, 174].
  - Nominal classification methods can be applied to ordinal datasets by simply ignoring the order of the labels. Generally, this behaviour leads these methods to require more data [4, 84].
  - Cost-sensitive approaches introduce the concept of giving different penalisation

to the misclassification errors [108, 176]. However, this group of methods has the difficulty of choosing the appropriate costs for each error.

- Ordinal binary decomposition methods decompose the ordinal target problem into binary subproblems, which are then estimated by single or multiple models. Generally, there are two different ideas:
    - ◦ Multiple models, in which a different model is trained for each subproblem. After that, the output of all the independent binary models is put in common and combined into an ordinal output [66].
    - ◦ Multiple output single model, in which a single model is trained for all the decomposed problems. In other words, models pertaining to this group can manage to solve each subproblem using a single structure. The best example is the use of artificial neural networks (ANNs), where each output neuron can solve a different subproblem [46].
- Threshold Models are, however, the most popular techniques. They consider the task as one itself and assume there is an unobserved latent variable representing the labels of the problem in a continuous function. These methods need to estimate both a function $f(x)$ that predicts the latent variable and a set of $J-1$ thresholds that define intervals (one for each class), taking into account the ordinal target information. Figure 1.2.3 shows an example of test patterns in a 3-class ordinal problem, where crosses represent correctly classified patterns and circles represent misclassification errors.

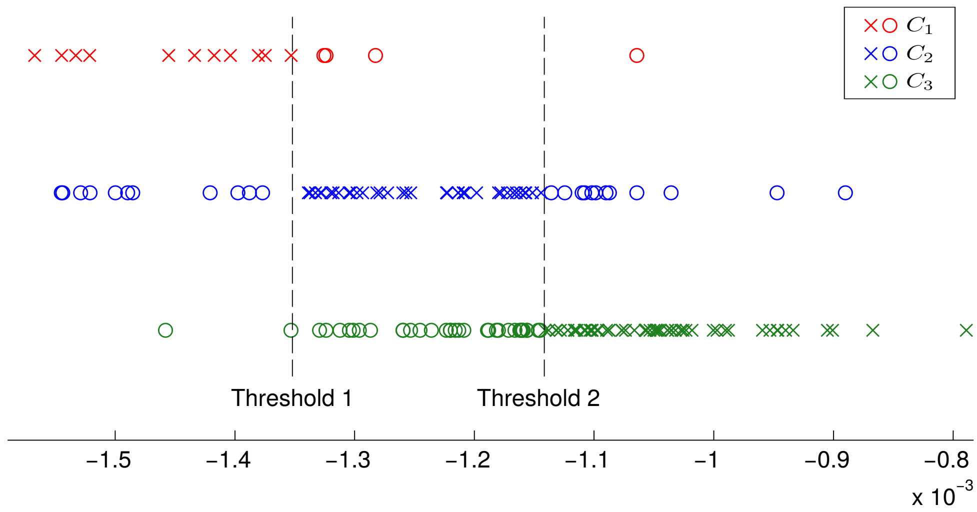

Figure 1.2.3: Projection of patterns according to the thresholds learnt from data.

There are several methods belonging to this category, but the most popular ones are the following:

- ◦ Proportional odds model (POM) [125], a linear model which was the first specifically-designed approximation to OC.
- ◦ Adaptations of support vector machines (SVMs) to OC. There are several different SVM-based techniques: support vector for ordinal regression with implicit constraints (SVORIM) [39], support vector for ordinal regression with explicit constraints (SVOREX) [39] or reduction applied to support vector machine (RedSVM) [114], among others.
- ◦ Kernel discriminant analysis for ordinal regression (KDLOR) [170] is based on discriminant learning, which is easily adapted to OC given that the definition of the thresholds can be used to discriminate between the classes.

1.3 Clustering

One of the main tasks of exploratory data mining is clustering, which consists in grouping a set of objects by how similar they are. In this way, similar objects are grouped in the same group (also called cluster). Apart from being an exploratory technique, clustering is also a well-known method for statistical data analysis with several paramount applications: pattern recognition [19] or information retrieval [191], among others.

There is a wide range of clustering techniques that differ in the way clusters are found [59]. However, the main objective of all the clustering methods is to minimise the intra-cluster distance, this is, the distance between all the objects belonging to the same cluster, while maximising the distance between the different clusters. In this way, Figure 1.3.1 shows an example of a cluster analysis applied to a synthetic problem. As can be seen in the figure, the three clusters are distant from the others and there are dense areas of objects close to the centroid, which is the most representative object of a cluster (it could be an existing object –$\mathcal{G}_3$–, or, usually, the mean or median of all the objects belonging to the cluster –$\mathcal{G}_1$ and $\mathcal{G}_2$–).

The main notion of what a cluster is varies significantly in its properties. Figure 1.3.2 shows a taxonomy with the main cluster models existing in the literature. In the taxonomy, we have only included the most popular clustering algorithms given that there are more than 100 published clustering algorithms, and, as the notion of clustering varies, its categorisation is not trivial[3].

Therefore, the main clustering methods can be described as follows:

- Partitional clustering [33] consists in finding a given number of clusters k, by assigning the objects to the nearest cluster centroid. The main two algorithms belonging to

[3] Further information about clustering techniques and taxonomies can be found in [192].

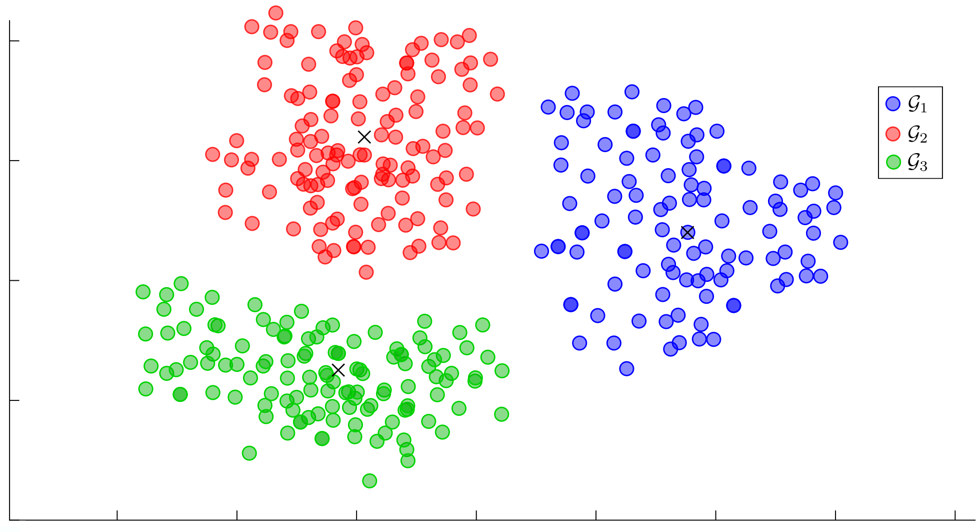

Figure 1.3.1: Clusters obtained for a synthetic problem. Centroids are represented by crosses.

this group are k-means [122] and k-medoids [105]. The main advantages of these iterative methods are their low computational cost and that they partition the data space into a regional structure (known as Voronoi diagram). However, they have several drawbacks, such as requiring the number of clusters k in advance or that they can only find a local optimum. Moreover, these two methods are stochastic, i.e. for two consecutive runs the results will vary, due to random initialisation.

- Hierarchical clustering [152] is based on finding clusters by building a hierarchy of them, also known as dendrogram. There are two main approaches [104]: 1) agglomerative, based on the bottom-up paradigm, in which each object starts in its own cluster, and, in each iteration, a pair of clusters are merged. And 2) divisive, based on the top-down paradigm, in which all the objects start in the same cluster, and splits are performed in each iteration. These methods require the use of a distance function or a similarity measure and a linkage criterion determining the way the distance between two clusters is computed [132]. Finally, the main disadvantages of these methods are the time and memory complexities. The main hierarchical clustering algorithms are: 1) BIRCH [203], which follows an agglomerative strategy; it firstly generates a small and compact summary of the dataset, which is what is clustered, instead of clustering the raw data. 2) CURE [78], which is another hierarchical clustering algorithm following the agglomerative strategy; it consists in adopting a middle ground between the cluster centroid and all objects located in the extremes. And, 3) Bisecting k-means [103], which is the most representative algorithm following the divisive scheme; it is based on the well-known k-means clustering method, but instead of partitioning the data into k clusters, it splits one cluster into two

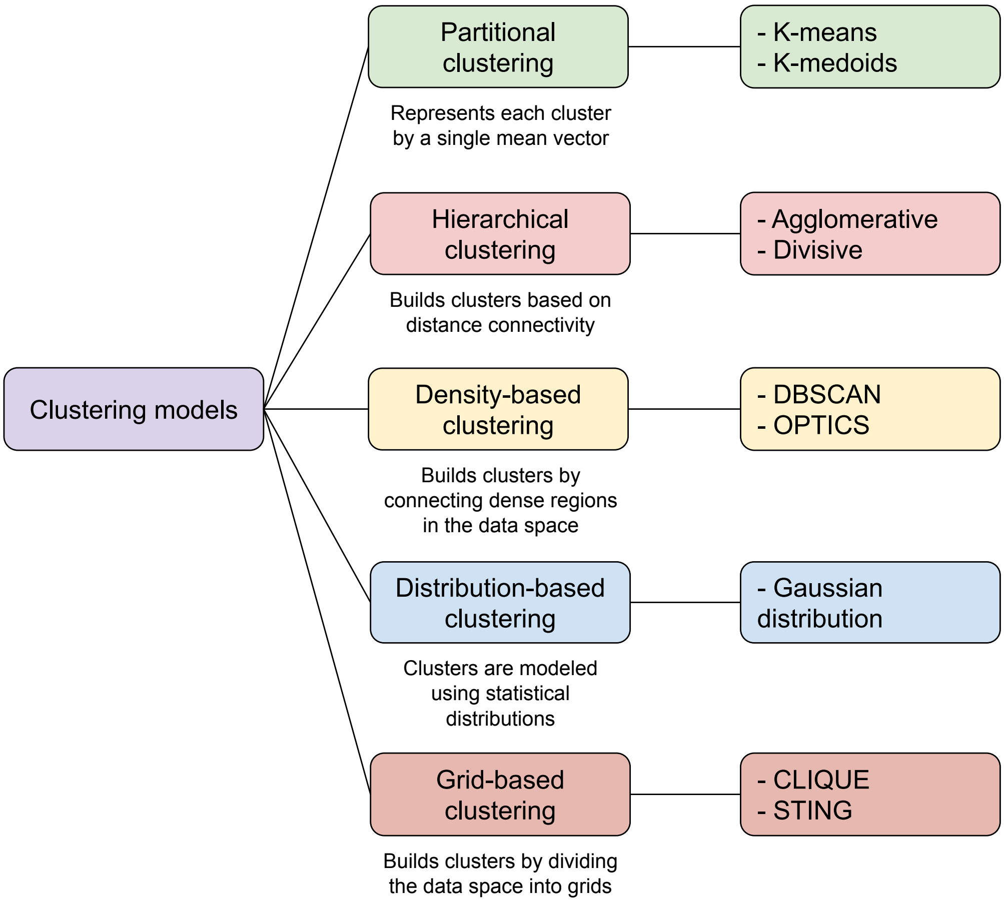

Figure 1.3.2: Clustering models taxonomy.

clusters in each bisecting step, until k clusters are obtained.

- Density-based clustering [110] defines clusters by areas whose density of points is higher than the remaining areas of the data space. This method avoids creating a cluster for objects in sparse areas or border points, considering them as outliers. The most popular methods are: 1) DBSCAN [58], which is based on grouping objects by their distance. However, they need to satisfy a density criterion (the original criterion is to include a minimum number of objects within a given radius). And, 2) OPTICS [9], which is a generalisation of DBSCAN removing the need to introduce the value of the radius beforehand.

- Distribution-based clustering [70] assumes that data objects follow a distribution, being the Gaussian distribution one of the most widely used. The more the distance from an object to the distribution centre, the less probability an object has to belong to that cluster. The main disadvantage of these methods is that they suffer from the overfitting problem, given that choosing a more complex distribution will usually explain the data better.

- Grid-based clustering [70] is a simple way to cluster data objects. It defines a grid structure, and data objects belong to that grid (also known as cell). There are two main approaches: 1) CLIQUE [3], in which clustering starts at single-dimension subspace and move upwards towards higher dimension subspaces, and, 2) STING [182], which processes many common region oriented queries on a set of objects, working as a hierarchical clustering method. However, the main disadvantage of both techniques is that they penalise their accuracy to increase the simplicity of the methods.

Assessing the quality of the clustering is a fundamental step of the clustering process [107]. For this purpose, a wide range of specific evaluation metrics have been presented in the literature [149]. They can be divided into two different categories:

- External measures [20], which make use of the class labels (also known as ground truth, if they are available) for evaluating the clusters extracted. It is important to clarify that the ground truth is not used during the clustering stage. Rand index (RI) [144] is the most common external evaluation metric. It penalises false positive and false negative decisions during clustering. An improvement (or correction) to the RI was introduced by Vinh *et al.* in [179], proposing the adjusted rand index (ARI). This external evaluation metric introduces a correction by using the expected similarity of all pair-wise comparisons between clusters specified by a random model.
- Internal measures [10], that assesses the goodness of the clusters extracted according to different criteria (depending on the metric to be used), but without using the ground truth. The most common internal evaluation metrics are the following[4]:
  - Caliński-Harabasz index (CH) [31] is a radio-type index, in which the cohesion is estimated based on the intra-cluster distance (i.e. the distances from the cluster centroid to the cluster objects), whereas the separation is based on the distance from all the cluster centroids to the global centroid.
  - Davies Boulding index (DB) [49] attempts to maximise the distance between clusters while minimising the intra-cluster distances.
  - Dunn index (DU) [55] aims to find compact and well-separated clusters. It is sensitive to noise, however, different corrections have been proposed in the literature to avoid this issue, by considering different notions of cluster distance or cluster diameter.
  - Silhouette index (SI) [153] measures the goodness of a cluster structure by computing the cohesion as the intra-cluster distance and considering the nearest neighbour distance for the separation.

[4]Further information about internal clustering evaluation metrics can be found in [140]

- Sum of squared error (SSE) is the most simple measure. It computes the error as the distance from the cluster centroid to each of the cluster objects.

Finally, as it was mentioned above, there is no uniform definition for clustering, and probably, it will not exist, given that there are many completely different applications for clustering. Focusing on the extraction-of-information point of view, clustering could be considered a preprocessing technique itself. Mainly, it can be used to discover useful information from the data or to discover relationships among all the variables [70].

1.4 Artificial neural networks

Artificial neural networks (ANNs) [26] are one of the most common modelling technique in the machine learning (ML) field, given their ability to learn and to correct errors, achieving outstanding results for real-world applications. They are based on simulating the biological neural networks from human beings brains, imitating the way humans analyse and process information. ANNs are composed by nodes (also known as artificial neurons), that emulate the neurons in a biological brain, also emulating the synapses by connections transmitting signals between these nodes. The first approach to neurons was the McCulloch–Pitts neuron [126], which consisted in the weighted sum of the inputs followed by the application of a non-linear function, also known as activation function.

The most basic model type of ANN is the feed-forward neural network (FNN). In this sort of ANN, the signals only move in the forward direction, from the inputs to the output nodes, going through the hidden nodes. A simple example of an FNN is shown in Figure 1.4.1, composed of an input layer with K nodes, a hidden layer with M neurons, and an output layer with $J-1$ nodes, where the number of neurons in the input layer matches with the number of independent variables from the model, and the number of neurons in the output layer is the number of classes minus one (classification model with J classes).

Depending on the ANN considered, the hidden layer will be composed of different sort of activation functions or basis functions. Regardless the hidden layer type, a one hidden layer FNN output can be written as:

$$f_j(\mathbf{x}, \mathbf{w}, \boldsymbol{\beta}) = \beta_{j0} + \sum_{m=1}^{M} \beta_{jm} B_m(\mathbf{x}, \mathbf{w}_m), \qquad j = 1, \ldots, J-1, \tag{1.1}$$

where $B_m(\mathbf{x}, \mathbf{w}_m)$ represents the set of non-linear transformations of the input vector $\mathbf{x}^{\mathrm{T}} = (x_1, x_2, \ldots, x_K)$, with K being its dimension; bias is considered in the output layer with the element $B_0(\mathbf{x}, \mathbf{w}_m) = 1$; $\boldsymbol{\beta}_j^{\mathrm{T}} = (\beta_{j1}, \beta_{j2}, \ldots, \beta_{jM})$ are the coefficients from the non-linear transformation estimated from the data; $\mathbf{w}_m^{\mathrm{T}} = (w_{m1}, w_{m2}, \ldots, w_{mK})$ are the

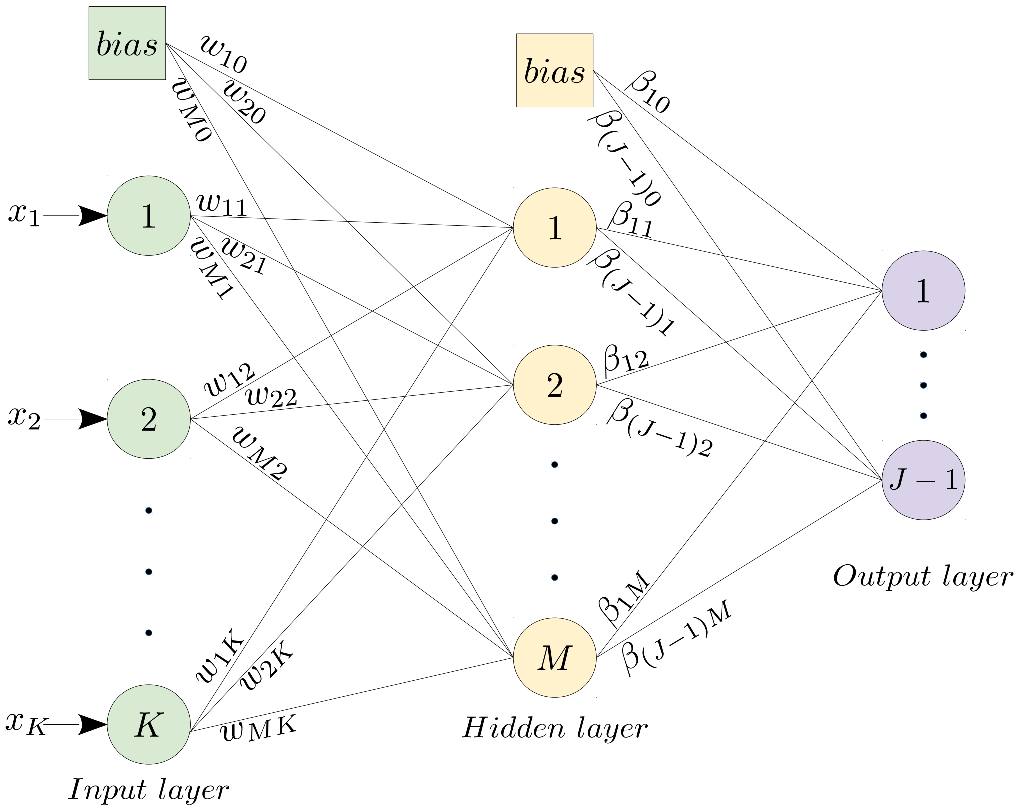

Figure 1.4.1: Example of a FNN.

parameters related to the basis functions; M is the number of basis functions required to minimise some defined error function and J is the number of outputs of the problem.

In the case of a classification problem, the outputs may be transformed to probabilities by using the softmax function [26]:

$$p_j(\mathbf{x}) = \frac{\exp f_j(\mathbf{x}, \mathbf{w}, \boldsymbol{\beta})}{\sum_{i=1}^{J} \exp f_i(\mathbf{x}, \mathbf{w}, \boldsymbol{\beta})}, \qquad j = 1, 2, \ldots, J, \tag{1.2}$$

where, given that all probabilities have to sum 1, it is assumed that $f_J(\mathbf{x}, \mathbf{w}, \boldsymbol{\beta}) = 0$.

The set of non-linear transformations of the input vector, also known as basis functions, $B_m(\mathbf{x}, \mathbf{w}_m)$, is formed by an activation function (the total input arriving to the node) and by an output or transfer function (the output of the neuron activation).

According to the sort of basis functions, two main groups can be considered:

- Local or kernel functions, which present a better performance when approximating isolated data. However, when dealing with global environment or a vast amount of inputs, the performance decreases considerably. The radial basis functions (RBFs) [23] are an example of local functions.

- Global or projection functions, which present difficulties when dealing with isolated data. By contrast, they are capable of achieving an outstanding performance on global environments, and on problems where the number of variables is high. Sigmoidal units (SUs) [121] and product units (PUs) [124] are examples of global functions.

On the other hand, according to the sort of activation function, two main types of networks can be differentiated:

- Additive model, which is the most common. Its output function is defined as:

$$B_m(\mathbf{x}, \mathbf{w}_m) = h(w_0^m + w_1^m x_1 + w_2^m x_2 + \ldots + w_K^m x_K), \tag{1.3}$$

  where K is the number of inputs, w_0^m is the bias and $h(\cdot)$ is the transfer function, both associated with neuron m. There are many different additive nodes: the perceptron [126], SUs and the identity function, among others.

- Multiplicative model, which is a recent strategy trying to lead with those situations in which there is an interaction between the variables or the decision regions are not separable by hyperplanes [162]:

$$B_m(\mathbf{x}, \mathbf{w}_m) = x_1^{w_1^m} \cdot x_2^{w_2^m} \cdot \ldots \cdot x_K^{w_K^m}, \tag{1.4}$$

  where a bias term makes no sense. Note that the general expression corresponds with the PU, being able to generalise other kinds of multiplicative units, given that the weights are real numbers.

Summarising, three main basis functions can be found in the literature:

- The SU is the most common basis function due to its ability to approximate any continuous function accurately. However, they fail in local optima frequently. Using the notation described above, SU is represented as:

$$B_m(\mathbf{x}, \mathbf{w}_m) = \frac{1}{1 + e^{-\left(w_{m0} + \sum_{i=1}^{K} w_{mi} x_i\right)}}, \qquad m = 1, ..., M. \tag{1.5}$$

- PU is a basis function that not only is it able to retain the properties of a universal approximator, but also it only uses a small number of multiplicative neurons [163]. According to the notation followed so far, the PU is expressed as:

$$B_m(\mathbf{x}, \mathbf{w}_m) = \prod_{i=1}^{K} x_i^{w_{mi}}, \qquad m = 1, \ldots, M. \tag{1.6}$$

- RBFs behave as different local elements, one for each hidden neuron, each pattern activating a different set of units. In this way, the number of local optima is decreased, making the training stage easier. Following the notation presented above, RBF is expressed as:

$$B_m(\mathbf{x}, \mathbf{w}_m) = e^{-\frac{1}{2}\left(\frac{\sum_{i=1}^{K}(x_i - c_{mi})^2}{r_m}\right)}, \qquad m = 1, \ldots, M, \tag{1.7}$$

  where the vector of weights of the m-th hidden neuron, $\mathbf{w}_m$, includes both a centroid $\mathbf{c}_m$ and a radium r_m for the corresponding Gaussian basis function, in such a way that $\mathbf{w}_m = \{r_m, \mathbf{c}_m\}$.

Therefore, there are three well-known FNNs with respect to the basis function used in the hidden layer:

- Sigmoidal unit neural networks (SUNNs) are also known as multilayer perceptrons (MLPs). SUs provide several advantages: they are able to approximate any given function with accuracy, being universal approximators [48, 92]. Moreover, even though reaching a local optimum is frequent, they are easy to be trained.
- Product unit neural networks (PUNNs) [56] are those ANNs following a multiplicative model of projection. PUNNs are also considered to be universal approximators [92]. It is worthy of mention that PUNNs benefit from the information capacity of individual PUs, which is much higher than additive neurons. Moreover, they have been proved to be excellent in problems with interactions of different order between the inputs. Nevertheless, PUNNs present a major disadvantage, the error surface is complex, leading frequently to local optima.
- Radial basis function neural networks (RBFNNs) are those ANNs considering kernel/local transfer functions. Each RBF node uses Gaussian functions to make local approximations to the input space. Note that the use of a linear output layer combines the effect of each hidden neuron, which are supposed to be adjusted to different regions of the input space (centre of the region) with a specific radius. This kind of ANN frequently uses the Euclidean distance as the activation function (the centre of the RBF is $\mathbf{w}_m$) and the Gaussian function as the transfer function. Finally, RBFNNs are also considered to be universal approximators [138].

1.4.1 Hybrid artificial neural networks

So far, ANNs presented only make use of one type of basis function in the hidden layer. However, the mixture of different basis functions could have several advantages, such as providing flexible decision borders or complementing the characteristics of the different families of basis functions. It has been demonstrated that any continuous function could be decomposed into two types of functions mutually exclusive, a projection function (SUs, PUs), and a kernel function (RBFs) [54]. A basic example of a hybrid FNN is shown in Figure 1.4.2, being composed of an input layer with K nodes, a hidden layer with M_1 hidden neurons of the first type and M_2 hidden neurons of the second type, and an output layer with $J-1$ nodes.

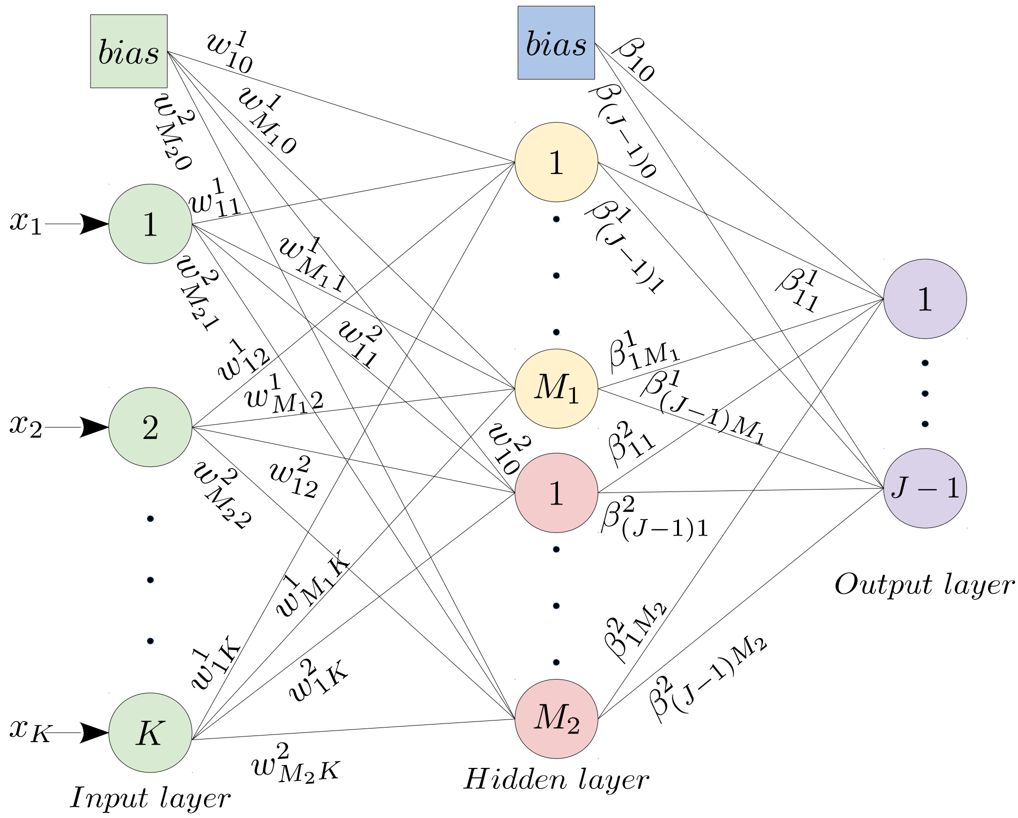

Figure 1.4.2: Structure of a hybrid ANN.

For these models, the ANN output can be written in the following way:

$$f_j(\mathbf{x}, \boldsymbol{\theta}) = \beta_{j0} + \sum_{m=1}^{M_1} \beta_{jm}^1 B_m^1(\mathbf{x}, \mathbf{w}_m^1) + \sum_{m=1}^{M_2} \beta_{jm}^2 B_m^2(\mathbf{x}, \mathbf{w}_m^2), \qquad j = 1, 2, \ldots, J-1, \quad (1.8)$$

where M_1 and M_2 are the number of hidden neurons of the first and the second type of hidden neuron, respectively, $\boldsymbol{\theta} = \{\boldsymbol{\beta}, \mathbf{w}_1^1, \ldots, \mathbf{w}_{M_1}^1, \mathbf{w}_1^2, \ldots, \mathbf{w}_{M_2}^2, \}$ is the vector containing all the coefficients of the neural network, $\boldsymbol{\beta} = \{\beta_{j0}, \beta_{j1}^1, \ldots, \beta_{jM_1}^1, \beta_{j1}^2, \ldots, \beta_{jM_2}^2\}$ includes the coefficients between hidden and output layers, and $\mathbf{w}_m^1$ and $\mathbf{w}_m^2$ represent the weights connecting the input layer to the m-th hidden neuron of the first and the second types, respectively. Any of the basis functions previously defined (Equations 1.5, 1.6 or 1.7) can be used for $B^1_m(\mathbf{x}, \mathbf{w}^1_m)$ and $B^2_m(\mathbf{x}, \mathbf{w}^2_m)$.

1.4.2 Artificial neural networks training

Focusing on the simplest ANN, the training step consists in estimating the values of $\mathbf{w}$ and $\boldsymbol{\beta}$, i.e. the coefficients of the ANN, and the structure of the network, i.e. the number of neurons in the hidden layers. M (note that in the case of hybrid ANNs, M is defined as $M = M_1 + M_2$) and the number of connections between layers. There are two main groups of methods:

- Classic methods, which typically lead to suboptimal solutions for the problem addressed. In this group, it can be found several different approaches:
  - Constructive methods, which start with a simple structure and add nodes and links trying to increase the performance previously achieved, according to an expert opinion, generally using a trial and error method [30]. Nevertheless, these methods frequently fall in local optima.
  - Destructive methods, which are the opposite methods, they start with a complex structure and delete nodes and connections aiming to improve the performance [148]. They have the same disadvantage of falling in local optima.
  - Gradient-based methods, are the most common technique for training ANNs. The back-propagation (BP) algorithm [154] is the most common one. In order to avoid falling in local optima, a group of parameters, such as the learning rate, the number of hidden layers or the weight initialisation, among others, are set in advance. However, they still suffer from some disadvantages, such as the impossibility of computing the gradient with non-derivable activation functions or the impossibility of reaching the global minimum if the error function is multimodal and/or non-derivable.
- Heuristic methods, which were developed in order to counteract the disadvantages presented by the classic methods. In this way, the most common techniques are:
  - Variants for the BP algorithm: using adaptative speed, optimising the learning parameters [200], or using the *iRprop+* algorithm [96], among others.
  - Standard approaches, such as simulated annealing [177] or tabu search [75].
  - More advanced techniques, such as evolutionary algorithms (EAs) [14], which have been proved to achieve excellent results.

In the last few years, EAs have been widely used for training ANNs, given that not only are they able to optimise the values of the links, but also they also consider modifications of the architecture of the network. Therefore, in this Thesis, evolutionary artificial neural networks (EANNs) [88] have been used.

1.4.3 Evolutionary artificial neural networks

EAs have been widely used in the literature for both training classification [151] and regression ANNs [43, 79]. EAs are highly efficient for searching a set of weights close to the optimum ones with two main advantages: 1) not using gradient information and 2) not forcing the fitness function to be derivable. EAs are able to automatically optimise the connection weights and the architecture of the network. The general structure of the EANN algorithm considered in this work is shown in Algorithm 1 [124].

Algorithm 1: EANN

```
generate a random population P
while stopping criteria is not satisfied do
    evaluate and rank the individuals
    keep the best individual
    for worst 10% of P do
        replace with the best 10% of P
    end
    parametric mutation to the best 10% of P
    structural mutation to the remaining 90% of P
    evaluate and rank the individuals
    the worst individual is replaced by the best individual of P stored previously
end
return the best individual
```

In Algorithm 1, the population P is composed of the individuals, which in this case, represent ANN models. Regarding the evaluation of the individuals, it could be done by means of any error function, such as the mean squared error (MSE) when dealing with regression problems, or using a performance measure, such as the correct classification rate (CCR) or the minimum sensitivity (MS) [64] or an error, such as the cross-entropy (E), when dealing with nominal classification problems. Moreover, the ranking or the selection of the best individual is done by maximising the fitness (minimising the error or maximising the performance metric). Furthermore, as can be seen, there are two mutation operators:

- The parametric mutator, which optimises the values of the network weights. This is done by adding a Gaussian noise to the connections of the individual, this is, a value following a normal distribution with mean zero and $\alpha \cdot T$ variance, $\mathrm{N}(0, \alpha \cdot T)$, where α is a dynamic parameter different for every kind of connection (α_1 for connections between the input and the hidden layer, and α_2 for connections between the hidden and the output layer). Moreover, T is the temperature value, defined as $T = 1 - A(g)$, where $A(g)$ is the fitness function for the network g, similar to a simulated annealing.

- The structural mutator, which modifies the architecture of the ANN model. There are five different types:

  - Node addition, which consists in adding neurons to the ANN model.
  - Node deletion, which consists in deleting exiting neurons of the ANN model.
  - Connection addition, which is based on adding new connections to the ANN model.
  - Connection deletion, which is based on deleting existing connections of the ANN model.
  - Node fusion, which consists in the fusion or union of two neurons of the ANN model.

  The ones related with connections can be applied to the hidden and output layer, whereas the remaining structural mutations can only be applied to the hidden layer.

  One of these structural mutations are applied to a $\beta\%$ of the population. For this, each individual has a temperature value, T, and, depending on whether this value is higher than a random value on the range $[0, 1]$, the first structural mutation is applied according to the order in which they have been previously described. Note that, if there is a special circumstance in which the structural mutation can not be applied (neurons without links or not having hidden layer, among others), the next type of structural mutation (following the pre-established order) will be applied.

Note that, at the beginning of the EA, the fitness value of the individuals is close to 0, and therefore, the temperature value, T, is high. In this way, the ratio of mutations is higher. As we advance in the evolutionary process, T decreases, and, therefore, the changes are smaller, thus refining the search.

On the other hand, sometimes there are problems where the ANNs have to be optimised according to more than one objective, which are known as multi-objective problems (MOPs). For these problems, the EAs used for optimisation are known as multi-objective evolutionary algorithms (MOEAs) [40]. Generally, MOEAs work in the same way as EAs, but now, they take into account at least two objective functions to guide the searching process. When MOEAs are used in the training step of ANNs, it is known as multi-objective evolutionary artificial neural networks (MOEANNs) [101].

1.4.4 Multi-objective evolutionary artificial neural networks

MOEANNs have been proved to be competitive for both classification [80] and regression problems [160]. In order to get an insight into MOEANNs, it is required to formally introduce the concept of MOPs. In this way, a MOP could be defined as a problem aiming to find a vector $\mathbf{z}^* = [z_1^*, z_2^*, \ldots, z_l^*]^{\mathrm{T}}$ that optimises a function vector $\mathbf{f}(\mathbf{z}) = [f_1(\mathbf{z}), f_2(\mathbf{z}), \ldots, f_t(\mathbf{z})]^{\mathrm{T}}$, satisfying b inequality constraints, $g_i(\mathbf{z}) \geq 0$ for $i = 1, 2, \ldots, b$, and d equality constraints, $h_i(\mathbf{z}) = 0$ for $i = 1, 2, \ldots, d$.

A MOP has t objectives and the functions $\mathbf{f}(\cdot) : \Omega \rightarrow A$ represent the relation between the search space Ω and the objective function space A. In this sense, the solution of a MOP is generally multiple, i.e. there are several optimal solutions, given that for most cases the objective functions are conflicting (optimising one of them leads the other one to worse results).

At this point, the concept of Pareto optimality need to be explained. A solution $\mathbf{z}^* \in \Omega$ can be considered Pareto optimal if there is no $\mathbf{z} \in \Omega$ whose function $\mathbf{f}(\mathbf{z})$ dominates $\mathbf{f}(\mathbf{z}^*)$. It is said that a vector $\mathbf{z}$ dominates another vector $\mathbf{z}'$ (denoted by $\mathbf{z} \succeq \mathbf{z}'$), if and only if $\forall_i \in \{1, 2, \ldots, t\}$, $z_i \leq z_i'$ and $\exists i \in \{1, 2, \ldots, t\} : z_i < z_i'$. In this way, the set of Pareto optimal ($\mathcal{P}^*$) is defined as:

$$\mathcal{P}^* := \{\mathbf{z}^* \in \Omega \mid \nexists \mathbf{z} \in \Omega, \;\; \mathbf{f}(\mathbf{z}) \succeq \mathbf{f}(\mathbf{z}^*)\}, \tag{1.9}$$

hence, the Pareto front ($\mathcal{PF}^*$) is expressed as:

$$\mathcal{PF}^* := \{\mathbf{u} = \mathbf{f}(\mathbf{z}) | \mathbf{z} \in \mathcal{P}^*\}. \tag{1.10}$$

Therefore, the Pareto front matches the global minimum of the MOP, i.e. given a vector of functions $\mathbf{f}(\cdot) : \Omega \subseteq \mathbb{R}^t \rightarrow \mathbb{R}^l$, $\Omega \neq 0$, and $t \geq 2$, the set $\mathcal{PF}^* : \mathbf{f}(\mathbf{z}^*)$ is called global minimum, if and only if $\forall \mathbf{z} \in \Omega : \mathbf{f}(\mathbf{z}^*) \succeq \mathbf{f}(\mathbf{z})$.

Finally, there are two types of solutions $\mathbf{z}^* \in \Omega$: 1) weakly non-dominated solution if there is no $\mathbf{z} \in \Omega$ such that $f_i(\mathbf{z}) < f_i(\mathbf{z}^*)$, for $i = 1, 2, \ldots, t$, and 2) strictly Pareto optimal if there is no $\mathbf{z} \in \Omega$, $\mathbf{z} \neq \mathbf{z}^*$, such that $f_i(\mathbf{z}) \leq f_i(\mathbf{z}^*)$, for $i = 1, 2, \ldots, t$.

From the previous definitions, we can extract that MOPs do not have a single solution able to optimise simultaneously all the objective functions, given that for most cases, they are conflicting. In this way, there are numerous Pareto optimal $\mathcal{P}^*$ solutions (possibly infinite). Note that all the Pareto optimal solutions $\mathcal{P}^*$ (also known as non-dominated solutions) are considered to be equally good, being all of them part of the Pareto front $\mathcal{PF}^*$. Non-dominated solutions are those that can not improve the value for one objective function, without decreasing any of the values for the remaining objective functions. An example of a Pareto front $\mathcal{PF}^*$ with two objective functions is shown in Figure 1.4.3.

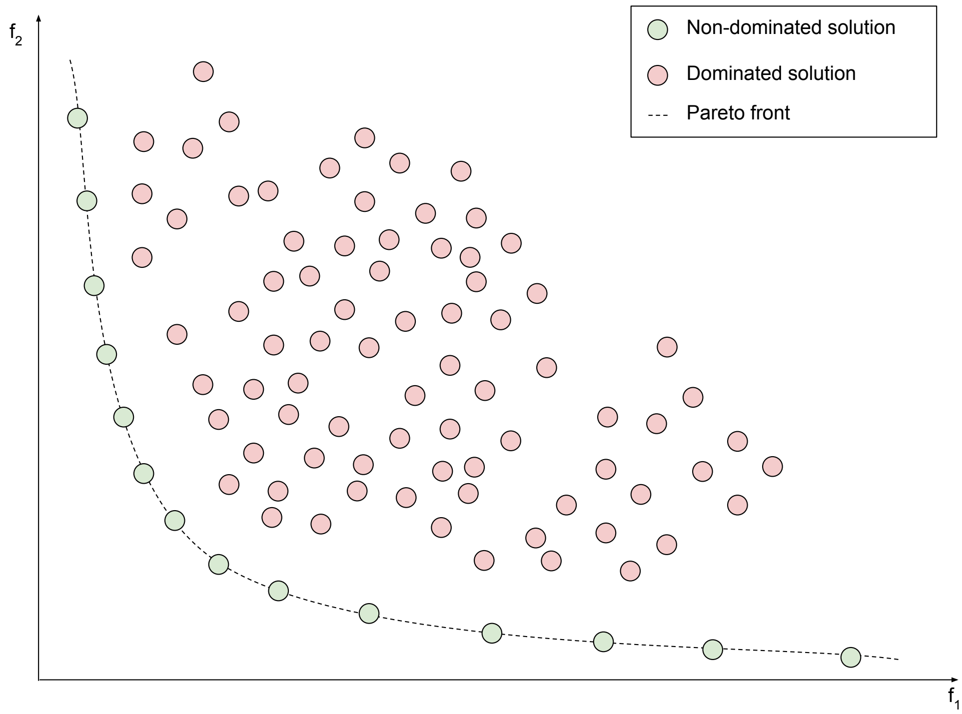

Figure 1.4.3: Pareto front obtained for a problem with two objective functions.

Problems with two or more objectives can also be optimised by EAs. They are known as MOEAs [40, 50]. One of the most popular MOEAs is the non-dominated sorting genetic algorithm II (NSGA-II) [51], which is an elitist method using different operators to preserve the diversity and the best solutions from one generation to the next.

NSGA-II is an improved version of the non-dominated sorting genetic algorithm (NSGA), using an operator known as crowding, which is used to select those disperse solutions among all the individuals (or solutions) from the last Pareto front $\mathcal{PF}^*$. The higher the crowding distance from one solution to the rest of the solutions belonging to the Pareto front, the best.

Moreover, for each generation, NSGA-II creates a set of individuals by joining the current population and the one created by selection, crossover and mutation. From this set, the different Pareto fronts are extracted (grouped and ordered by the number of dominated solutions), i.e. the first Pareto front is formed by all the non-dominated solutions.

As with EAs, MOEAs can be used to optimise ANNs, which are known as MOEANNs. A general structure of the MOEANN algorithm is shown in Algorithm 2.

Algorithm 2 starts with a random generation of individuals (in this case, ANNs),

Algorithm 2: MOEANN

generate a random population P
evaluate and rank the individuals according to Pareto front strategy
while *stopping criteria is not satisfied* **do**
  tournament selection to choose P individuals for the mutations
  random parametric and structural mutations (offspring of size P: population Q)
  parents and mutated offspring $(P + Q)$
  evaluate and rank the individuals
  keep the P best individuals
end
return *the first Pareto front*

which are evaluated according to at least two objective functions, and then, sorted in Pareto fronts following the concept of Pareto dominance. After that, until satisfying the stopping criteria, new individuals are generated by tournament techniques and parametric and structural mutations are probabilistically applied (these mutations are those described in Section 1.4.3). Best individuals are kept in each generation. Finally, once the evolutionary process is finished, the set of solutions providing the best performance need to be selected. For this purpose, the strategy of choosing the two extremes of the front is followed (for those MOPs with two objective functions), i.e. the best model in terms of each objective functions are chosen, given that these models are those maximising (or minimising) each of the objective functions.

1.5 Time series

Time series are a special kind of data, in which data points are collected chronologically. Time series are also considered as a function varying across time. Examples of time series can be found in several fields, such as the height of ocean tides, the daily closing value of the different stock market indices, or the number of daily sales for a given product. There are numerous tasks that could be applied to time series [67], depending on the purpose and the area of application. In the context of statistics, econometrics or meteorology, the main goal of time series analysis is forecasting [36]. Regarding the signal processing area, it is mainly used for detection and estimation [178]. Finally, in the primary area concerning this Thesis (data mining, pattern recognition and machine learning (ML)), time series analysis can be used for querying by contents (time series motifs) [115], anomaly detection [34], classification [17, 62], clustering [1], segmentation [106] or prediction [186], among others.

As previously stated, generally, this Thesis is focused on time series analysis. More precisely, it concerns time series clustering (where time series segmentation is used), time series prediction, and finally, time series classification (both nominal and ordinal).

1.5.1 Time series prediction

Time series prediction (also known as time series forecasting) is one of the most popular time series data mining techniques. It consists in estimating the next value of a time series, which is quite often the ultimate goal of time series data mining. More formally, if a time series is defined as $X = \{x_n\}_{n=1}^{N}$, the prediction task will estimate the value x_{N+l}, where l is the number of points to be estimated.

In time series prediction, historical data (i.e. information about past) can be used to predict future values. Traditionally, time series prediction has been approached considering the concept of stationarity. A time series is considered stationary when its data varies around the same mean value, with constant variance, and the relationship between values depends only on the number of points between them. This is, the distribution of the values for a given time series $X = \{x_1, x_2, \dots, x_k\}$ is the same distribution for the values $\{x_{1+h}, x_{2+h}, \dots, x_{k+h}\}$, where the relationship between these values depends only on the h values separating them.

We can find the following time series models in the state-of-the-art that are considered traditional prediction models: autoregressive model (AR), moving average model (MA), autoregressive moving average model (ARMA), and autoregressive integrated moving average model (ARIMA).

Autoregressive models

ARs are one of the simplest form of prediction, also known by the notation AR(p). An AR model of order p consists in performing the forecasting via a regression equation, where the independent variables are the p previous lagged values of the dependent variable:

$$x_n = \delta + \beta_1 x_{n-1} + \beta_2 x_{n-2} + \cdots + \beta_p x_{n-p} + \varepsilon_n, \tag{1.11}$$

where δ is a constant, $\upbeta = (\beta_1, \beta_2, \dots, \beta_p)$ are the model parameters, and ε_n is white noise (a white noise time series $\upvarepsilon = \{\varepsilon_n\}_{n=1}^{N}$ has zero mean and constant variance).

Moving average models

The MA is the other simplest form of forecasting, based on the idea of computing the average of past values to forecast future values. The notation MA(q) refers to a moving average model of order q, i.e. the past q values are used to predict future values in the following way:

$$x_n = \mu + \varepsilon_n + \alpha_1 \varepsilon_{n-1} + \alpha_2 \varepsilon_{n-2} + \cdots + \alpha_q \varepsilon_{n-q}, \tag{1.12}$$

where μ is the mean of the time series, $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_q)$ are the parameters of the model, and $\varepsilon_n, \varepsilon_{n-1}, \ldots, \varepsilon_{n-q}$ are white noise error terms (with mean zero and σ^2 variance).

Autoregressive moving average models

The polynomial combination of an AR(p) term and an MA(q) term is known as ARMA model with order (p, q). ARMA(p, q) is given by:

$$x_n = \delta + \beta_1 x_{n-1} + \beta_2 x_{n-2} + \cdots + \beta_p x_{n-p} + \varepsilon_n + \alpha_1 \varepsilon_{n-1} + \alpha_2 \varepsilon_{n-2} + \cdots + \alpha_q \varepsilon_{n-q}. \tag{1.13}$$

Autoregressive integrated moving average models

A generalisation of the ARMA model is presented in ARIMA models, which are typically applied to non-stationary time series. A non-stationary time series can be transformed into a stationary time series by including an initial differencing step (which corresponds to the integrated feature of the model). This differencing step can be applied one or more times in order to remove the non-stationarity. ARIMA models are denoted by ARIMA(p, d, q), where p is the order of the AR model, d is the number of differencing steps, and q is the order of the MA model. An ARIMA(p,d,q) model can be written as:

$$x_n^d = \delta + \beta_1 x_{n-1}^d + \beta_2 x_{n-2}^d + \cdots + \beta_p x_{n-p}^d + \varepsilon_n + \alpha_1 \varepsilon_{n-1} + \alpha_2 \varepsilon_{n-2} + \cdots + \alpha_q \varepsilon_{n-q}, \tag{1.14}$$

where x_n^d is the differenced time series (d differencing steps).

Order selection for traditional prediction models

Choosing appropriate values for these previous traditional prediction models is not a trivial task. However, there are two plots (or functions) that help in the decision of appropriate values for p and q. These functions are the autocorrelation function (ACF) and the partial autocorrelation function (PACF). The former returns the values of autocorrelation of any time series with its lagged values, whereas the latter returns the correlation of the residuals (being what remains after removing the effects explained by earlier lags). Note that the first PACF coefficient is identical to the first ACF coefficient, given that no lagged values are removed. In this way, ACF and PACF are functions that help to determine the appropriate values for p and q.

In order to choose the orders p and q, we need to assure that the time series is stationary. If not, the time series need to be transformed to stationary by differencing it. Once the time series is stabilised, according to the ACF and PACF plots, we have three possible ways to determine these orders:

1. If the ACF shows a gradual decrease of its terms, and simultaneously the PACF shows a sharp drop after p significant lags, we can consider that the time series can be modelled by an AR(p).
2. On the other hand, if the PACF shows the gradual decrease, and at the same time, the ACF shows the sharp drop after q number of lags, we can consider that the time series can be modelled by a MA(q).
3. Otherwise, if both ACF and PACF plots show a gradual decrease, then an ARMA(p,q) process should be considered for modelling the time series.

There are other widely used measures for identifying the model, such as the Akaike information criterion (AIC) [6] or the bayesian information criterion (BIC) [164].

1.5.2 Time series classification

Time series classification (TSC) is the most popular technique in time series data mining. It consists in developing a mapping function from the space of inputs to a probability distribution over the class labels. More formally, a TSC dataset is defined as $D = \{\mathbf{t}_i, y_i\}_{i=1}^{M}$, with M patterns. $\mathbf{t}_i = \{x_{i1}, x_{i2}, \dots x_{iN_i}\}$ represents a time series pattern (note that the existence of an individual N_i for each pattern implies that unequal-length time series could be considered), and y_i is the i-th discrete class value, $y_i \in \mathcal{Y} = \{C_1, C_2, ..., C_J\}$, with J possible values.

According to the nature of the labels assigned to the time series, the ordinal classification (OC) methodology could result in better results by considering their ordinal nature. In this sense, TSC field could be divided into two separate fields: nominal (known in the literature as TSC) and ordinal (in this Thesis, we propose the use of TSOC as the abbreviation for time series ordinal classification (TSOC) field). TSOC, up to the knowledge of the author, is an unexplored field, which has just started to receive some attention. In this Section, TSC is revised in big detail given its trajectory in the literature.

Time series nominal classification

In the last decade, the TSC field has been through an enormous rise of interest. Numerous approaches have been proposed to solve this task, many of them involving some processing or filtering of the time series before constructing the classifier. Bagnall *et al.* in [17] proposed a taxonomy to improve the understanding of the advantages and disadvantages of existing approaches. Figure 1.5.1 shows a taxonomy including novel algorithms proposed in the literature.

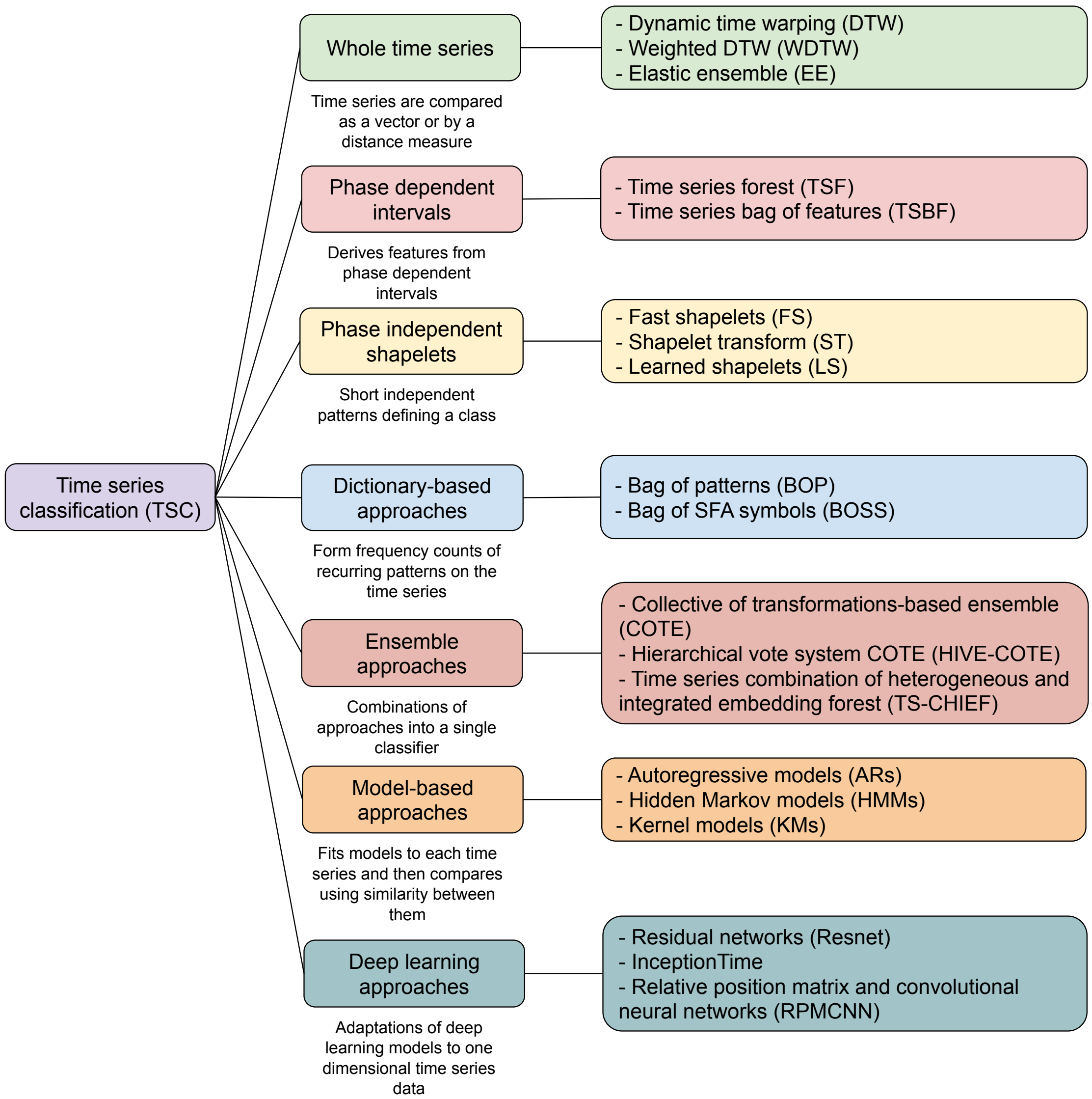

Figure 1.5.1: TSC techniques taxonomy.

The main TSC techniques can be described as follows:

- Whole time series: this group of techniques employs similarity/distance measures with a nearest neighbour classifier. This kind of techniques is appropriate only when there are discriminatory features over the whole time series. Dynamic time warping (DTW) [98] and its weighted version, weighted dynamic time warping (WDTW) [100] are the most popular distance measures in the state-of-the-art. Even though many novel alternatives have been presented in the literature, Wang *et al.* in [183] demonstrated that none of them were significantly better than DTW. Moreover, Lines and Bagnall in [119] proposed the elastic ensemble (EE), which is an ensemble of 11 nearest neighbours classifiers with different distance measures, demonstrating that this approach outperformed significantly every single component.

- Phase dependent intervals: techniques belonging to this group are excellent when applied to long time series datasets, including phase dependent discriminatory patterns and regions of noise. Main algorithms are time series forest (TSF) [52] and an extension known as time series bag of features (TSBF) [21]. Both techniques employ a random forest approach along with summary statistics of each interval as features.

- Phase independent shapelets: in this case, the location of the patterns extracted from the time series is irrelevant, i.e. the patterns define the classes but the location of the pattern could be in any point of the measurement. These patterns or subseries are known as shapelets. Main approaches involving the use of shapelets are fast shapelets (FS) [143], shapelet transform (ST) [89], or learned shapelets (LS) [76].

- Dictionary-based approaches: methods belonging to this group classify time series by the frequency of repetition of subseries, i.e. train the classifiers from the histograms which result from the frequency counts of recurring patterns. The most popular algorithms in this group are bag of patterns (BOP) [117], which works by representing time series as symbols (using the SAX [116] method), and then build the classifier, and the bag of symbolic Fourier approximation symbols (BOSS) [161], which is a improved version of BOP using the discrete Fourier transform for the windows creations, among other differences.

- Ensemble approaches: this group covers combinations from several individual approaches into a single classifier. The main approaches of this kind are the collective of transformation-based ensemble (COTE) [18], its extension, hierarchical vote system collective of transformation-based ensemble (HIVE-COTE) [120], and the time series combination of heterogeneous and integrated embedding forest (TS-CHIEF) [168]. On the one hand, HIVE-COTE is the state-of-the-art in the TSC field, consisting in an ensemble encapsulating classifiers built on the different previously explained data representations. On the other hand, TS-CHIEF is a novel ensemble approach, in which embeddings of time series are integrated using tree-structured classifiers.

- Model-based approaches: this set of techniques fit generative models to each time series and then compare them by using the distance between models. The most well-known approaches in this field are based on fitting AR models [16], applying hidden Markov models (HMM) [169], or kernel model (KM) [37], among others.

- Deep learning (DL) approaches: in the last years, DL have been proved to be excellent for TSC [62]. Adaptations of DL methods to one-dimensional time series data are also the state-of-the-art in this area. The main techniques include the application of residual networks (Resnet) [185], which conform the most deepest architecture applied to TSC task, or inceptionTime [63], which is proved to be on pair of HIVE-COTE accuracy but with higher scalability. Another approach is relative position matrix and convolutional neural network (RPMCNN) [38], which consists in transforming time series to structural images and then applying a simplified version of the popular VGGNet architecture.

Time series ordinal classification

For those time series datasets where there is a natural order between the labels and the number of classes is higher than 2, nominal classifiers are not the best option. In this context, this Thesis proposes to consider TSOC, which bridges the gap by applying the ordinal classification paradigm in order to maximise the performance on those ordinal time series datasets. TSOC requires time series datasets in which the class attribute Y is defined as $Y \in \mathcal{Y} = \{C_1, C_2, \ldots, C_J\}$, where J is the number of categories and $\{C_1, C_2, \ldots, C_J\}$ are the ordinal labels, satisfying the constraint $C_1 \prec C_2 \prec \ldots \prec C_J$. As aforementioned, up-to-the-knowledge of the authors, there are no specific ordinal approaches in the literature, this Thesis proposing some techniques for this novel field.

Time series shapelets

In this Thesis, we focus on shapelets, which have been used for both TSC and TSOC tasks. Shapelets are phase independent subsequences of the time series forming a new primitive for TSC. They were firstly proposed by Ye and Keogh [198, 199], yet improved versions and new perspectives have been presented in the literature [27, 76, 89]. An example of a shapelet is shown in Figure 1.5.2, in which each time series represents the distances from the border points to the centre of the each image. In this example, the shapelet is extracted from the first time series (in red colour), and it is a unique characteristic from that time series (in this case, something representative from the tractors is the metal fender and hood). Then, the matches are shown for the remaining testing time series.

More formally, a shapelet $\mathbf{s} = \{s_1, s_2, \ldots, s_v\}$ is a subsequence of a time series $\mathbf{t}_i$, where $v \leq N_i$. One of the main approaches with shapelets is using them to build a transformed dataset, in which the transformed attributes represent the shape-similarity between the original time series and the shapelets. This approach is known as ST, and it begins by the shapelet generation procedure, with the main steps shown in Algorithm 3.

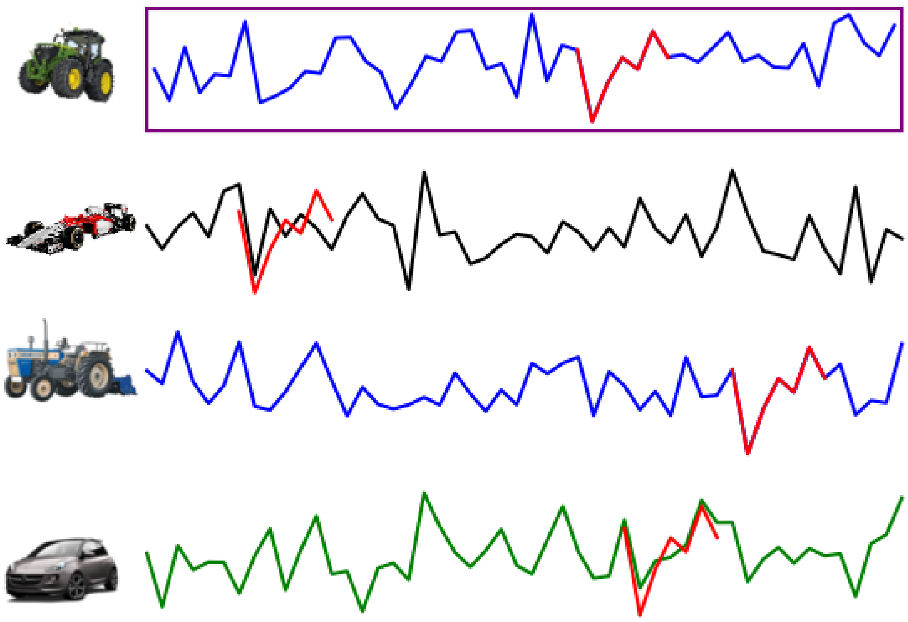

Figure 1.5.2: Example of a shapelet matching procedure.

Specifically, the shapelet generation procedure is divided into three main parts: 1) generation of candidates (subsequences) satisfying the previous length constraint ($v \leq N_i$), 2) computation of distances between the candidate and the time series to measure their similarity, and 3) evaluation of the candidate quality. The most recent version of ST [27] uses the the euclidean distance to measure the distance between the shapelets and the time series (note that this distance is computed as the minimum of the distances between the shapelet, and all possible subsequences of the time series with the same length of the shapelet), and the information gain (IG) to evaluate the quality of each shapelet. In this sense, the IG measures how well the shapelet class is discriminated from the rest, according to the set of distances between the shapelet $\mathbf{s}$ and the time series $\mathbf{t}$. To compute the IG the set of distances, $d_{\mathbf{s}}$, from the evaluated shapelet $\mathbf{s}$ to all the time series $\mathbf{t}$ need to be calculated and sorted. Then, all the possible split points (average points between two consecutive distances) are evaluated, keeping the best IG split point. The IG of a shapelet $\mathbf{s}$ is defined as:

$$IG(\mathbf{s}) = max_{s_{\mathrm{p}} \in d_{\mathbf{s}}} IG(s_{\mathrm{p}}), \tag{1.15}$$

where $IG(s_{\mathrm{p}})$ is the IG for an specific split point (s_{p}), and it is expressed as:

$$IG(s_{\mathrm{p}}) = H(d_{\mathbf{s}}) - \left(\frac{|s_{\mathrm{p}}^{-}|}{|d_{\mathbf{s}}|} H(|s_{\mathrm{p}}^{-}|) + \frac{|s_{\mathrm{p}}^{+}|}{|d_{\mathbf{s}}|} H(|s_{\mathrm{p}}^{+}|) \right), \tag{1.16}$$

Algorithm 3: Main steps of the best shapelet set $\mathcal{S}$ generation.

$\mathcal{S} \leftarrow \emptyset$ // Shapelet set
for *Each time series* $\mathbf{t}_i$ **do**
  $\mathcal{S}_{\mathbf{t}_i} \leftarrow \emptyset$
  $bestQuality \leftarrow 0$
  for $v \leftarrow min\ to\ max$ **do**
    $P_v \leftarrow$ Generate candidates($\mathbf{t}_i$, v)
    for *Candidate* $\mathbf{s}$ *in* P_v **do**
      $d_\mathbf{s} \leftarrow$ Calculate distances($\mathbf{s}$, $\mathbf{t}$)
      $quality \leftarrow$ Evaluate candidate ($\mathbf{s}$, $\mathbf{t}$)
      if $quality > bestQuality$ **then**
        $\mathcal{S}_{\mathbf{t}_i} \leftarrow S$
        $bestQuality \leftarrow quality$
      end
    end
  end
  $\mathcal{S} \leftarrow \mathcal{S}_{\mathbf{t}_i}$
  Sort $\mathcal{S}$ by quality
  Remove similar shapelets in $\mathcal{S}$
end
return *Best shapelet set* $\mathcal{S}$

where s_p^- are the elements of the sorted distance set located at the left of the split point, s_p, whereas s_p^+ are the remaining elements. Moreover, $|d_\mathbf{s}|$ and $H(d_\mathbf{s})$ are the cardinality and the entropy of the set $d_\mathbf{s}$, respectively, being the entropy defined as:

$$H(d_\mathbf{s}) = -\sum_{c \in \mathcal{Y}} p_c\ log\ p_c, \tag{1.17}$$

where p_c is the a priori probability of class c.

Once the best shapelets set $\mathcal{S}$ is obtained following the main steps explained in Algorithm 3, a new transformed dataset is built, in which each attribute represents a shapelet, S, and the value of the attribute is the distance between the shapelet s and all the original time series $\mathbf{t}$, $d_\mathbf{s}$. In this sense, any classifier could be applied to the transformed dataset, disassociating the shapelet extraction procedure from the classification stage.

1.5.3 Time series clustering

Time series clustering consists in grouping time series aiming to discover interesting patterns in the time series datasets. This field has received a lot of attention during this last decade, hence, there are several recent review papers regarding time series clustering [1, 67, 113]. Time series clustering is commonly applied as a preprocessing step for several tasks: 1) anomaly detection [112], 2) recognition of dynamic changes [87], 3) prediction [77], or 4) classification [2], among others.

More formally, given a time series clustering dataset defined as $D = \{T_i\}_{i=1}^{M}$, where $T_i = \{t_j\}_{j=1}^{N_i}$ is a time series of length N, the objective of time series clustering is to organize them into L groups, $\mathcal{G} = \{\mathcal{G}_1, \mathcal{G}_2, \dots \mathcal{G}_L\}$, optimizing the clustering quality, where $\forall \mathcal{G}_i \neq \mathcal{G}_j, \mathcal{G}_i \cap \mathcal{G}_j = \emptyset$ and $\bigcup_{l=1}^{L} \mathcal{G}_l = \mathcal{G}$.

Time series clustering algorithms can be classified into 3 different sets according to [1]. Figure 1.5.3 shows a taxonomy including the main techniques proposed in the literature.

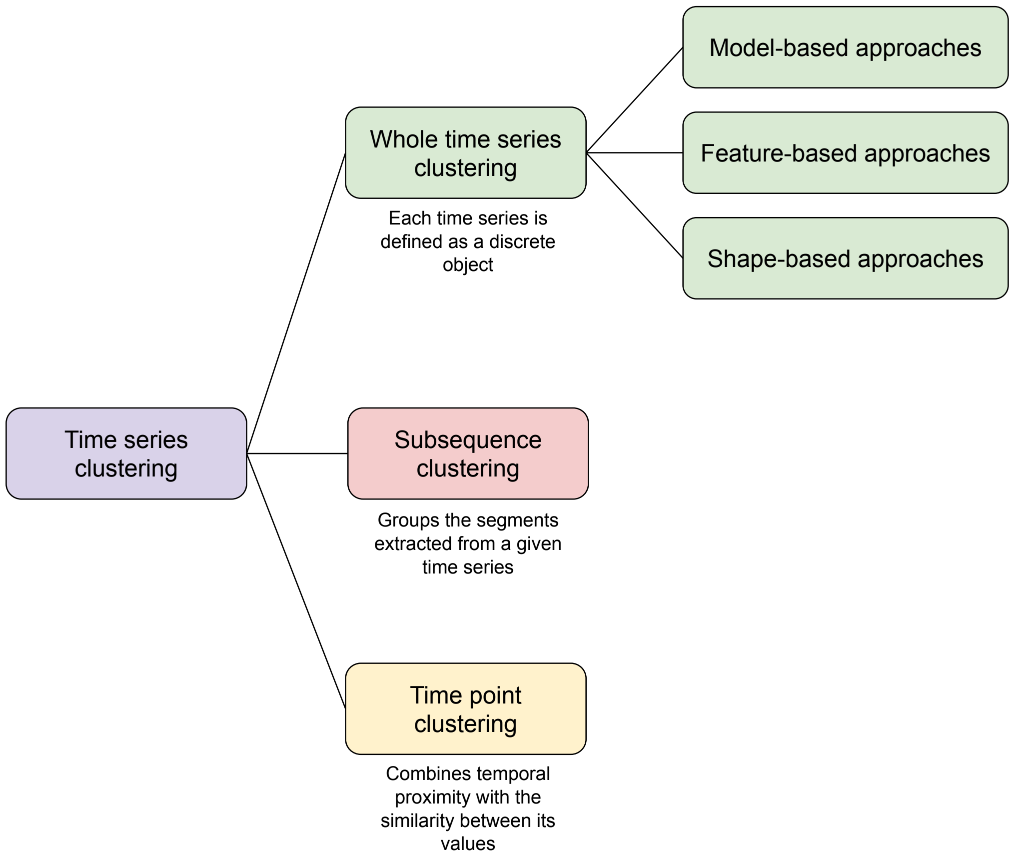

Figure 1.5.3: Time series clustering techniques taxonomy.

The taxonomy can be described as follows:

- Whole time series clustering defines each time series as a discrete object and clusters a set of time series measuring their similarity and applying a conventional clustering algorithm.
    - Model-based approaches: techniques pertaining to this group convert the original raw time series into a set of model parameters. The distance between mod-

els is measured, and a classic clustering algorithm is then applied. There are two main algorithms in this group: Yang *et al.* in [195] proposed the combination of rival penalised competitive learning with other representations into an unsupervised ensemble, and Yang and Jiang in [196] presented a novel HMM with bi-weighting scheme to solve problems related with the initialisation and model selection of time series clustering algorithms.

  ◦ Feature-based approaches: in this group, methods create a new representation of the time series by transforming them into a set of statistical characteristics. Note that each time series is converted to a new representation whose size is smaller than the original time series. One of the advantages of this group is its application to unequal-length time series datasets, given that the feature vector have the same length. Then, a standard distance measure is calculated, and a clustering algorithm is applied. Two algorithms belonging to this category are the one proposed by Räsänen *et al.* in [146], which is based on an efficient computational method for statistical feature-based clustering, and the one proposed by Hautamaki *et al.* in [86], who presented a raw time series clustering using the DTW distance for hierarchical and partitional clustering algorithms. The problem of DTW is that it can be highly sensitive to noise.

  ◦ Shape-based approaches: methods belonging to this group aim to match the shapes of the different time series. A conventional clustering algorithm is applied using an appropriate distance measure. Paparrizos *et al.* in [137] presented an approach using a normalized version of the cross-correlation measure that bears in mind the shapes of the time series, the proposal made by Asadi *et al.* in [11] consists in a novel method based on HMMs ensembles, and, recently, Wang *et al.* in [184] developed a new technique focused on studying the trend of time series. In their proposal, the similarity between two time series is measured by using an area-based shape distance.

- Subsequence clustering is considered as the clustering of segments obtained from a time series segmentation algorithm.

- Time point clustering combine the temporal proximity of time points with the similarity between their corresponding values.

1.5.4 Time series to image transformation

The transformation of 1D time series to 2D image-like representation is interesting given its use for other posterior tasks, such as classification or clustering. The transformation of time series to 2D image-like representation is highly motivated by the subsequent application of DL techniques, such as convolutional neural networks (CNNs). This line of research is being tackled at the moment of writing this Thesis, and therefore, is a future research. More concretely, it is related to the development of a novel technique for TSC based on the use of different CNN techniques applied to the 2D image-like representation of the time series.

This idea has been previously approached in the literature in many and varied forms. For instance, Sezer and Ozbayoglu in [166] presented a novel technique in which 1D financial time series are transformed to images by computing 15 different technical indicator instances with various parameter settings, resulting each time series in 15×15 pixels sized images. Furthermore, Hatami *et al.* in [85] developed a framework for TSC in which recurrence plots (RPs) are used for encoding 1D time series into texture images; once the time series are transformed to images, a CNN is applied. One of the main advantages of using RPs is the visualisation of significant details of the phase space trajectory through the 2D images. Moreover, Chen and Shi in [38] proposed a novel technique using the relative position matrix which converts preprocessed time series to texture images. The preprocessing step applied to the time series consists in a dimensionality reduction stage, carried out by using the piecewise aggregation approximation (PAA) method.

There are some other approaches following the idea of converting 1D time series data to 2D image-like data. Said and Erradi in [157], a deep learning framework was proposed for the gap forecasting between mobile crowdsourced service supply and the demand at a given time and space. In this paper, the time series are encoded by using the gramian angular summation field (GASF), the gramian angular difference field (GADF) and RPs. After the conversion of time series, CNNs are applied. Mo *et al.* in [128], the time series are firstly mapped into two-dimensional greyscale images by a sliding window approach. This approach is applied to classify, locate and detect network traffic data, aimed to detect network traffic outliers. Besides, Olivier and Aldrich in [135] presents a novel approach for dynamic monitoring of grinding circuits, in which a modified version of RPs is used to generate the images from the original time series. This version, also known as global RP, consists in removing the threshold function in order to provide a global view of the trajectories in all neighbourhoods. This idea has also been applied to multivariate time series by Yang *et al.* in [194], in which sensor classification is carried out by converting multivariate time series sensor data into two-dimensional coloured images, concatenating them for the subsequent application of CNN. The methods used to create the images were GASF, GADF and Markov transition fields (MTFs).

To sum up, although the way original time series are converted to images changes from one paper to another, the main idea behind all these approaches is to apply CNNs to the obtained images. In this way, the subsequent application of CNN architectures aims to automatically learn a high-level abstract representation of low-level raw time series data. Moreover, many of the previous works follow the general workflow presented in

Figure 1.5.4, in case TSC was the task to be performed. To carry out the dimensionality reduction stage, any of the segmentation procedures existing in the literature [106] could be applied. Once the dimensionality-reduced time series has been obtained, the next step is developing the mapping function to get the image-like representation. It could be done by computing the mean of the segments or by using any distance matrix, and then, applying a colour mapping to the obtained matrix.

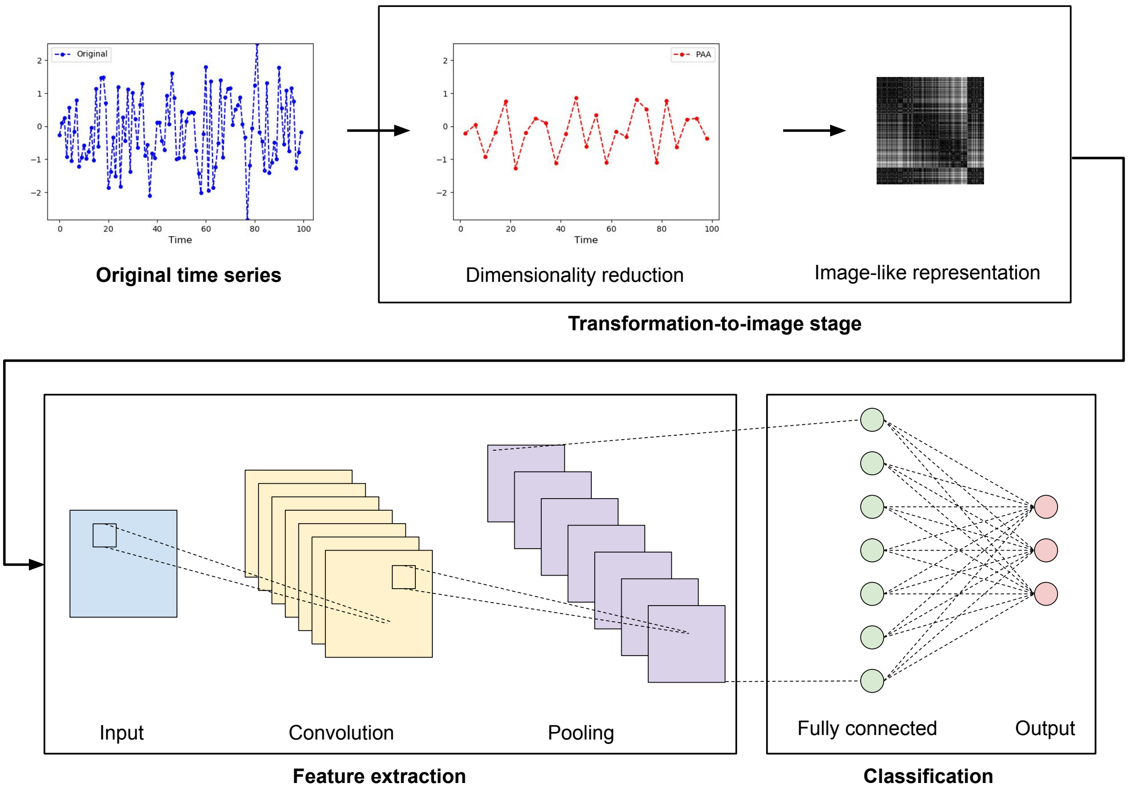

Figure 1.5.4: Overall workflow for TSC via time series to image transformation.

1.6 Applications in real-world problems

In this Thesis, several real-world problems have been considered in order to validate the methodologies proposed, and to address the difficulties or issues found for these applications. In this way, the real-world problems selected could be divided into the following groups:

1. Atmospheric events: prediction of fog formation in the airport of Valladolid (Spain), and prediction of convective cloud formation in the airport of Madrid (Spain).

2. Engineering applications: prediction of solar energy and prediction of energy-flux, which are related with renewable energy processes, and, on the other hand, mod-

elling of desiccant wheels and modelling of the acoustic behaviour produced by induction motors.

3. Health: determination of a patient typology with human immunodeficiency virus (HIV)/hepatitis C virus (HCV), and donor-recipient matching in liver transplantation.

1.6.1 Atmospheric events

This group of applications concerns the prediction of different meteorological situations, which occur constantly and in many forms. In this way, the prediction of low-visibility events due to fog or the prediction of convective cloud formation are of significant impact. The data for both events are collected in airports, specifically, in the Valladolid and Madrid airports (Spain), respectively.

Fog prediction

Predicting low-visibility events is very important in many human activities, and crucial in transportation facilities such as airports, where they can cause severe impact in flight scheduling and safety. One of the main factors reducing the visibility is the presence of fog [22]. Fog is a meteorological phenomenon consisting of the suspension of tiny, usually microscopic, water droplets in the air, reducing the horizontal visibility at the Earth's surface to less than 1 km [190].

The presence of fog makes the airport managers activate specific low-visibility procedures to sustain safely operations in reduced visibility. In this way, there a lot of potential problems such as larger time-intervals during landing and taking-off operations, increases in traffic controllers and pilots workload, and suspension of runway operations, among others.

There are two main ways for addressing this problem: 1) predicting the fog formation in real-time, where an example in this line was presented by Ahmed *et al.* in [5], who proposed the use of a bi-spectral brightness temperature difference technique along with satellite images, while another example is the one presented by Dey in [53], in which the use of the brightness temperature difference technique is discussed; and 2) predicting visibility at different time-horizons, where, for instance, Colabone *et al.* in [41] proposed the use of a multilayer perceptron (MLP) for hourly fog forecasting, whereas [28, 57, 60] predict the occurrence of fog with a prediction time-horizon up to 3h, 6h and 18h, respectively. For both approaches, the prediction can be based both on previous values, and on external data related to fog prediction. However, the main drawback associated with autoregressive models (ARs) is, as described in Section 1.5.1, the selection of the order p.

Generally, low-visibility events are characterised through the runway visual range (RVR), which is a meteorological variable defined as the vision range of a pilot for seeing the runway surface markings, or the lights delineating the runway [136]. Following the guidelines provided in [95], RVR values above 2000 metres are not significant for airport operational purposes regarding visibility. However, if the RVR is below 2000 metres, three categories can be differentiated: poor visibility ($0 \leq RVR < 1000$), medium visibility ($1000 \leq RVR < 1990$), and situations that do not have a significant impact on the aeronautical operations ($1990 \leq RVR$). It can be seen that these categories (visibility condition at the airport) show a natural order. Figure 1.6.1 includes a multivariate time series of all the input atmospheric time series, in which the category of the day is extracted from categorising the RVR. Note that most of the days are misty or foggy, and that standard regression could not be possible as RVR is truncated (when the value is higher that 2000, the exact distance is not known).

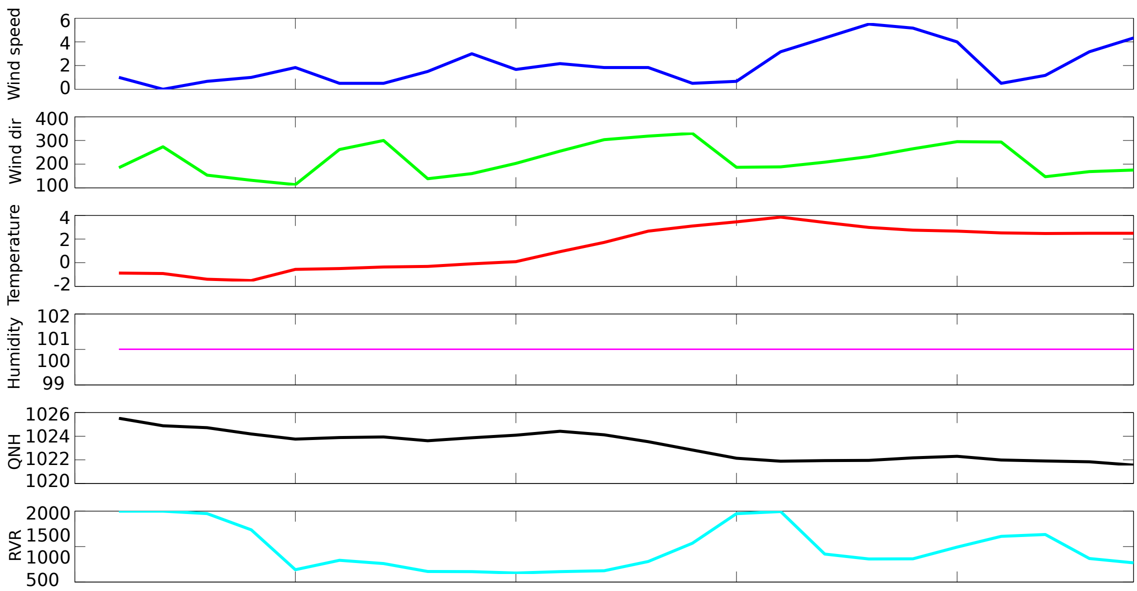

Figure 1.6.1: Example of a multivariate time series, in which the output time series RVR reflects foggy days, poor visibility days corresponding to $0 \leq RVR < 1000$.

Most of the works published in the literature mainly tackle the problem of fog prediction as a regression approach. Therefore, in this Thesis, we tackled this problem from the ordinal classification (OC) point of view.

Convective cloud prediction

Another atmospheric phenomenon vastly affecting human activities and transportation facilities are convective cloud formations. As previously mentioned, anticipating to extreme weather conditions such as convective cloud formation is an arduous task for the operational weather forecasters [32]. It is well-known in the forecaster community that stability indices derived from temperature and humidity data are widely used for predicting convective situations. Some of them are the lifted index [69], the total of totals index [127], or the convective available potential energy [130], among others.

Normally, the combination of these indices is used to improve the prediction results obtained individually [159]. In this sense, there are several approaches. For example, the one proposed by Sánchez *et al.* in [158], who performed an analysis of the pre-convective conditions in Argentina based on the information of 713 days of radiosonde data. Moreover, Púčik *et al.* in [141] used 16421 proximity soundings taken at 32 central European stations to assess environments of severe and non-severe thunderstorms.

Regarding the kind of data characterising the occurrence of convection, it has been assumed that the resolution of surface-station observations is not enough, and that the numerical weather prediction (NWP) models are not able to accurately predict the exact location and time of the convective situation [205]. Therefore, satellite information to forecast the occurrence of convection could help standard approaches identify atmospheric conditions that favours convection.

Thus, the combination of satellite information and stability indices with machine learning (ML) techniques could be adequate, given that they have been proved to be excellent in predicting accurately a wide range of local atmospheric phenomenal [123].

In this Thesis, we propose two different ways of tackling this problem: 1) from the multi-objective point of view: as the dataset is highly imbalanced, both accuracy (which measures the global performance) and minimum sensitivity (MS) (which measures the performance of the most difficult class) should be optimised, with the difficulty that both metrics are conflictive under some circumstances [64]. And, 2) by tackling the problem from the OC perspective: the forecasting is done considering the ordinal nature of the different convective situations.

1.6.2 Engineering applications

This group of applications concerns different engineering applications. The first two main applications consider the high environmental impact of current energy resources together with the need for addressing the impact of climate change. In this sense, an important development of renewable energy sources has been observed during past years. The international energy agency stated that electricity generation from renewable energy is expected to rise up to 39% by 2050 [25]. Hence, in this Thesis, these two main applications are related to harvesting renewable energy: the prediction of solar radiation at the radiometric station of Toledo (Spain), and the flux of energy prediction in the Gulf of Alaska (USA). Furthermore, other two engineering applications are also addressed. They are related with the modelling of desiccant wheels, widely used as dehumidification systems, and the modelling of the acoustic behaviour of induction motors.

Solar energy prediction

Among renewable resources, solar energy is one of the most developed ones, being widely accepted as the future renewable source, mainly due to the high availability of the solar resource. Furthermore, solar energy, as source of energy, is clean, sustainable, and extremely abundant [73]. One of the main drawbacks of solar energy, when compared to others renewable resources, is the difficulty for managing its inherent intermittence. In this sense, the estimation of solar energy availability and the optimal management of energy system in any part of the world are challenging tasks.

In the last years, several approaches have been presented in the literature to improve the performance achieved by standard methods [8, 165] or by classical astronomical equations [97]. However, they have several disadvantages that NWP models counteracts by being able to model the dynamics of the atmosphere, as well as the physical processes involved, employing a set of equations based on physical laws of motion and thermodynamics [139].

Furthermore, in recent years, ML techniques has become an important alternative/complement to NWP in solar energy prediction problems, given its ability to obtain reliable and accurate forecasts. Some techniques combine NWP models with ML algorithms, such as [180], in which hourly global radiation for five places in Mediterranean area was performed combining a hybrid autoregressive moving average model (ARMA)/artificial neural network (ANN) model with ALADIN, a well-known NWP model. Ghimire *et al.* in [72] proposed a deep learning (DL) hybrid model to predict solar radiation in two phases, firstly a convolutional network is used to extract features, and secondly, a long-short-time memory (LSTM) network is applied for the prediction stage. Moreover, Alharbi in [7] presented a case study of solar radiation prediction in Arabia Saudi comparing the performance of ANN with classical training and extreme learning machines (ELMs).

The use of satellite data also contributes to an increase in the performance of the prediction stage. For example, Zarzalejo *et al.* in [201] used fuzzy logic and ANNs for estimating hourly global radiation from satellite images, or Şahin *et al.* in [156] presented an approach combining ELMs with satellite data and geographic variables to predict solar radiation over Turkey.

In this Thesis, given the wide variety of methods proposed in the state-of-the-art, we present a robust comparison between several ML regressors when applied to solar radiation estimation problem. The comparison should be performed not only against ML algorithms, but also against NWP models based on satellite measurements such a Copernicus atmosphere monitoring service (CAMS) or *SolarGIS©*.

Energy flux prediction

Other renewable and eco-friendly sources of energy are tides and waves. These sources have three main advantages: 1) they are always available (i.e. their nature is not intermittent), 2) they can be used in many parts of the world, including oceans or big seas, and 3) given that the energy storage is limited and that around 40% of the world's population live within 100 km of the coast, a cost-effective transmission of electricity is possible.

The behaviour of wave energy have been modelled from different points of view [133]: 1) physical models [94], 2) statistical models [118], and 3) by the application of ML techniques [47], among others. Focusing on ML techniques, there are a wide range of approaches in the state-of-the-art. Cornejo-Bueno *et al.* in [44] presented a grouping genetic algorithm applied to ELMs to predict the significant wave height and the flux of energy. Fernández *et al.* in [65] proposed the use of meteorological data, obtained from national center for atmospheric research (NCAR) Reanalysis project, and applied several ordinal classifiers for the prediction of both significant wave height and flux of energy. Kumar *et al.* in [111] used an ensemble of ELMs for the prediction of significant wave height in 10 stations of varying terrains from Gulf of Mexico, Brazil and Korean region. Note that, for the estimation of the flux of energy, buoys parameters (significant wave height and average wave period) need to be computed beforehand.

Furthermore, given that the data is collected from sensors located at the buoys, it is typical that the buoys break down due to unexpected events, resulting in discontinuities in the buoys data time series [145]. Therefore, using reanalysis variables as input data, where no observed data is required, not only does it improve its applicability to other sites, but also it avoids problems related with missing data. Figure 1.6.2 shows the energy flux time series collected during 2018 from buoy 46001 located in the Gulf of Alaska.

Regarding this problem, in this Thesis we propose a multi-task evolutionary artificial neural network (MTEANN) model able to tackle short- and long-term energy flux prediction (6h, 12h, 24h and 48h) simultaneously considering ML algorithms and reanalysis data for developing accurate flux of energy prediction models.

Desiccant wheels modelling

In order to maintain the required indoor conditions in buildings with high latent loads, for example in the food and pharmaceutical industries, dehumidification systems are necessary. Conventional dehumidification systems depend mainly on electrical energy, which should be decreased to reduce the environmental impact associated with air dehumidifica-

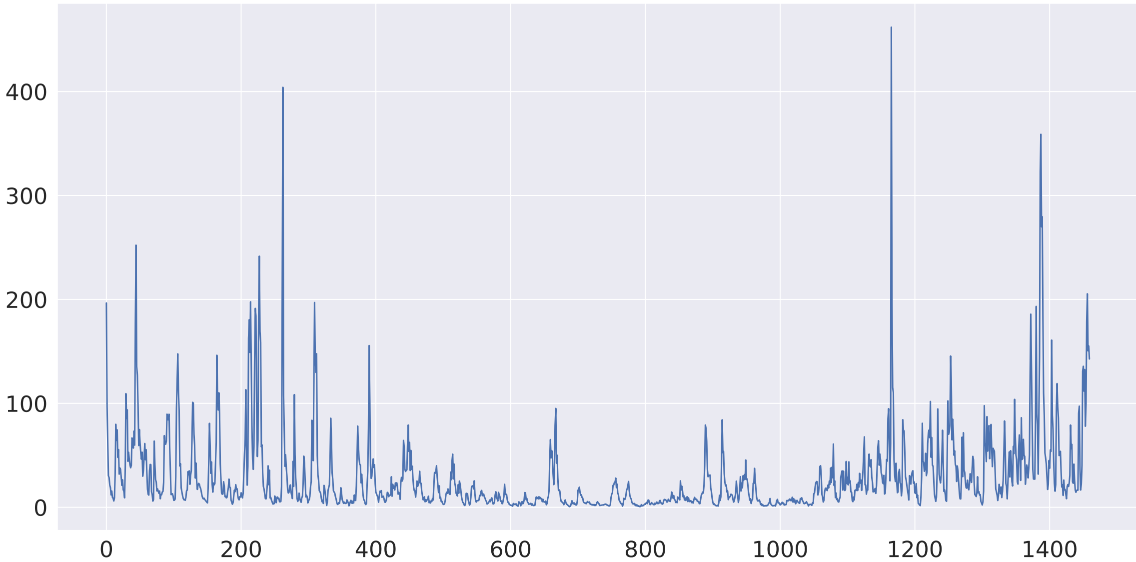

Figure 1.6.2: Energy flux time series collected from buoy 46001 located in the Gulf of Alaska.

tion. In this sense, desiccant wheels (DWs) could be a serious alternative to conventional dehumidification systems. They are based on direct expansion units, which also depend on electrical energy. However, an acceptable dehumidification capacity is achieved when the DW are activated at low temperature, hence reducing the pollution associated to air dehumidification.

In this sense, modelling the behaviour of these devices is an interesting research line, given that the mathematical models require low computational load. Several different mathematical models have been used in the literature to adjust the behaviour of the DWs, from first- or second-order equations [42] to multi-objective genetic algorithms combined with response surface methodology [202]. More recently, complex models have been applied, such as a multi-objective genetic algorithm combined with response surface methodology [202], or [99], in which, ANNs are trained by means of a standard back-propagation (BP) algorithm.

Based on the limitations of the previous articles, in this Thesis, we propose a MTEANN able to predict, simultaneously, the outlet process air temperature and the outlet process air humidity ratio.

Induction motors acoustic behaviour modelling

Electric induction motors are widely used in industrial and household applications, from small electrical devices to large industrial machinery and vehicles. One of the objects of research regarding induction motors is their noise component, given that people suffer from physical and psychological discomfort when they are exposed to their noise during a long time.

Typically, in the literature, the noise generated by electric motors has been modelled in two main ways: by the electric power signal and mechanical vibration [24], and by recording the noise in a semi-anechoic chamber [61]. However, tackling noise prediction without separating the noise of the electric motor from the environmental noise is of significant interest. Therefore, the equivalent sound pressure level is used to measure the physical discomfort, whereas the loudness, roughness and sharpness are used to measure the psychological discomfort. Note that it is easier to measure the electrical parameter than to register the complete sound and then calculate the sound quality parameters.

ANNs have been successfully used in this field, Huang *et al.* in [93] proposed the use of a regression-based deep belief network combined with a support vector machine (SVM) for assessing the sound quality of vehicle interior. Tenenbaum *et al.* in [172] presented an approach based on ANNs to determine the annoyance of different electric vehicle sounds for a constant speed, single car pass-by situation. Furthermore, ANNs have been also used to develop a new technique to produce fast and robust auralizations for room acoustics simulation [172]. Regarding the prediction of the acoustic behaviour of induction motors, in this Thesis, we propose the use of MTEANNs in order to predict the four aforementioned measures simultaneously.

1.6.3 Health

This group of applications includes two main health problems. The first one is related with determining the typology of patients with HIV/HCV to be treated with antivirals, whereas the second one concerns organ transplantation, more concretely addressing the donor-recipient matching in liver transplantation (LT).

HIV/HCV patient typology

Chronic HCV is a major cause of cirrhosis, LT, and liver-related deaths worldwide [187]. Given that HCV and HIV are found to be transmitted together, patients are usually infected with both viruses [188]. In the last years, several researchers and doctors have claimed for an universal treatment of this disease [175]. However, given the huge number of patients requiring the treatment, prioritisation criteria has been established by scientific societies and health authorities.

These prioritisation criteria are aimed to achieve the maximum survival rate and best clinical benefits for the patients. Nevertheless, they have not been assessed. In this way, identifying the typology of patient requiring the treatment first is an important issue. Moreover, recognising those patient-related variables limiting treatment uptake in

HIV/HCV co-infected patients is also an interesting task. For these purposes, in this Thesis, we propose the use of ANNs, given that these models could achieve exceptional accuracy.

ANNs have been used in the field previously: Wang *et al.* [181] presented an ANN approach to predict virological response to therapy from HIV genotype. On the other hand, Resino *et al.* in [150] proposed the use of ANNs for the prediction of significant fibrosis among HIV/HCV co-infected patients, reaching to the conclusion that ANNs are useful and helpful for guiding therapeutic decisions in HIV/HCV co-infected patients.

Machine learning in organ transplantation

ML techniques play an important role in organ transplantation. In the last decade, there has been an explosion of interest in data mining associated to organ transplantation, with numerous approaches proposed in the literature aiming to find universal models or, at least, models for multi-centre cohorts or from different countries [74].

ML can be applied to donor-recipient matching, whose objective is to maximise the probability of graft survival after a certain time. One of the main benefits of using ML for this task is that the decision process is objective, given that there is no human subjectivity in the selection of the donors and recipients, among others [12]. Numerous works have been published in the literature: Shadabi *et al.* [167] estimated the graft survival probability in kidney transplantation using an ANN to predict if the graft survived after a certain time. Briceño *et al.* [29] presented a novel donor-recipient matching model to improve the performance of current clinical decision-making systems in LT.

Moreover, the rationale of assigning a given donor to a potential candidate on the waiting list is causing some controversy, specially for LT, where there is an increasing number of candidates and a scarce number of available donors. Several scores have been introduced in the literature aiming to solve this issue, such as the model for end-stage liver disease (MELD) [102], among others. However, all these scores have their supporters and detractors, for instance, in the case of MELD, its aim is to decrease the mortality in the waiting list without affecting the result of the transplant. Therefore, there is a need of a balanced score to maximise the survival benefit among all possible candidates, without perjuring the most critical receipts on the waiting list.

In this Thesis, we present the current opinion, and a review of the state-of-the-art in ML methods when applied to organ transplantation. Moreover, we carry out a comparison between statistical methods and ML techniques for donor-recipient matching, when applied to LT.

The goal is to turn data into information,
and information into insight.

Carly Fiorina

2 Motivation and objectives

This chapter introduces the motivation, challenges, and objectives considered in this Thesis, as well as the main publications derived from it.

2.1 Motivation and challenges

First of all, the development of this Thesis begins with works based on applying artificial neural networks (ANNs) to real-world problems, mainly using time series data, given that several research projects involve providing machine learning (ML) based solutions for this kind of data. Getting deep into ANNs models will lead to other challenges and objectives more specifically related with time series. In this sense, from the previous chapter, we can differentiate between three main large areas based on time series data mining. Firstly, time series clustering is a challenging task based on identifying interesting groups in time series datasets. Secondly, one of the most popular goals regarding time series data mining is time series classification (TSC). Specifically, during the last years, many approaches have been published in the state-of-the-art, mainly applied to nominal time series, however, when dealing with ordinal time series, the performance of these classification tasks could be improved by means of ordinal techniques, nevertheless, time series ordinal classification (TSOC) is yet an unexplored field. And, finally, the development of time series prediction models taking nominal, ordinal and regression predictions into account is interesting for comparing the difference of performance between the different points of view. Furthermore, and from a transversal point of view common to the three previous areas, the preprocessing of time series is a field with significant impact given that, nowadays, time series datasets include a vast amount of information, and this field can help alleviate the burden of subsequent tasks. Moreover, apart from time series data mining approaches, non-temporal data regression has also been considered, providing interesting ideas and techniques to solve engineering and health applications.

Based on these comments, we can synthesise the following open challenges:

- Time series preprocessing: previous to the application of other tasks such as classification, prediction or clustering, preprocessing the time series has a significant impact on the performance of subsequent tasks. In this sense, the preprocessing could be done in several ways: recovering missing data from the time series or mapping the original time series to a different form of representation by projecting the segments obtained during a time series segmentation into vectors of statistical features, among others.

- Time series clustering: one of the main contributions of this Thesis will be made in the field of time series clustering. There are several ways for performing a clustering task applied to time series, depending on the way time series are treated. Specifically, standard approaches only search for similarities between the different time series, and then, cluster them following these distances. However, it has been demonstrated that the application of preprocessing techniques, such as a time series segmentation, has some advantages. In this way, we consider that it could be interesting to exploit the similarities found in subsequences of the time series. Besides, few works can be found in the literature for clustering unequal-length time series. This could be affordable given that unequal-length time series are segmented into unequal-length segments, which are then projected into feature vectors of the same length, vastly reducing the size of the data without a large loss of information, in such a way that the subsequent clustering task increases its quality and performance.

- Time series classification: as stated previously, classification is the most popular task in the context of time series. The state-of-the-art includes a wide variety of approaches, from standard techniques (similarity/distance measures with a nearest neighbour classifier, for instance) to more complex deep learning (DL) models (residual networks (Resnet), for instance). The last approaches presented in the area try to maintain the accuracy achieved by the best methods, while being computationally effective. In this sense, hierarchical vote system collective of transformation-based ensemble (HIVE-COTE) is the one achieving the highest accuracy, but its computational cost is huge, given that it is an ensemble of classifiers built on different representations. Therefore, efforts should be made on alleviating the computational load associated to these techniques. Specifically, in this Thesis, we will focus on the shapelet transform (ST) [89], which has raised a lot of attention given its versatility. In this way, the main idea behind this transformation will be modified for both TSC and TSOC, respectively. Note that the field of TSOC has not been previously considered in the literature.

- Time series prediction: it is the most frequent task concerning time series, consisting in predicting the next values taking into account the previous values of the time series. An interesting approach is transforming the prediction model to a classification model in which both the next value and the next event could be determined. In this sense, several approaches could be considered, such as the use of autoregressive models (ARs) or the transformation from longitudinal data to transversal data. Besides, apart from some previous ideas, adapting novel ML techniques for classification and optimising them by advanced algorithms, such as evolutionary algorithms (EAs), will result in an increase of the performance measures.

- Real-world applications: the development of methods should also include its application to real-world problems. In this sense, we have considered a wide range of problems from different areas, such as atmospheric events, engineering applications and health.

2.2 Objectives

The aforementioned challenges will be addressed by different works presented in this Thesis. In order to specify more formally the previous challenges, the following objectives have been defined:

1. To propose different artificial neural network (ANN) architectures by hybridising activation functions or combining them in hidden and output layers, aiming to search for a good balance and to improve the performance achieved by standard models.

2. To adapt evolutionary artificial neural networks (EANNs) for its application to two different sorts of problems: multi-objective problems (MOPs), in which two objectives are considered during the optimisation of the ANNs, and multi-task problems, in which two problems are learned at the same time, providing useful information and making contributions to both tasks.

3. To review the state-of-the-art in preprocessing and analysis techniques for time series, with the aim of studying new representation forms alleviating the difficulty of subsequent tasks.

4. To study and develop a novel approach to time series clustering by preprocessing the time series with time series segmentation, reducing their dimensionality by carrying out a statistical feature extraction process.

5. To analyse and survey the shapelet transform (ST) methodology, in order to provide improvements to this methodology by developing a new proposal in the time series classification (TSC) field.

6. To adapt and develop a novel approach based on the ST technique for its application to ordinal data, opening a new branch in TSC known as time series ordinal classification (TSOC).

7. To apply the methods described above to the following real-world problems:

   (a) Prediction of fog formation in airports.
   (b) Prediction of convective situations formation in airports.
   (c) Prediction of solar radiation.
   (d) Prediction of energy flux from ocean waves.
   (e) Modelling of desiccant wheels (DWs).
   (f) Modelling of the acoustic behaviour of induction motors.
   (g) Identification of human immunodeficiency virus (HIV)/hepatitis C virus (HCV) co-infected patient typology.
   (h) Donor-recipient matching in liver transplantation (LT).

2.3 Summary of the Thesis

In the last years, there has been an increase in the number of fields improving their standard processes by using machine learning (ML) techniques. The main reason for this is that the vast amount of data generated by these processes is difficult to be processed by humans. Therefore, the development of automatic methods to process and extract relevant information from these data processes is of great necessity, giving that these approaches could lead to an increase in the economic benefit of enterprises or to a reduction in the workload of some current employments. Concretely, in this Thesis, ML approaches are applied to problems concerning time series data. Time series is a special kind of data in which data points are collected chronologically. Time series are present in a wide variety of fields, such as atmospheric events or engineering applications. Besides, according to the main objective to be satisfied, there are different tasks in the literature applied to time series. Some of them are those on which this Thesis is mainly focused: clustering, classification, prediction and, in general, analysis.

Generally, the amount of data to be processed is huge, arising the need of methods able to reduce the dimensionality of time series without decreasing the amount of information. In this sense, the application of time series segmentation procedures dividing the time series into different subsequences is a good option, given that each segment defines a specific behaviour. Once the different segments are obtained, the use of statistical features to characterise them is an excellent way to maximise the information of the time series and simultaneously reducing considerably their dimensionality.

In the case of time series clustering, the objective is to find groups of similar time series with the idea of discovering interesting patterns in time series datasets. In this Thesis, we have developed a novel time series clustering technique. The aim of this proposal is twofold: to reduce as much as possible the dimensionality and to develop a time series clustering approach able to outperform current state-of-the-art techniques. In this sense, for the first objective, the time series are segmented in order to divide the them identifying different behaviours. Then, these segments are projected into a vector of statistical features aiming to reduce the dimensionality of the time series. Once this preprocessing step is done, the clustering of the time series is carried out, with a significantly lower computational load. This novel approach has been tested on all the time series datasets available in the University of East Anglia and University of California Riverside (UEA/UCR) time series classification (TSC) repository.

Regarding time series classification, two main paths could be differentiated: firstly, nominal TSC, which is a well-known field involving a wide variety of proposals and transformations applied to time series. Concretely, one of the most popular transformation is the shapelet transform (ST), which has been widely used in this field. The original method extracts shapelets from the original time series and uses them for classification purposes. Nevertheless, the full enumeration of all possible shapelets is very time consuming. Therefore, in this Thesis, we have developed a hybrid method that starts with the best shapelets extracted by using the original approach with a time constraint and then tunes these shapelets by using a convolutional neural network (CNN) model. Secondly, time series ordinal classification (TSOC) is an unexplored field beginning with this Thesis. In this way, we have adapted the original ST to the ordinal classification (OC) paradigm by proposing several shapelet quality measures taking advantage of the ordinal information of the time series. This methodology leads to better results than the state-of-the-art TSC techniques for those ordinal time series datasets. All these proposals have been tested on all the time series datasets available in the UEA/UCR TSC repository.

With respect to time series prediction, it is based on estimating the next value or values of the time series by considering the previous ones. In this Thesis, several different approaches have been considered depending on the problem to be solved. Firstly, the prediction of low-visibility events produced by fog conditions is carried out by means of hybrid autoregressive models (ARs) combining fixed-size and dynamic windows, adapting itself to the dynamics of the time series. Secondly, the prediction of convective cloud formation (which is a highly imbalance problem given that the number of convective cloud events is much lower than that of non-convective situations) is performed in two completely different ways: 1) tackling the problem as a multi-objective classification task by the use of multi-objective evolutionary artificial neural networks (MOEANNs), in which the two conflictive objectives are accuracy of the minority class and the global accuracy, and 2) tackling the problem from the OC point of view, in which, in order to reduce the imbalance degree, an oversampling approach is proposed along with the use of OC techniques. Thirdly, the prediction of solar radiation is carried out by means of evolutionary artificial neural networks (EANNs) with different combinations of basis functions in the hidden and output layers. Finally, the last challenging problem is the prediction of energy flux from waves and tides. For this, a multitask EANN has been proposed aiming to predict the energy flux at several prediction time horizons (from 6h to 48h). All these proposals and techniques have been corroborated and discussed according to physical and atmospheric models.

The work developed in this Thesis is supported by 11 JCR-indexed papers in international journals (7 Q1, 3 Q2, 1 Q3), 11 papers in international conferences, and 4 papers in national conferences.

2.4 Publications

The following papers have been published in international journals (J):

J1 **D. Guijo-Rubio**, P.A. Gutiérrez, C. Casanova-Mateo, J. Sanz-Justo, S. Salcedo-Sanz and C. Hervás-Martínez. "Prediction of low-visibility events due to fog using ordinal classification", Atmospheric Research, Vol. 214, 2018, pp. $64-73$.
JCR (2018): 4.114. Position: 13/86 (Q1).
DOI: 10.1016/j.atmosres.2018.07.017

J2 F. Comino, **D. Guijo-Rubio**, M. R. de Adana and C. Hervás-Martínez. "Validation of multitask artificial neural networks to model desiccant wheels activated at low temperature", International Journal of Refrigeration, Vol. 100. 2019, pp. $434-442$.
JCR (2019): 3.461 Position: 11/61 (Q1).
DOI: 10.1016/j.ijrefrig.2019.02.002

J3 A. Rivero-Juárez, **D. Guijo-Rubio**, F. Téllez, R. Palacios, D. Merino, J. Macías, J.C. Fernández, P.A. Gutiérrez, A. Rivero and C. Hervás-Martínez. "Using machine learning methods to determine a typology of patients with HIV-HCV infection to be treated

with antivirals", PLoS One, Vol. 15(1). 2020, pp. e0227188.
JCR (2019): 2.740 Position: 27/71 (Q2).
DOI: 10.1371/journal.pone.0227188

J4 **D. Guijo-Rubio**, A.M. Durán-Rosal, P.A. Gutiérrez, A. Troncoso and C. Hervás-Martínez. "Time series clustering based on the characterisation of segment typologies", IEEE Transactions on Cybernetics. 2020.
JCR (2019): 11.079 Position: 5/136 (Q1).
DOI: 10.1109/TCYB.2019.2962584

J5 **D. Guijo-Rubio**, P.A. Gutiérrez, C. Casanova-Mateo, J.C. Fernández, A.M. Gómez-Orellana, P. Salvador-González, S. Salcedo-Sanz and C. Hervás-Martínez. "Prediction of convective clouds formation using evolutionary neural computation techniques", Neural Computing and Applications, Vol. 32, 2020, pp. 13917 – 13929.
JCR (2019): 4.774 Position: 23/136 (Q1).
DOI: 10.1007/s00521-020-04795-w

J6 **D. Guijo-Rubio**, C. Casanova-Mateo, J. Sanz-Justo, P.A. Gutiérrez, S. Cornejo-Bueno, C. Hervás-Martínez y S. Salcedo-Sanz. "Ordinal regression algorithms for the analysis of convective situations over Madrid-Barajas airport", Atmospheric Research, Vol. 236, 2020, pp. 104798.
JCR (2019): 4.676 Position: 13/93 (Q1).
DOI: 10.1016/j.atmosres.2019.104798

J7 **D. Guijo-Rubio**, P.A. Gutiérrez y C. Hervás-Martínez. "Machine learning methods in organ transplantation", Current Opinion in Organ Transplantation, Vol. 25(4), 2020, pp. 399 – 405.
JCR (2019): 2.571 Position: 13/24 (Q3).
DOI: 10.1097/MOT.0000000000000774

J8 F.J. Jiménez-Romero, **D. Guijo-Rubio**, F.R. Lara-Raya, A. Ruiz-González y C. Hervás-Martínez. "Validation of artificial neural networks to model the acoustic behaviour of induction motors", Applied Acoustics, Vol. 166, 2020, pp. 107332.
JCR (2019): 2.440 Position: 9/32 (Q2).
DOI: 10.1016/j.apacoust.2020.107332

J9 **D. Guijo-Rubio**, A.M. Durán-Rosal, P.A. Gutiérrez, A.M. Gómez-Orellana, C. Casanova-Mateo, J. Sanz-Justo, S. Salcedo-Sanz y C. Hervás-Martínez. "Evolutionary artificial neural networks for accurate solar radiation prediction", Energy, Vol. 210, 2020, pp. 1 – 11.
JCR (2019): 6.082 Position: 3/61 (Q1).
DOI: 10.1016/j.energy.2020.118374

J10 **D. Guijo-Rubio**, A.M. Gómez-Orellana, P.A. Gutiérrez y C. Hervás-Martínez. "Short- and long-term energy flux prediction using Multi-Task Evolutionary Artificial Neural Networks", Ocean Engineering, Vol. 216, 2020, pp.108089.
JCR (2019): 3.068 Position: 1/14 (Q1).
DOI: 10.1016/j.oceaneng.2020.108089

J11 **D. Guijo-Rubio**, J. Briceño, P. A. Gutiérrez, M.D. Ayllón, R. Ciria, C. Hervás Martínez. "Comparison of statistical methods and machine learning techniques for donor-recipient matching in liver transplantation". PLos One, 2021.
JCR (2019): 2.740 Position: 27/71 (Q2). Accepted.

Also, some related works have also been published in the proceedings of international conferences (C):

C1 M. Dorado-Moreno, A. M. Durán-Rosal, **D. Guijo-Rubio**, P. A. Gutiérrez, L. Prieto, S. Salcedo-Sanz, and C. Hervás-Martínez. "Multiclass prediction of wind power ramp events combining reservoir computing and support vector machines". Conference of the Spanish Association for Artificial Intelligence (CAEPIA 2016). 2016. pp. 300 – 309.
DOI: 10.1007/978-3-319-44636-3_28

C2 A. M. Durán-Rosal, **D. Guijo-Rubio**, P. A. Gutiérrez, S. Salcedo-Sanz, and C. Hervás-Martínez. "A coral reef optimization algorithm for wave height time series segmentation problems". International Work-Conference on Artificial and Natural Neural Networks (IWANN 2017). 2017. LNCS. Vol. 10305, pp. 673 – 684.
DOI: 10.1007/978-3-319-59153-7_58

C3 A. M. Durán-Rosal, **D. Guijo-Rubio**, P. A. Gutiérrez, and C. Hervás-Martínez. "Hybrid Weighted Barebones Exploiting Particle Swarm Optimization Algorithm for Time Series Representation". Bioinspired Optimization Methods and their Applications (BIOMA 2018). 2018. LNCS. Vol. 10835, pp. 126 – 137.
DOI: 10.1007/978-3-319-91641-5_11

C4 **D. Guijo-Rubio**, A. M. Durán-Rosal, A. M. Gómez-Orellana, P. A. Gutiérrez, and C. Hervás-Martínez. "Distribution-based discretisation and ordinal classification applied to wave height prediction". 19th International Conference on Intelligence Data Engineering and Automated Learning (IDEAL 2018). 2018. LNCS, Vol. 11315, pp. 171 – 179.
DOI: 10.1007/978-3-030-03496-2_20

C5 **D. Guijo-Rubio**, P.J. Villalón-Vaquero, P.A. Gutiérrez, M.D. Ayllón, J. Briceño y C. Hervás-Martínez. "Modelling survival by machine learning methods in liver trans-

plantation: application to the UNOS dataset". 20th International Conference on Intelligent Data Engineering and Automated Learning (IDEAL 2019). 2019. LNCS, Vol. 11872, pp. $97-104$.
DOI: 10.1007/978-3-030-33617-2_11

C6 **D. Guijo-Rubio**, P.A. Gutiérrez, R. Tavenard y A. Bagnall. "A hybrid approach to time series classification with shapelets". 20th International Conference on Intelligent Data Engineering and Automated Learning (IDEAL 2019). 2019. LNCS, Vol. 11871, pp. $137-144$.
DOI: https://doi.org/10.1007/978-3-030-33607-3_16

C7 **D. Guijo-Rubio**, P.A. Gutiérrez, A. Bagnall y C. Hervás-Martínez. "Time series ordinal classification via shapelets". 2020 IEEE International Joint Conference on Neural Networks (IJCNN 2020). 2020. Glasgow, UK. pp. $1-8$.
DOI: 10.1109/IJCNN48605.2020.9207200

C8 **D. Guijo-Rubio**, P.A. Gutiérrez, A. Bagnall y C. Hervás-Martínez. "Ordinal versus nominal time series classification". 5th Workshop on Advances Analytics and Learning on Temporal Data (AALTD 2020). 2020. LNAI, Vol. 12588, pp. $19-29$.
DOI: 10.1007/978-3-030-65742-0_2

C9 F. J. Rodríguez-Lozano, **D. Guijo-Rubio**, P.A. Gutiérrez, J. M. Soto-Hidalgo y J. C. Gámez-Granados. "Enhancing the ORCA framework with a new Fuzzy Rule Base System implementation compatible with the JFML library". IEEE International Conference on Fuzzy Systems 2021 (Fuzz-IEEE 2021). 2021. Accepted.

C10 **D. Guijo-Rubio**, V. M. Vargas, P. A. Gutiérrez, and C. Hervás-Martínez. "Studying the effect of different Lp norms in the context of Time Series Ordinal Classification". Conference of the Spanish Association for Artificial Intelligence (CAEPIA 2021). 2021. Accepted.

C11 V. M. Vargas, **D. Guijo-Rubio**, P. A. Gutiérrez, and C. Hervás-Martínez. "ReLU-based activations: analysis and experimental study for deep learning". Conference of the Spanish Association for Artificial Intelligence (CAEPIA 2021). 2021. Accepted.

And, finally, the following contributions have been made in national conferences (NC):

NC1 **D. Guijo-Rubio**, A.M. Durán-Rosal, P.A. Gutiérrez y C. Hervás-Martínez. "Clustering de Series Temporales basado en la Extracción de Tipologías de Segmentos". I Congreso de Investigadores Noveles de la Universidad de Córdoba (CIN-UCO 2016). 2016. pp. $201-204$.

URL: http://www.uco.es/investigacion/ucci/es/congreso-cientifico-de-investigadores-noveles

NC2 M. Diaz-Lozano, **D. Guijo-Rubio**, P.A. Gutiérrez, C. Casanova-Mateo, S. Salcedo-Sanz y C. Hervás-Martínez. "Algoritmos de aprendizaje automático para predicción de niveles de niebla usando ventanas estáticas y dinámicas". IX Simposio Teoría y Aplicaciones de Minería de Datos (TAMIDA 2018). 2018. pp. 833 – 838.
URL: https://sci2s.ugr.es/caepia18/proceedings/docs/CAEPIA2018_paper_122.pdf

NC3 **D. Guijo-Rubio**, A.M. Durán-Rosal, P.A. Gutiérrez, A. Troncoso y C. Hervás-Martínez. "Time series clustering based on the characterisation of segment typologies". 3er Bilbao Data Science Workshop (BiDAS 3). 2018.
URL: https://wp.bcamath.org/bidas3/

NC4 **D. Guijo-Rubio**, P.A. Gutiérrez y C. Hervás-Martínez. "Predicción de altura de ola mediante discretización basada en distribuciones utilizando clasificación ordinal". VII Congreso Científico de Investigadores en Formación (CCIF-UCO 2019). 2019. vol. 3, pp. 641 – 644.
URL: https://www.uco.es/idep/doctorado-novedades/263-vii-congreso-cientifico-de-investigadores-en-formacion

Without big data, you are blind and deaf and in the middle of a freeway.

Geoffrey Moore

3 Time series preprocessing

This chapter presents a novel technique for time series clustering, which could be considered as a preprocessing technique for subsequent tasks such as classification or prediction of time series.

3.1 Time series clustering

Time series clustering is a field receiving a lot of attention in the last decade. In this sense, a novel approach has been presented based on the characterisation of segments obtained after carrying out a segmentation strategy over the time series.

Main publication associated to this section:

- **D. Guijo-Rubio**, A.M. Durán-Rosal, P.A. Gutiérrez, A. Troncoso and C. Hervás-Martínez. "Time series clustering based on the characterisation of segment typologies", IEEE Transactions on Cybernetics. 2020.
  JCR (2019): 11.079 Position: 5/136 (Q1).
  DOI: 10.1109/TCYB.2019.2962584

Other publications associated to this section:

- **D. Guijo-Rubio**, A.M. Durán-Rosal, P.A. Gutiérrez y C. Hervás-Martínez. "Clustering

de Series Temporales basado en la Extracción de Tipologías de Segmentos". I Congreso de Investigadores Noveles de la Universidad de Córdoba (CIN-UCO 2016). 2016. pp. $201-204$.
URL: http://www.uco.es/investigacion/ucci/es/congreso-cientifico-de-investigadores-noveles

- **D. Guijo-Rubio**, A.M. Durán-Rosal, P.A. Gutiérrez, A. Troncoso y C. Hervás-Martínez. "Time series clustering based on the characterisation of segment typologies". 3er Bilbao Data Science Workshop (BiDAS 3). 2018.
URL: https://wp.bcamath.org/bidas3/

3.1.1 Time series clustering based on the characterisation of segment typologies

In this paper, we present a novel technique for time series clustering, testing its performance over the whole set of datasets belonging to the University of East Anglia and University of California Riverside (UEA/UCR) time series classification (TSC) repository. The method proposed consists of two clustering stages: the first one consists in the application of a segmentation technique used to simplify the time series, whereas the second stage consists in the application of a final clustering algorithm for grouping the time series.

More concretely, the first stage is associated with the simplification of the time series. First of all, the SwiftSeg segmentation procedure [68] is carried out to divide each time series into segments. This segmentation algorithm introduces points of the time series iteratively into a growing window and generates the corresponding least-squares polynomial approximation of the segment. This segmentation involves the use of an error threshold for the segmentation procedure, SEP_{max}, which is the maximum error allowed for which the window is not further grown. After that, all these unequal-length segments are projected into a fixed-size vector in which each position corresponds with a different statistical feature extracted from the segments. At this point, all these vector of characteristics are equal-length and are clustered into $k = 2$ groups (higher values for k resulted in lower clustering quality, increasing the computational cost).

On the other hand, the second stage is associated with creating a common structure for all the time series belonging to the dataset, and clustering them. In this sense, first of all, a mapping step is carried out to generate a common representation from the cluster stage applied to the individual time series. This common representation consists in including the following information from each of the clusters: 1) the centroid, which is the average of all cluster points, and 2) the mapping of the segment with higher variance (which is one of the statistical features extracted from the segments) in order to include the most characteristic segment of the cluster. Besides, apart from this information, two more characteristics are included into this time series representation: 1) the error difference between the segment least similar to its centroid (farthest segment) and the segment most similar to its centroid (closest segment), and 2) the number of segments extracted from the time series. Therefore, this common representation for each time series has length $(w \times k) + v$, where w is the length of the mapped cluster, expressed as $w = (l \times 2)$ being l the length of the segment representation (all the statistical features extracted from the segment), it is $\times 2$ due to information of the centroid and the extreme segment are both included, k is the number of clusters, and $v = 2$ given that this is the extra information we are considering for the time series. After all this mapping, a final hierarchical clustering stage is performed with the idea of grouping similar time series in the same cluster.

According to the way of adjusting the parameter SEP_{max}, we defined two different strategies: 1) selecting the SEP_{max} leading to the best Caliński-Harabasz index (CH) index (known as $TS3C_{CH}$), and 2) selecting the SEP_{max} by means of a majority voting system in which a wide variety of internal measures is used (known as $TS3C_{MV}$).

This methodology is applied to 84 datasets from the UEA/UCR TSC repository, and the results achieved are compared against 3 state-of-the-art techniques. From the results, we can highlight that, for larger datasets, the methodology proposed, known as $TS3C_{CH}$, obtains better solutions than the rest of the methods, not only in terms of rand index (RI), but also achieving significantly lowest computational time. On the other hand, for medium-size time series, the results achieved by $TS3C_{MV}$ are similar to those achieved by $WDTW$ [100] and better than the rest of the approaches.

Some conclusions can be extracted from these results: 1) for large datasets with long time series, our method achieves a good performance in terms of RI; 2) the computational load associated with the segmentation process and the first hierarchical clustering is high, but the global cost is more than acceptable given that the final clustering does not depend on the length of the original time series; 3) for datasets with a large number of medium-length time series, the performance of our method is competitive, achieving good results in terms of RI; 4) our approach is designed for large time series datasets with medium-to-large-length time series, given that one of the advantages is the huge summarisation of information carried out in the first stage.

Time-Series Clustering Based on the Characterization of Segment Typologies

David Guijo-Rubio, Antonio Manuel Durán-Rosal, Pedro Antonio Gutiérrez, *Senior Member, IEEE*, Alicia Troncoso, and César Hervás-Martínez, *Senior Member, IEEE*

***Abstract*—Time-series clustering is the process of grouping time series with respect to their similarity or characteristics. Previous approaches usually combine a specific distance measure for time series and a standard clustering method. However, these approaches do not take the similarity of the different subsequences of each time series into account, which can be used to better compare the time-series objects of the dataset. In this article, we propose a novel technique of time-series clustering consisting of two clustering stages. In a first step, a least-squares polynomial segmentation procedure is applied to each time series, which is based on a growing window technique that returns different-length segments. Then, all of the segments are projected into the same dimensional space, based on the coefficients of the model that approximates the segment and a set of statistical features. After mapping, a first hierarchical clustering phase is applied to all mapped segments, returning groups of segments for each time series. These clusters are used to represent all time series in the same dimensional space, after defining another specific mapping process. In a second and final clustering stage, all the time-series objects are grouped. We consider internal clustering quality to automatically adjust the main parameter of the algorithm, which is an error threshold for the segmentation. The results obtained on 84 datasets from the UCR Time Series Classification Archive have been compared against three state-of-the-art methods, showing that the performance of this methodology is very promising, especially on larger datasets.**

***Index Terms*—Data mining, feature extraction, segmentation, time-series clustering.**

I. Introduction

TIME series are an important class of temporal data objects collected chronologically [1]. Given that they tend to be high dimensional, directly dealing with them in its raw format is very expensive in terms of processing and storage cost, which makes them difficult to analyze. However, time series have applications in many different fields of science, engineering, economics, finance, etc.

In recent years, there has been a high explosion of interest in mining time-series databases. Clustering is one of these data mining techniques, where similar data are organized into related or homogeneous groups without specific knowledge of the group definitions [2]. Usually, clustering is used as a preprocessing step for other data mining tasks.

Time-series clustering consists of grouping time series. There are several recent review papers dealing with time-series clustering [3]–[5]. It can be used as a preprocessing step for anomaly detection [6], for recognizing dynamic changes in the time series [7], for prediction [8], and for classification [9]. For example, the application of these techniques can be used to discover common patterns preceding important paleoclimate events [10] or mining gene expression patterns [11].

Time-series clustering can be approached by considering specific distance measures for time series combined with the standard clustering techniques [4], [12]. Some of these metrics are designed for equal-length time series, such as the standard Euclidean distance, which is applied to time series in [13], while others, such as the dynamic time warping (DTW) [14], [15], can be used for time series of different size, allowing the comparison of one-to-many points (i.e., it is an elastic measure). There have been many attempts to obtain better time-series distance metrics as extensions of DTW [16]–[19]. Moreover, apart from adapting distance measures, some authors propose specific versions of the clustering algorithm to deal with their special characteristics [20]. Interesting work about other types of similarity measures in time series and its effect over dataset size, accuracy, and speed can be found in [21] and [22].

On the other hand, time-series segmentation consists in cutting the series in some specific points, trying to achieve two different objectives: 1) dividing time series in segments as a procedure for discovering useful patterns (homogeneous segments) [10], [23]–[25] or 2) approximating the time series with a set of simple models for each segment without losing too much information [26]–[29].

These works of time-series segmentation open a new perspective for time-series clustering, given that previous time-series clustering proposals only search for similarities between

Manuscript received April 2, 2019; revised July 29, 2019; accepted December 22, 2019. This work was supported by the Spanish Ministry of Economy and Competitiveness and FEDER funds (EU) under Grant TIN2017-85887-C2-1-P and Grant TIN2017-90567-REDT. The work of David Guijo-Rubio was supported by the FPU Predoctoral Program (Spanish Ministry of Education and Science) under Grant FPU16/02128. This article was recommended by Associate Editor H. Wang. *(Corresponding author: David Guijo-Rubio.)*

David Guijo-Rubio, Pedro Antonio Gutiérrez, and César Hervás-Martínez are with the Department of Computer Science and Numerical Analysis, University of Córdoba, 14071 Córdoba, Spain (e-mail: dguijo@uco.es; pagutierrez@uco.es; chervas@uco.es).

Antonio Manuel Durán-Rosal is with the Department of Quantitative Methods, Universidad Loyola Andalucía, 14004 Córdoba, Spain (e-mail: amduran@uloyola.es).

Alicia Troncoso is with the Department of Computer Languages and Systems, University Pablo de Olavide, 41013 Seville, Spain (e-mail: atrolor@upo.es).

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCYB.2019.2962584

Prediction is very difficult, especially if it's about the future.

Niels Bohr

4
Time series prediction

The prediction of time series is considered the most important field regarding time series data mining, not only due the wide variety of objectives that need to be fulfilled, but also given that most of the tasks are highly challenging. In this sense, there are a huge variety of approaches to this problem. Traditionally, time series prediction has been accomplished by means of standard statistical procedures (such as autoregressive models (ARs), moving average models (MAs) or a mixture of them, such as autoregressive moving average model (ARMA) or autoregressive integrated moving average model (ARIMA), among others). Thus, in this Thesis, we have focused on time series prediction by transforming the prediction problem into classification or regression tasks. According to the strategy followed, more concretely, it could be accomplished by using ordinal classifications (OCs) or by using a regression point of view.

4.1 Time series ordinal prediction

This chapter shows some of the contributions of this Thesis based on the topic of time series prediction following the ordinal classifications (OCs) paradigm. In this way, this chapter includes the prediction of low-visibility events produced by atmospheric conditions, causing inconveniences in the proper operating conditions in the airports of Valladolid and Madrid-Barajas, respectively.

Main publications associated to this section:

- **D. Guijo-Rubio**, P.A. Gutiérrez, C. Casanova-Mateo, J. Sanz-Justo, S. Salcedo-Sanz and C. Hervás-Martínez. "Prediction of low-visibility events due to fog using ordinal classification", Atmospheric Research, Vol. 214, 2018, pp. 64 – 73.
  JCR (2018): 4.114. Position: 13/86 (Q1).
  DOI: 10.1016/j.atmosres.2018.07.017

- **D. Guijo-Rubio**, C. Casanova-Mateo, J. Sanz-Justo, P.A. Gutiérrez, S. Cornejo-Bueno, C. Hervás-Martínez y S. Salcedo-Sanz. "Ordinal regression algorithms for the analysis of convective situations over Madrid-Barajas airport", Atmospheric Research, Vol. 236, 2020, pp. 104798.
  JCR (2019): 4.676 Position: 13/93 (Q1).
  DOI: 10.1016/j.atmosres.2019.104798

Other publications associated to this section:

- **D. Guijo-Rubio**, A. M. Durán-Rosal, A. M. Gómez-Orellana, P. A. Gutiérrez, and C. Hervás-Martínez. "Distribution-based discretisation and ordinal classification applied to wave height prediction". 19th International Conference on Intelligence Data Engineering and Automated Learning (IDEAL 2018). 2018. LNCS, Vol. 11315, pp. 171 – 179.
  DOI: 10.1007/978-3-030-03496-2_20

- M. Diaz-Lozano, **D. Guijo-Rubio**, P.A. Gutiérrez, C. Casanova-Mateo, S. Salcedo-Sanz y C. Hervás-Martínez. "Algoritmos de aprendizaje automático para predicción de niveles de niebla usando ventanas estáticas y dinámicas". IX Simposio Teoría y Aplicaciones de Minería de Datos (TAMIDA 2018). 2018. pp. 833 – 838.
  URL: https://sci2s.ugr.es/caepia18/proceedings/docs/CAEPIA2018_paper_122.pdf

- **D. Guijo-Rubio**, P.A. Gutiérrez y C. Hervás-Martínez. "Predicción de altura de ola mediante discretización basada en distribuciones utilizando clasificación ordinal". VII Congreso Científico de Investigadores en Formación (CCIF-UCO 2019). 2019. vol. 3, pp. 641 – 644.
  URL: https://www.uco.es/idep/doctorado-novedades/263-vii-congreso-cientifico-de-investigadores-en-formacion

4.1.1 Prediction of low-visibility events due to fog using ordinal classification

In this paper, the prediction of low-visibility events due to fog is carried out. Anticipating to meteorological phenomena that can have a severe impact on daily human activities is crucial, especially for transportation facilities such as airports, in which altering schedules or safety has a significant repercussion on the economical side (lower traffic load for the airspace capacity), producing, in this sense, larger time-intervals between landings and take-offs, increases in the controllers and pilots workload or suspension of the runway operations, among others.

Given the importance of predicting the low-visibility events due to the presence of fog, we have centred our study in one of the airports in which the number of reduced visibility days is large, the Valladolid airport. The low visibility events have been recorded hourly and can be characterised by the runway visual range (RVR), measured from visibilimetres located along the runway. Furthermore, as fog can be modelled by meteorological variables, wind speed and direction, temperature, relative humidity and atmospheric pressure (known as QNH) have been used.

In this work, rather than predicting the RVR from a regression point of view, we propose using a categorical perspective, with three main categories: FOG ($0 \leq \overline{RVR} < 1000$), MIST ($1000 \leq \overline{RVR} < 1990$) and CLEAR ($1990 \leq \overline{RVR}$), where $\overline{RVR}$ is the average of the RVR of a day. In this sense, the prediction is performed on a daily basis. Furthermore, we propose three different kinds of windows for generating a set of input variables, based on previous values of the original time series and focused on autoregressive models (ARs). The first approach is the fixed window, which resembles the AR structure, i.e. it includes the n previous events of the time series. On the other hand, we have also proposed the use of two dynamic windows: 1) dynamic windows based on label changes, i.e. past values of the time series are included until there is a change in the RVR label, and 2) dynamic windows based on variance change, i.e. the same previous idea is carried out, but rather than expecting a RVR label change, past values are added to the window until the variance of the information included does not exceed a given threshold.

Then, the transformed dataset is built using the different windows and four statistical features (mean, variance, skewness and autocorrelation). For instance, the fixed windows are constructed by computing the variance, the skewness and the autocorrelation coefficient, whereas the dynamic windows are constructed by computing the average. Once the original dataset has been transformed, the corresponding classifier is trained. In this way, given that the outcome is categorical, and due to the nature of the variable, this problem is tackled as an OC one. Four different ordinal classifiers and a nominal one are applied to the transformed dataset. The results demonstrate that the combination of the three previously described windows along with the use of the kernel discriminant analysis for ordinal regression (KDLOR) classifier, achieves the best performance in terms of both minimum sensitivity (MS) and average mean absolute error (AMAE), not only against the rest of the proposals, but also against the persistence approach (these models predict the same category that the one observed for the previous event, in this case, the previous day), whose performance is considerably high, given that it will be right for long runs, only failing when there are changes.

Atmospheric Research 214 (2018) 64–73

Contents lists available at ScienceDirect

Atmospheric Research

journal homepage: www.elsevier.com/locate/atmosres

Prediction of low-visibility events due to fog using ordinal classification

D. Guijo-Rubio[a,*], P.A. Gutiérrez[a], C. Casanova-Mateo[b,c], J. Sanz-Justo[b], S. Salcedo-Sanz[d], C. Hervás-Martínez[a]

[a] Department of Computer Science and Numerical Analysis, Universidad de Córdoba, Córdoba, Spain
[b] LATUV: Remote Sensing Laboratory, Universidad de Valladolid, Valladolid, Spain
[c] Department of Civil Engineering: Construction, Infrastructure and Transport, Universidad Politécnica de Madrid, Madrid, Spain
[d] Department of Signal Processing and Communications, Universidad de Alcalá, Alcalá de Henares, Madrid, Spain

ARTICLE INFO

Keyword:
Airports
Fog events prediction
Time series
Forecasting
Ordinal classification
Time series preprocessing.

ABSTRACT

The prediction of low-visibility events is very important in many human activities, and crucial in transportation facilities such as airports, where they can cause severe impact in flight scheduling and safety. The design of accurate predictors for low-visibility events can be approached by modelling future visibility conditions based on past values of different input variables, recorded at the airport. The use of autoregressive time series forecasters involves adjusting the order of the model (number of past series values or size of the sliding window), which usually depends on the dynamical nature of the time series. Moreover, the same window size is normally used for all the data, thought it would be reasonable to use different sliding windows. In this paper, we propose a hybrid prediction model for daily low-visibility events, which combines fixed-size and dynamic windows, and adapts its size according to the dynamics of the time series. Moreover, visibility is labelled using three ordered categories (FOG, MIST and CLEAR), and the prediction is then carried out by means of ordinal classifiers, in order to take advantage of the ordinal nature of low-visibility events. We evaluate the model using a dataset from Valladolid airport (Spain), where radiation fog is very common in autumn and winter months. The considered data set includes five different meteorological input variables (wind speed and direction, temperature, relative humidity and QNH – pressure adjusted at mean sea level) and the Runway Visual Range (RVR), which is used to characterize the low-visibility events at the airport. The results show that the proposed hybrid window model with ordinal classification leads to very robust performance prediction in daily time-horizon, improving the results obtained by the persistence model and alternative prediction schemes tested.

1. Introduction

Fog is a meteorological phenomenon consisting of the suspension of very small, usually microscopic water droplets in the air, generally reducing the horizontal visibility at the Earth's surface to < 1 km (WMO 2011). Basically, fog is a cloud at ground level, that has been studied extensively from different points of view (Román Cascón 2015; Román-Cascón et al. 2012). Different fog types can be classified according to the two main physical processes which produce saturation of the air: cooling and the addition of water vapour. In the first group we have the radiation fog, that can occur in long nights and clear skies as a result of the thermal radiation cooling, the advection fog, that can arise when warm and moist air is moving over a colder surface, and the up-slope or orographic fog, formed when the air is forced up a slope undergoing an adiabatic cooling process. The second group includes in turn two main types: the steam fog, produced by rapid evaporation from an underlying warm water surface, and the frontal fog, caused by rain falling into cold air and moistening it. In all the cases, when the horizontal visibility is at least 1 km, but not > 5 km, the phenomenon is called *mist* (WMO 2011).

As it is well known, adverse weather phenomena can strongly affect air traffic management and flight operations (da Rocha et al. 2015; Dey 2018). Thunderstorms, icing, turbulence or wind shear conditions can greatly disrupt air traffic flows, leading to flight delays, diversions or cancellations. However, fog is perhaps the most important local visibility-reducing phenomenon that affects airport operations, since it can strongly reduce runway capacity (Bergot et al. 2007).

The number of runways fully available is a key element in the global airspace capacity, since they largely determine the number of departing or arriving flights within an airspace area. Therefore, any local weather

* Corresponding author at: Department of Computer Science and Numerical Analysis, University of Córdoba, Rabanales Campus, Albert Einstein Building 3rd Floor, 14071 Córdoba, Spain.
E-mail address: dguijo@uco.es (D. Guijo-Rubio).

https://doi.org/10.1016/j.atmosres.2018.07.017
Received 12 April 2018; Received in revised form 13 July 2018; Accepted 16 July 2018
Available online 20 July 2018
0169-8095/

4.1.2 Ordinal regression algorithms for the analysis of convective situations over Madrid-Barajas airport

In this paper, we tackle another problem related with meteorological phenomenon affecting daily human activities, concretely in airports. As stated before, anticipating to these meteorological events makes a great improvement in air and land operations at the airport. In this work, we deal with 12-h time horizon in the prediction of convective clouds present at the Madrid-Barajas airport, the one with most air traffic given the huge number of connections available.

As can be imagined, these meteorological events are not common in Spain and, therefore, this is why anticipating to them is of an enormous importance. The presence of convective clouds, especially of cumulus congestus and cumulonimbus, typically leads to airspace constraints, affects fuelling and handling operations, and, what is more significant, it could produce flight diversions. Handling with these circumstances cause overwhelming workload situations, in such a way that pilots need to change the planned routes or to activate special safety operations to avoid unnecessary conflicts. These events can be characterised by atmospheric data such as vertical temperature, wind speed and direction, water vapour content, and so on. Therefore, for this work, we have considered two sorts of variables: 1) derived from the radiosonde station located at the airport, and 2) derived from ECMWF ERA-Interim reanalysis project. Moreover, in order to avoid including overlapping information, we carried out correlation tests and removed all the variables with high correlation values.

On the other hand, the objective variable of our problem is ordinal, due to the nature of the problem. In this way, each event representing the presence of a given type of weather situation in the following 12h time-horizon is encoded as follows: CLEAR (no convective clouds sighted in the next 12h), TCU (cumulus congestus sighted in the next 12h, but neither cumulonimbus nor thunderstorms), CB (cumulonimbus sighted in the next 12h, but no thunderstorms) and TS (thunderstorm sighted in the next 12h). Note that there is an obvious order in which CLEAR periods are preferred than TS periods.

As was previously stated, CLEAR days are by far more common than any of the other categories, and, therefore, we have carried out an undersampling procedure removing a 30% of the training patterns with the CLEAR class. After that, an ordinal oversampling has been applied, given that the dataset is still highly imbalanced. Moreover, due to the ordinal nature of the problem, the oversampling technique must be ordinal given that otherwise the ordinal information present in the classes would be ignored.

Once the final oversampled dataset has been built, an ordinal classifier is applied. In this work, 13 ordinal classifiers are applied and compared against 4 nominal techniques, in order to demonstrate that ordinal approaches achieve better results. Furthermore, the results after applying the ordinal oversampling procedure are better than without applying it.

In addition, a comparison and verification of the results achieved is carried out to judge the quality of the proposed methodology. In this way, the results achieved by the best ordinal classifier are compared against terminal aerodrome forecasts (TAFs), which are tools that contribute towards the security, safety, and efficiency of international air navigation. Hence, the comparison demonstrated that our approach achieves better results than TAFs when dealing with extreme categories (TCU, CB and TS), whereas TAFs are generally better avoiding false alarms.

Atmospheric Research 236 (2020) 104798

Contents lists available at ScienceDirect

Atmospheric Research

journal homepage: www.elsevier.com/locate/atmosres

Ordinal regression algorithms for the analysis of convective situations over Madrid-Barajas airport

D. Guijo-Rubio[a], C. Casanova-Mateo[b], J. Sanz-Justo[c], P.A. Gutiérrez[a], S. Cornejo-Bueno[d], C. Hervás[a], S. Salcedo-Sanz[d,*]

[a] Department of Computer Science and Numerical Analysis, Universidad de Córdoba, Córdoba, Spain
[b] Department of Civil Engineering: Construction, Infrastructure and Transport, Universidad Politécnica de Madrid, Madrid, Spain
[c] LATUV, Remote Sensing Laboratory, Universidad de Valladolid, Valladolid, Spain
[d] Department of Signal Processing and Communications, Universidad de Alcalá, Alcalá de Henares, Spain

ARTICLE INFO

Keywords:
Convective clouds
Convective analysis
Airports
Machine learning techniques
Ordinal regression

ABSTRACT

In this paper we tackle a problem of convective situations analysis at Adolfo-Suarez Madrid-Barajas International Airport (Spain), based on Ordinal Regression algorithms. The diagnosis of convective clouds is key in a large airport like Barajas, since these meteorological events are associated with strong winds and local precipitation, which may affect air and land operations at the airport. In this work, we deal with a 12-h time horizon in the analysis of convective clouds, using as input variables data from a radiosonde station and also from numerical weather models. The information about the objective variable (convective clouds presence at the airport) has been obtained from the Madrid-Barajas METAR and SPECI aeronautical reports. We treat the problem as an ordinal regression task, where there exist a natural order among the classes. Moreover, the classification problem is highly imbalanced, since there are very few convective clouds events compared to clear days. Thus, a process of oversampling is applied to the database in order to obtain a better balance of the samples for this specific problem. An important number of ordinal regression methods are then tested in the experimental part of the work, showing that the best approach for this problem is the SVORIM algorithm, based on the Support Vector Machine strategy, but adapted for ordinal regression problems. The SVORIM algorithm shows a good accuracy in the case of thunderstorms and Cumulonimbus clouds, which represent a real hazard for the airport operations.

1. Introduction

Air traffic can be greatly disrupted by adverse weather conditions as pointed out in Borsky and Unterberger (2019). Turbulence, wind shear, low-visibility events or icing represent a serious risk to aviation. Specifically, convective clouds have a significant impact on airports operations (Lee and Craun, 2013; Bolgiani et al., 2018). Among the different types of convective clouds, Cumulus Congestus (also known as Towering Cumulus) and Cumulonimbus are particularly threatening when they form nearby the Terminal Manoeuvring Areas, or TMA (area of controlled airspace surrounding a major airport where there is a high volume of traffic). Convective clouds can lead to airspace constraints, increase airborne holding, affect fueling and handling operations, produce flight diversions (with the added problem that if the number of flights diverted is significantly high, alternative airports could not accommodate all the aircrafts) and, in summary, they lead to a scenario where pilots and air traffic controllers have to tackle with an overwhelming workload situation. Under these circumstances, air traffic controllers are expected to provide the most appropriate advice to pilots in order to avoid flight areas of adverse convective conditions, or to set an appropriate safe separation to accommodate arriving and departing traffics at airports to prevent them from unnecessary conflicts. Consequently, the International Civil Aviation Organization (ICAO) has determined that it is necessary to inform about their existence, when they are observed, in the airport local weather reports, and to forecast their possible formation in and around airports (ICAO, 2016). Nevertheless, the forecasting of these convective situations is still a difficult and challenging task (Sánchez et al., 2009).

Towering Cumulus clouds are strongly sprouting Cumulus with generally sharp outlines and often great vertical extent (WMO, 2017). They may produce precipitation as well as turbulence, wind shear and icing. Towering Cumulus often develops into Cumulonimbus, which are heavy and dense clouds, with a considerable vertical extent, in the form of a mountain or huge towers (WMO, 2017). Note that Cumulonimbus

* Corresponding author at: Department of Signal Processing and Communications, Universidad de Alcalá, Alcalá de Henares 28871, Madrid, Spain.
E-mail address: sancho.salcedo@uah.es (S. Salcedo-Sanz).

https://doi.org/10.1016/j.atmosres.2019.104798
Received 12 July 2019; Received in revised form 26 November 2019; Accepted 27 November 2019
Available online 03 December 2019

4.2 Time series forecasting

This chapters specifies some contributions of the Thesis focused on the prediction of time series from a more traditional point of view (nominal classification or regression), but incorporating advanced methods for this. Diverse approaches have been considered in this Thesis, including three different perspectives: 1) nominal classification based on a multi-objective paradigm, 2) standard regression using advanced artificial neural networks (ANNs) models, and 3) multi-task models for regression.

Main publications associated to this section:

- **D. Guijo-Rubio**, P.A. Gutiérrez, C. Casanova-Mateo, J.C. Fernández, A.M. Gómez-Orellana, P. Salvador-González, S. Salcedo-Sanz and C. Hervás-Martínez. "Prediction of convective clouds formation using evolutionary neural computation techniques", Neural Computing and Applications, Vol. 32, 2020, pp. $13917 - 13929$.
JCR (2019): 4.774 Position: 23/136 (Q1).
DOI: 10.1007/s00521-020-04795-w
- **D. Guijo-Rubio**, A.M. Durán-Rosal, P.A. Gutiérrez, A.M. Gómez-Orellana, C. Casanova-Mateo, J. Sanz-Justo, S. Salcedo-Sanz y C. Hervás-Martínez. "Evolutionary artificial neural networks for accurate solar radiation prediction", Energy, Vol. 210, 2020, pp. $1 - 11$.
JCR (2019): 6.082 Position: 3/61 (Q1).
DOI: 10.1016/j.energy.2020.118374
- **D. Guijo-Rubio**, A.M. Gómez-Orellana, P.A. Gutiérrez y C. Hervás-Martínez. "Short- and long-term energy flux prediction using Multi-Task Evolutionary Artificial Neural Networks", Ocean Engineering, Vol. 216, 2020, pp. 108089.
JCR (2019): 3.068 Position: 1/14 (Q1).
DOI: 10.1016/j.oceaneng.2020.108089

4.2.1 Prediction of convective clouds formation using multi-objective evolutionary neural computation techniques

In this paper, we also tackle the prediction of convective clouds formation, but rather than predicting it from the ordinal classification (OC) point of view, we approach it by means of evolutionary algorithms (EAs) applied to ANNs. One of the reasons behind the use of ANNs is that they have been successfully used for previous atmospheric events prediction problems. As we stated before, the dataset is highly imbalanced given that the number of CLEAR days is much higher than the number of days of the remaining classes.

Therefore, the use of mono-objective evolutionary artificial neural networks (EANNs) will aim to classify most of the patterns in the majority class (in this case the CLEAR class), achieving, in this case, a high performance in terms of accuracy, which represents the global performance.

On the other hand, it is worthy of mention that, when dealing with highly imbalanced datasets, accuracy is not the best performance measure, at least when the models are only optimised by using it. In this sense, if this measure is combined with any other metric aiming to achieve a good performance for the minority classes, the model will improve significantly its quality. One of the metrics following the idea of reaching a notable performance for minority classes is the minimum sensitivity (MS). Note that accuracy and MS are opposite, given that if one objective increases, it leads to a decrease in the other one, i.e. an increase in the global accuracy does not necessarily mean that the accuracy of the minority classes will increase (in general, the majority class will increase its accuracy), and, otherwise, an increase in the MS could not correspond to an increase in the global accuracy (in general, the majority class will decrease its accuracy). Therefore, in this work we propose the use of a multi-objective methodology, aiming to get the highest MS without disregarding the global performance. The proposed methodology is a multi-objective evolutionary artificial neural network (MOEANN).

Regarding the structure of the ANNs, we propose the use of the most common basis functions in the literature: sigmoidal units (SUs), product units (PUs) and radial basis functions (RBFs). Apart from these basis functions, we also considered mixtures of them, in such a way that projection functions (SUs and PUs) are combined with the kernel function RBF, giving rise to SURBF and PURBF, respectively. Furthermore, in order to robustly check the performance of the MOEANN, we also run two types of mono-objective EANNs depending on the objective function (one optimises accuracy whereas the second one optimises the MS).

The results achieved demonstrate that, if we only consider the global accuracy, using a mono-objective EANNs optimised by this metric will lead to the best results (disregarding the minority classes). However, given that we are more interested in anticipating the convective clouds formation, we should aim to correctly predict the minority classes (without losing too much global performance). Therefore, MOEANNs are able to improve the performance on these opposite measures (accuracy and MS), achieving balanced results. Finally, these results are compared against standard terminal aerodrome forecasts (TAFs), obtaining improvements in the classification of the most difficult convective situation prediction.

Neural Computing and Applications (2020) 32:13917–13929
https://doi.org/10.1007/s00521-020-04795-w

ORIGINAL ARTICLE

Prediction of convective clouds formation using evolutionary neural computation techniques

David Guijo-Rubio[1] · Pedro A. Gutiérrez[1] · Carlos Casanova-Mateo[2] · Juan Carlos Fernández[1] · Antonio Manuel Gómez-Orellana[1] · Pablo Salvador-González[3] · Sancho Salcedo-Sanz[4] · César Hervás-Martínez[1]

Received: 6 July 2019 / Accepted: 17 February 2020 / Published online: 10 March 2020

Abstract

The prediction of convective clouds formation is a very important problem in different areas such as agriculture, natural hazards prevention or transport-related facilities. In this paper, we evaluate the capacity of different types of evolutionary artificial neural networks to predict the formation of convective clouds, tackling the problem as a classification task. We use data from Madrid-Barajas airport, including variables and indices derived from the Madrid-Barajas airport radiosonde station. As objective variable, we use the cloud information contained in the METAR and SPECI meteorological reports from the same airport and we consider a prediction time horizon of 12 h. The performance of different types of evolutionary artificial neural networks has been discussed and analysed, including three types of basis functions (sigmoidal unit, product unit and radial basis function) and two types of models, a mono-objective evolutionary algorithm with two objective functions and a multi-objective evolutionary algorithm optimised by the two objective functions simultaneously. We show that some of the developed neuro-evolutionary models obtain high quality solutions to this problem, due to its high unbalance characteristic.

Keywords Convection initialisation prediction · Machine learning algorithms · Neural networks · Unbalanced databases

1 Introduction

Convective weather conditions can significantly affect many strategic economical areas such as electric supply, communications, logistic services and transport, especially air traffic. The identification of atmospheric situations that favours the initiation of convection as well as the accurate prediction of its timing and location is still a difficult task for the operational weather forecasters [1]. The basic ingredients for convection to occur stated by Johns and Doswell [2] (a sufficiently moist and deep layer in the low or mid-atmosphere, conditional instability and a triggering mechanism) are still the base of many tools and applications developed to support the forecast of convection initiation. However, several works discussing the results obtained in many field studies have been published in the last years [3–6], which shows that there is still room for improving our knowledge about convective initiation.

As it is well known throughout the weather forecasters community, one of the most widely used tools for predicting the occurrence of convection is the stability indices derived, in most cases, from the temperature and humidity data measured by the sounding (upper-air) stations. These indices, such as the lifted index [7], the total of totals index [8] or the Convective Available Potential Energy (CAPE) Index [9], try to characterise, with varying degrees of success, if the current atmospheric conditions favour convection and, eventually, the formation of cumulus congestus and cumulonimbus clouds that can bring adverse

✉ David Guijo-Rubio
dguijo@uco.es

1 Department of Computer Science and Numerical Analysis, University of Córdoba, Rabanales Campus, Albert Einstein Building 3rd Floor, 14071 Córdoba, Spain

2 Department of Civil Engineering: Construction, Infrastructure and Transport, Universidad Politécnica de Madrid, Madrid, Spain

3 LATUV, Remote Sensing Laboratory, Universidad de Valladolid, Valladolid, Spain

4 Department of Signal Processing and Communications, Universidad de Alcalá, Alcalá de Henares, Spain

4.2.2 Evolutionary artificial neural networks for accurate solar radiation prediction

In this paper, we carry out the prediction of solar radiation. In the last decades, the demand of electricity has been increasing every year, and the availability of fossil fuels, such as petroleum or coal, is getting compromised progressively. Apart from that, we should bear in mind that these sources of energy have a significant impact on climate change. Therefore, the development of renewable energy sources not only does it provide an enormous benefit for the environment, but also is the unique way to cover the increase of energy demand, projected to be around a 60% from 2002 to 2030.

Among all the renewable sources of energy, there are two that highlight from the rest: the tidal or waves energy and the solar energy. More concretely, the second one is thought as the future renewable energy source, not only given the extraordinary solar resource we have available, but also due to the fact that it is a clean energy without risks for the environment. The main aim of the paper is the prediction of solar radiation in order to enable optimal management of energy systems.

In recent years, the use of machine learning (ML) techniques has increased in this field given its ability and accuracy to obtain reliable forecasts. Therefore, in this paper, we focus on the use of satellite data and EANNs, which have been proved to achieve excellent results in other renewable energy prediction works. Regarding the architecture of the ANNs, apart from the typical linear output, we propose the use of an ANN with a PU in the output layer. Hence, three different ANNs are compared: SU-LO (SU in the hidden layer with a linear output), RBF-LO (RBF in the hidden layer with a linear output) and SU-PU (SU in the hidden layer with PU in the output layer). Note that the combination of RBF in the hidden layer with a PU output layer is not adequate, given that RBFs are local functions and assessing their interaction would make no sense.

Regarding the data used in the study, the estimation of solar radiation is performed at the radiometric station of Toledo (Spain) with a 1h time-horizon. Furthermore, for the predictive variables, we have integrated different sources of satellite data: 1) from the Meteosat satellites, we have obtained the reflectivity in the visible channels (0.6 and 0.8 μm) and the clear sky radiance. 2) From the Copernicus atmosphere monitoring service (CAMS), the cloud index as well as the CAMS solar radiation were obtained. And 3) solar radiation data from the *SolarGIS* model has also been included. Once all these predictive variables are obtained, different configurations for the dataset are built, in order to analyse the behaviour of EANNs under different sets of input data. Therefore, four scenarios are considered including a variable number of inputs, from 5 to 40, given that satellite data (i.e. reflectivity and clear sky radiance variables) could be included in several ways: from including just the central pixel value to including all the pixels values, in this case, 9.

The results achieved by the different EANNs are then compared against some of the main state-of-the-art ML algorithms (extreme learning machines (ELMs), support vector regressors (SVRs) or multilayer perceptrons (MLPs), among others), outperforming all of them with a large difference of accuracy. Finally, a statistical analysis is carried out comparing the results obtained by the EANNs and the state-of-the-art ML techniques, concluding that SU-PU shows significant differences with respect to the rest of techniques, except with RBF-LO and ELM, for which SU-PU shows an improvement, although it is not significant.

Energy 210 (2020) 118374

Contents lists available at ScienceDirect

Energy

journal homepage: www.elsevier.com/locate/energy

Evolutionary artificial neural networks for accurate solar radiation prediction

D. Guijo-Rubio [a, *, 1], A.M. Durán-Rosal [b, 1], P.A. Gutiérrez [a], A.M. Gómez-Orellana [a], C. Casanova-Mateo [c, d], J. Sanz-Justo [c], S. Salcedo-Sanz [e], C. Hervás-Martínez [a]

[a] Department of Computer Science and Numerical Analysis, University of Córdoba, Córdoba, Spain
[b] Department of Quantitative Methods, Universidad Loyola Andalucía, Córdoba, Spain
[c] LATUV: Remote Sensing Laboratory, University of Valladolid, Valladolid, Spain
[d] Department of Civil Engineering: Construction, Infrastructure and Transport, Polytechnic University of Madrid, Madrid, Spain
[e] Department of Signal Processing and Communications, University of Alcalá, Alcalá de Henares, Spain

ARTICLE INFO

Article history:
Received 26 March 2020
Received in revised form
9 July 2020
Accepted 11 July 2020
Available online 6 August 2020

Keywords:
Solar radiation estimation
Evolutionary artificial neural networks
Satellite data
Physical models

ABSTRACT

This paper evaluates the performance of different evolutionary neural network models in a problem of solar radiation prediction at Toledo, Spain. The prediction problem has been tackled exclusively from satellite-based measurements and variables, which avoids the use of data from ground stations or atmospheric soundings. Specifically, three types of neural computation approaches are considered: neural networks with sigmoid-based neurons, radial basis function units and product units. In all cases these neural computation algorithms are trained by means of evolutionary algorithms, leading to robust and accurate models for solar radiation prediction. The results obtained in the solar radiation estimation at the radiometric station of Toledo show an excellent performance of evolutionary neural networks tested. The structure sigmoid unit-product unit with evolutionary training has been shown as the best model among all tested in this paper, able to obtain an extremely accurate prediction of the solar radiation from satellite images data, and outperforming all other evolutionary neural networks tested, and alternative Machine Learning approaches such as Support Vector Regressors or Extreme Learning Machines.

1. Introduction

Today's world is completely dependent on electricity, mainly produced based on fossil fuels such as petroleum, natural gas or coal. Moreover, according to Ref. [1], the world primary energy demand is projected to expand by almost 60% from 2002 to 2030, with an average increase of 1.7% per year. However, the high environmental impact of current energy resources, together with the need for addressing the impact of climate change, led to an important development of renewable energy sources [2]. The International Energy Agency (IEA) has stated that electricity generation from renewable energy is expected to rise up to 39% by 2050 [3]. In fact, renewable sources have experienced a huge growth in the last two decades, and they are today thought as the resources which will erase fossil fuels from our society in the next fifty years. The prevalence of renewable resources has several challenges, that must be solved before renewable energies overtake fossils as primal energy resources. The most important issue with renewable energy sources is to manage their inherent intermittence, which currently avoid that renewable energies overpass 40% of penetration in the energetic mix. Among renewable resources, wind and solar energies are the two most developed ones, but solar resource is currently thought as the future renewable source, due to the extraordinary solar resource we have available. More specifically, solar energy is a clean, extremely abundant and sustainable source of energy [4], that poses a low risk to the environment. With this background in mind, it clearly follows that new techniques and methodologies need to be developed in order to accurately estimate solar energy availability at any part of the world, as well as to improve its prediction to enable optimal management of energy systems.

The basis to estimate solar radiation at any given location is to apply the classical astronomical equations [5]. In addition, the well-

* Corresponding author. Department of Computer Science and Numerical Analysis, University of Cordoba, Rabanales Campus, Albert Einstein Building, 3rd Floor, 14014, Córdoba, Spain.
E-mail address: dguijo@uco.es (D. Guijo-Rubio).
[1] D. Guijo-Rubio and A. M. Durán-Rosal contributed at the same level to this work.

https://doi.org/10.1016/j.energy.2020.118374

4.2.3 Short- and long-term energy flux prediction using Multi-Task Evolutionary Artificial Neural Networks

In this paper, a novel approach to tackle simultaneously short- and long-term energy flux prediction is presented. As stated before, in the last years, the vast use of fossil fuels has led to a rise in the global mean temperatures, surface temperatures increasing about $0.2°$C per decade. In this sense, the development of renewable and eco-friendly sources of energy is of urgent necessity. On the other hand, as aforementioned, solar and tidal and waves energies are the most popular sources of renewable energy. However, tidal and waves energies have an import advantage: being constantly available, they do not depend on the intermittent nature of sun.

The main aim of this paper is to predict the wave energy flux in order to develop an stable source of energy, given that waves could exhibit a stochastic nature produced by different environmental elements. Moreover, anticipating to excesses and shortcomings has a significant impact not only on the economic side, but also on the people's daily life. Therefore, we have carried out the prediction at short and long time horizons, from 6h to 48h. For this, we propose the use of EANNs, which have been previously applied to this field of science, achieving an excellent performance in comparison against standard statistical, meteorological and physical models. The main novelty of this work is to tackle simultaneously short- and long-term energy flux prediction, considering multi-task evolutionary artificial neural networks (MTEANNs).

With respect to the data used in this study, two different sources of information have been integrated: 1) data from the national data buoy center (NDBC), which has been collected by buoys located in the oceans and seas, including meteorological and oceanographic observations, such as significant wave height or sea level pressure, among others; and 2) data from the national center for atmospheric research (NCAR), which provides reanalysis data of meteorological variables. On the other hand, the predictive variable is not directly measured by the buoys sensors, but it is derived from two wave parameters, which are in fact measured by the buoys. More concretely, in this work, we use data from three different buoys located at Gulf of Alaska.

Regarding the EANNs, the three most popular basis functions (SUs, PUs and RBFs) are used in the hidden layer. On the other hand, for the output layer, both linear and PUs are proposed. The results obtained demonstrate that MTEANNs are the best in terms of both mean squared error (MSE) and standard error of prediction (SEP). Moreover, the number of connections of the models involves a notable simplicity. Specifically, SU-LI, i.e. MTEANNs with SUs in the hidden layer and linear output, is the model achieving the best performance for all the time prediction horizons. Apart from the comparison between the different proposals of MTEANNs, another comparison against the main state-of-the-art ML techniques in regression, including SVRs and ELMs, is carried out, concluding that MTEANNs are an excellent approach for the prediction of energy flux at both short- and long-term time prediction horizons.

Ocean Engineering 216 (2020) 108089

Contents lists available at ScienceDirect

Ocean Engineering

journal homepage: www.elsevier.com/locate/oceaneng

Short- and long-term energy flux prediction using Multi-Task Evolutionary Artificial Neural Networks

David Guijo-Rubio[1], Antonio M. Gómez-Orellana[1,*], Pedro A. Gutiérrez, César Hervás-Martínez

Department of Computer Science and Numerical Analysis, University of Córdoba, Córdoba, Spain

ARTICLE INFO

Keywords:
Wave energy flux prediction
Marine energy
Multi-task machine learning
Evolutionary artificial neural networks
Reanalysis data

ABSTRACT

This paper presents a novel approach to tackle simultaneously short- and long-term energy flux prediction (specifically, at $6h$, $12h$, $24h$ and $48h$ time horizons). The methodology proposed is based on the Multi-Task Learning paradigm in order to solve the four problems with a single model. We consider Multi-Task Evolutionary Artificial Neural Networks (MTEANN) with four outputs, one for each time prediction horizon. For this purpose, three buoys located at the Gulf of Alaska are considered. Measurements collected by these buoys are used to obtain the target values of energy flux, whereas, only reanalysis data are used as input values, allowing the applicability to other locations. The performance of three different basis functions (Sigmoidal Unit, Radial Basis Function and Product Unit) are compared against some popular state-of-the-art approaches such as Extreme Learning Machines and Support Vector Regressors. The results show that MTEANN methodology using Sigmoidal Units in the hidden layer and a linear output achieves the best performance. In this way, the multi-task methodology is an excellent and lower-complexity approach for energy flux prediction at both short- and long-term prediction time horizons. Furthermore, the results also confirm that reanalysis data is enough for describing well the problem tackled.

1. Introduction

In the last decades, there has been an increase in the carbon dioxide levels, and in consequence, the global mean temperatures have also risen. According to the special report regarding global warming done by the Intergovernmental Panel on Climate Change (IPCC) (Hoegh-Guldberg et al., 2018) the surface temperatures are increasing by about 0.2°C per decade. The ongoing rise of the average temperature of the Earth is leading to wildfires, the expansion of deserts or more intense storms, among others (Council et al., 2012). Therefore, the interest in renewable and eco-friendly energy sources such as solar, wind, tides and waves is in continuous growth (Ellabban et al., 2014). Although solar and wind are considered the most popular alternative sources of energy, they have the setback of not being constantly available, due to the intermittent nature of wind and sun. On the other hand, tides and waves not only benefit from being always in movement, but they also can be used in many parts of the world, including oceans or big seas. Furthermore, given that the energy storage systems are a major challenge (Palmer and Floyd, 2020) and that around 40% of the world's population live within 100 km of the coast, the wave and tidal renewable energies allow for a cost-effective transmission of electricity (Esteban and Leary, 2012).

Waves exhibit a stochastic nature, due to the influence of a great number of environmental elements. Therefore, they cannot be predicted straightforwardly as tides. Due to this stochastic behaviour, the reliability and confidence of wave energy generation need to be predicted beforehand in order to develop an stable source of energy. For the generation of electricity from wave energy, Wave Energy Converters (WECs) (Falcão, 2010; Aderinto and Li, 2018) are applied in order to transform the kinetic energy directly generated by the waves into electricity (Falnes and Kurniawan, 2020). To model the behaviour of wave energy, the two most important parameters are the significant wave height (H_s) and the wave energy flux (F_e), which have been widely studied in the literature from different perspectives (Nitsure et al., 2012): physical models (Ibarra-Berastegi et al., 2015), statistical models (Lin et al., 2020) or by the application of Machine Learning (ML) techniques (Cuadra et al., 2016), among others.

Focusing on ML approaches, some of the first works published were those of Deo and Naidu (1998) and Deo et al. (2001), which consisted in real-time forecasting of H_s by using Artificial Neural Network

* Corresponding author.
E-mail addresses: dguijo@uco.es (D. Guijo-Rubio), am.gomez@uco.es (A.M. Gómez-Orellana), pagutierrez@uco.es (P.A. Gutiérrez), chervas@uco.es (C. Hervás-Martínez).
[1] Contributed equally to this work.

https://doi.org/10.1016/j.oceaneng.2020.108089
Received 5 June 2020; Received in revised form 24 July 2020; Accepted 9 September 2020
Available online 16 September 2020

Computers are able to see, hear and learn. Welcome to the future.

Dave Waters

5 Time series classification

This chapter presents the works related with time series classification (TSC). More concretely, TSC is divided into two main blocks according to the nature of the time series dataset. When the target variable takes categorical values without an order, it is known as standard TSC. On the other hand, if the class attribute values (being also categorical values) follow a natural order relationship among them, the problem is tackled applying the ordinal paradigm, and we propose to name this field time series ordinal classification (TSOC).

5.1 Time series classification

Time series classification (TSC) is the most popular field regarding time series data mining, given the interest raised in the literature. In the last two decades, a huge number of proposals have been presented to the literature, trying to outperform the approaches in the state-of-the-art. In this sense, the main open question to be solved regarding TSC is: *Is there any approach reaching the state-of-the-art performance with the lowest time complexity?* Two different paths could be differentiated: 1) trying to improve the performance of the hierarchical vote system collective of transformation-based ensemble (HIVE-COTE) technique, which is one of the methods achieving the best results in terms of correct classification rate (CCR), and 2) achieving a non-significant lower performance than the

HIVE-COTE but significantly decreasing the computational time. Bearing in mind the outstanding performance of HIVE-COTE and considering the first path, one of the main steps is to improve the individual classifiers belonging to this ensemble technique. Therefore, a hybrid approach with shapelets has been presented.

Main publications associated to this section:

- **D. Guijo-Rubio**, P.A. Gutiérrez, R. Tavenard y A. Bagnall. "A hybrid approach to time series classification with shapelets". 20th International Conference on Intelligent Data Engineering and Automated Learning (IDEAL 2019). 2019. LNCS, Vol. 11871, pp. 137 – 144.
  DOI: https://doi.org/10.1007/978-3-030-33607-3_16

5.1.1 A hybrid approach to time series classification with shapelets

In this paper, we present a novel approach to time series classification. As stated above, the HIVE-COTE is the state-of-the-art in performance for TSC. It consists in an ensemble encapsulating classifiers built on several data representations. One of the most popular representations is the shapelet transform (ST), which consists in the transformation of the original time series dataset by means of phase independent patterns or subseries extracted from the original time series, known as shapelets.

The traditional extraction procedure of shapelets is divided in four different steps: 1) search, in which random samples of shapelets are extracted from the training set and the best are kept; 2) transform, in which a completely new dataset is created, and the attributes are obtained from computing the distances from the shapelets to the original time series; 3) fit model, in which a classifier is trained over the transformed train dataset; and 4) predict, which consists in estimating the class values on the test set.

Furthermore, given that shapelets have been a popular research topic, several approaches have been presented to the literature. One of these proposals was based on searching for shapelets over the whole space of all possible shapelets, in such a way that shapelets were no longer subsequences existing in the training set. This method is known as learned shapelets (LS) [76].

Following this idea, the proposal presented in this paper is a hybrid approach between original ST and the LS, implementing the LS model as a variant of a convolutional neural network (CNN), taking time series as inputs and class probabilities as outputs. This model is composed of two layers: 1) the first extracts a ST-like representation, known as shapelet layer, and 2) a logistic regression (LR) layer. The shapelet layer contains several shapelet blocks (grouping shapelets by their length), which in turn are made of two stages: 1) a feature extraction step, computing the pairwise distances between the shapelet and the original time series, and 2) a pooling step, retaining the minimum of all these distances. Besides, the optimisation procedure consists in tuning both the shapelet values and the LR weights.

Finally, the proposal consisted in applying CNNs, but rather than using the standard LR classifier, it used the rotation forest (RF) classifier, which was recently proved to be outstanding when dealing with only numerical attributes [15]. This method has been evaluated on 92 datasets from the University of East Anglia and University of California Riverside (UEA/UCR) TSC repository. The hybrid approach is then compared against the standard ST, using the RF as final classifier, and against the hybrid approach, using the standard LR. In conclusion, tuning the shapelets significantly improved accuracy achieving non-significantly worse results than the ST module from HIVE-COTE, which enumerates all possible shapelets, but decreasing enormously the computational time.

A Hybrid Approach to Time Series Classification with Shapelets

David Guijo-Rubio[1,2], Pedro A. Gutiérrez[1], Romain Tavenard[3], and Anthony Bagnall[2]

[1] Department of Computer Sciences, Universidad de Córdoba, Córdoba, Spain
dguijo@uco.es
[2] University of East Anglia, Norwich NR47TJ, UK
[3] Univ. Rennes, CNRS, LETG/IRISA, Rennes, France

Abstract. Shapelets are phase independent subseries that can be used to discriminate between time series. Shapelets have proved to be very effective primitives for time series classification. The two most prominent shapelet based classification algorithms are the shapelet transform (ST) and learned shapelets (LS). One significant difference between these approaches is that ST is data driven, whereas LS searches the entire shapelet space through stochastic gradient descent. The weakness of the former is that full enumeration of possible shapelets is very time consuming. The problem with the latter is that it is very dependent on the initialisation of the shapelets. We propose hybridising the two approaches through a pipeline that includes a time constrained data driven shapelet search which is then passed to a neural network architecture of learned shapelets for tuning. The tuned shapelets are extracted and formed into a transform, which is then classified with a rotation forest. We show that this hybrid approach is significantly better than either approach in isolation, and that the resulting classifier is not significantly worse than a full shapelet search.

Keywords: Time series classification · Shapelets · Convolutional neural networks

1 Introduction

Shapelets [1] are discriminatory phase independent subsequences that form a basic primitive in many time series algorithms. For classification, shapelets are assessed using their distance to train set time series and the usefulness of these distances in discriminating between classes. Shapelet based features define a distinct form of discrimination which can be characterised as quantifying whether a particular shape exists in a series or not (at any location). Shapelets have proved an effective tool for classification [2] and have been a popular research topic. One key distinction between research threads is whether shapelets are extracted from the training data or whether the space of all possible shapelets is searched. The

H. Yin et al. (Eds.): IDEAL 2019, LNCS 11871, pp. 137–144, 2019.
https://doi.org/10.1007/978-3-030-33607-3_16

5.2 Time series ordinal classification

A special and unexplored field of time series classification (TSC) is time series ordinal classification (TSOC). It consists in the classification of time series in which the class values follow a natural order relationship. Up-to-the-knowledge of the authors, this field has not been tackled previously. Therefore, focusing on the shapelet transform (ST), we firstly proposed an adaptation of this methodology to the ordinal paradigm, and, secondly, we performed a comparison against the state-of-the-art techniques in TSC for those ordinal datasets identified from the University of East Anglia and University of California Riverside (UEA/UCR) TSC repository.

Main publications associated to this section:

- **D. Guijo-Rubio**, P.A. Gutiérrez, A. Bagnall y C. Hervás-Martínez. "Time series ordinal classification via shapelets". 2020 IEEE International Joint Conference on Neural Networks (IJCNN 2020). 2020. Glasgow, UK. pp. $1-8$.
  DOI: 10.1109/IJCNN48605.2020.9207200
- **D. Guijo-Rubio**, P.A. Gutiérrez, A. Bagnall y C. Hervás-Martínez. "Ordinal versus nominal time series classification". 5th Workshop on Advances Analytics and Learning on Temporal Data (AALTD 2020). 2020. LNAI, Vol. 12588, pp. $19-29$.
  DOI: 10.1007/978-3-030-65742-0_2

Other publications associated to this section:

- **D. Guijo-Rubio**, V. M. Vargas, P. A. Gutiérrez, and C. Hervás-Martínez. "Studying the effect of different Lp norms in the context of Time Series Ordinal Classification". Conference of the Spanish Association for Artificial Intelligence (CAEPIA 2021). 2021. Accepted.

5.2.1 Time series ordinal classification via shapelets

This is the first paper considering TSOC in a general sense (i.e. not focused to any specific application). The proposal is focused on the ST algorithm, which has been adapted to extract the maximum quantity of ordinal information from the time series dataset. Focusing on the ST algorithm, it is made of the three following main steps: 1) candidate generation, in which a subsequence from the training set satisfying all the constraint is extracted; 2) similarity measurement between the candidate and the original time series; and 3) quality measurement of the candidate. After that, once the best k candidates are extracted, a new representation of the dataset is built, in which the attributes are the distances between these k best candidates and the original time series.

In this sense, in order to incorporate the ordinal nature of the labels into the ST extraction stage, several shapelet quality measures are considered instead of the standard information gain (IG), which is the quality measure originally used by the ST algorithm. The first shapelet quality measure proposed is the Fisher score, typically used for feature selection. A reformulation of the original Fisher score is defined adapting it to ordinal classification by including higher costs for distant classes. The second approach is a modified version of the Pearson's determination coefficient, that computes the correlation between the distances obtained from the shapelet to the original time series and the difference of their class indices. And finally, the last approach is a modified version of the Spearman's determination coefficient, which computes the correlation in the same way as Pearson's but considers that any of the variables could be either categorical or continuous.

Moreover, rather than applying the rotation forest (RF) as final classifier, we introduce the use of ordinal classifiers such as proportional odds model (POM) or support vector for ordinal regression with implicit constraints (SVORIM), given that these methods take advantage of the nature order between the labels. Apart from proposing a novel technique for TSOC, a first selection of ordinal time series datasets is carried out, identifying 7 ordinal datasets from the whole TSC repository of the UEA/UCR.

The results show that, in terms of correct classification rate (CCR), the best performance is achieved by the ST using the Pearson's determination coefficient. Moreover, in terms of average mean absolute error (AMAE) (which measures the ordinal classification errors made for every class), the best results are also achieved by ST along with the Pearson's determination coefficient. On the other hand, regarding the classifiers, the average CCR ranks demonstrate that the use of a nominal classifier achieves the best results, which is natural as CCR does not consider the order between the labels. Besides, the average AMAE ranks show that SVORIM is the best. Finally, it can be concluded that the use of the Pearson's determination coefficient as shapelet quality measure leads to the best results, the differences being statistically significant in terms of AMAE.

Time series ordinal classification via shapelets

David Guijo-Rubio
Dept. of Computer Science and Numerical Analysis
University of Córdoba
Córdoba, Spain.
dguijo@uco.es
0000-0002-8035-4057

Pedro A. Gutiérrez
Dept. of Computer Science and Numerical Analysis
University of Córdoba
Córdoba, Spain.
pagutierrez@uco.es
0000-0002-2657-776X

Anthony Bagnall
School of Computing Sciences
University of East Anglia
Norwich Research Park
Norwich, United Kingdom.
ajb@uea.ac.uk
0000-0003-2360-8994

César Hervás-Martínez
Dept. of Computer Science and Numerical Analysis
University of Córdoba
Córdoba, Spain.
chervas@uco.es
0000-0003-4564-1816

***Abstract*—Nominal time series classification has been widely developed over the last years. However, to the best of our knowledge, ordinal classification of time series is an unexplored field, and this paper proposes a first approach in the context of the shapelet transform (ST). For those time series dataset where there is a natural order between the labels and the number of classes is higher than 2, nominal classifiers are not capable of achieving the best results, because the models impose the same cost of misclassification to all the errors, regardless the difference between the predicted and the ground-truth. In this sense, we consider four different evaluation metrics to do so, three of them of an ordinal nature. The first one is the widely known Information Gain (IG), proved to be very competitive for ST methods, whereas the remaining three measures try to boost the order information by refining the quality measure. These three measures are a reformulation of the Fisher score, the Spearman's correlation coefficient (ρ), and finally, the Pearson's correlation coefficient (R^2). An empirical evaluation is carried out, considering 7 ordinal datasets from the UEA & UCR time series classification repository, 4 classifiers (2 of them of nominal nature, whereas the other 2 are of ordinal nature) and 2 performance measures (correct classification rate, CCR, and average mean absolute error, $AMAE$). The results show that, for both performance metrics, the ST quality metric based on R^2 is able to obtain the best results, specially for $AMAE$, for which the differences are statistically significant in favour of R^2.**

***Index Terms*—Time Series, Ordinal Classification, Ordinal regression, Shapelet Quality Measures**

I. INTRODUCTION

Time series are a widely used sort of temporal data in which objects are collected over time. In the last years, time series have been a hot topic in machine learning and data mining, and can be found in a vast number of fields such as: fog prediction [1], stock indices [2] or forged-alcohol detection [3]. Time series classification is a task in which a label is given to a set of chronologically ordered points. We focus on a specific case, those problems in which there are three or more possible categories and they follow an order relationship.

This kind of classification is known as ordinal classification or ordinal regression, being a field of machine learning tackling problems in which the target variables are discrete and present a natural order between their labels [4]. An example is the prediction of the stage of a disease state, in which a patient could be labelled as $none$, $mild$, $moderate$, $severe$ or $extreme$. Obviously, misclassifying a $mild$ patient as $severe$, should be far more penalised than misclassifying that patient as $none$ or $moderate$. This problem can be tackled in several ways: 1) as a nominal classification problem, which ignores the natural order between the labels, 2) as a regression problem, which implies assigning each label a numerical value (which requires assuming a distance between values that can hinder the performance of the regressor), or 3) as an ordinal classification problem, which is the approach we consider. This special kind of classification can be found in several fields, such as meteorological prediction [5], medical research [6], [7] and wave height prediction [8]. The datasets used in these projects include an ordered target variable, and thus, specialised ordinal classifiers are able to achieve higher performances than nominal classifiers or regressors, by constructing more accurate models.

Traditionally, nominal time series have been classified using a similarity measure in conjunction with a standard classifier, such as k-Nearest Neighbours [9]. This similarity can be assessed from several points of view: by considering time, change or shape. We focus on shape based similarity, in which time series are compared by using phase independent sub-sequences generally much shorter than the original time series. These sub-sequences, known as shapelets,were first proposed as a time series primitive by Ye and Keogh [10]. The original proposal embedded the shapelet extraction into a decision tree that used Information Gain (IG) to assess the candidates. Moreover, this time series primitive has been used in some other ways in the literature: Hills *et al.* [11] proposed the Shapelet Transformation (ST), in which the k best shapelets are used to convert the original time series dataset into a new transformed dataset. In this new representation, the attributes are the distances between the shapelets and the time series being evaluated. The reason for this is that the transformation allows the application of any classifiers and avoids the sequential search for shapelets at each node of the tree. Grabocka *et al.* [12] proposed a new perspective in which shapelets are learned. This method enables the learning of shapelets without the need of searching for a vast number of candidates.

978-1-7281-6926-2/20/$31.00

5.2.2 Ordinal versus nominal time series classification

In this paper, we continue the research line established by the previous paper. The main aim of this paper is to put in value TSOC and to establish a robust comparison against the main state-of-the-art approach in TSC. Nevertheless, in this work, the use of the ordinal classifier POM is removed given that it is a linear method and its implementation does not consider a regularisation term, which seems to be necessary due to the numerous shapelets extracted. In the same way than the previous paper, we have concluded that the Pearson's determination coefficient along with the use of the ordinal classifier SVORIM achieves the best rank, not only for CCR, but also in terms of AMAE.

The main novelty of this paper is the comparison against the three main state-of-the-art techniques in TSC, which are: 1) hierarchical vote system collective of transformation-based ensemble (HIVE-COTE), a meta-ensemble composed of 5 modules, among which ST is one of them; 2) InceptionTime, which is an ensemble of convolutional neural network (CNN) models in which several filters of different lengths are applied at the same time to the input time series; and finally, 3) time series combination of heterogeneous and integrated embedding forest (TS-CHIEF), which is another ensemble classifier integrating the most effective embeddings of time series. Note that all these three approaches are based on the idea of ensembles, which are commonly known to reach outstanding performances. However, they suffer from the computational load associated to them, in such a way that they are computationally intensive.

The results of the comparison demonstrate that SVORIM achieves the best results or the second best for most of the datasets, obtaining the best average ranking. Moreover, it is worth mentioning that all the classifiers (ordinal or nominal), when applied to the ST combined with the Pearson's determination coefficient as shapelet quality measure, achieve a better average rank than the state-of-the-art TSC techniques. This is due to these methods are taking advantage of the ordinal information induced by ST using the Pearson's determination coefficient as shapelet quality measure.

Ordinal Versus Nominal Time Series Classification

David Guijo-Rubio[1]( ), Pedro Antonio Gutiérrez[1], Anthony Bagnall[2], and César Hervás-Martínez[1]

[1] Department of Computer Sciences, Universidad de Córdoba, Córdoba, Spain
{dguijo,pagutierrez,chervas}@uco.es

[2] School of Computing Sciences, University of East Anglia, Norwich Research Park, Norwich, UK
ajb@uea.ac.uk

Abstract. Time series ordinal classification is one of the less studied problems in time series data mining. This problem consists in classifying time series with labels that show a natural order between them. In this paper, an approach is proposed based on the Shapelet Transform (ST) specifically adapted to ordinal classification. ST consists of two different steps: 1) the shapelet extraction procedure and its evaluation; and 2) the classifier learning using the transformed dataset. In this way, regarding the first step, 3 ordinal shapelet quality measures are proposed to assess the shapelets extracted, and, for the second step, an ordinal classifier is applied once the transformed dataset has been constructed. An empirical evaluation is carried out, considering 7 ordinal datasets from the UEA & UCR Time Series Classification (TSC) repository. The results show that a support vector ordinal classifier applied to the ST using the Pearson's correlation coefficient (R^2) is the combination achieving the best results in terms of two evaluation metrics: accuracy and average mean absolute error. A final comparison against three of the most popular and competitive nominal TSC techniques is performed, demonstrating that ordinal approaches can achieve higher performances even in terms of accuracy.

Keywords: Time series · Ordinal classification · Ordinal regression · Shapelet quality measures

1 Introduction

Time Series Ordinal Classification (TSOC) refers to a prediction problem where the objective is to classify time series with an ordinal label, i.e. the set of labels includes a natural order relationship. In this context, ordinal classification [12] covers those supervised problems where the target variable is discrete and includes a natural order relationship among the labels. Ordinal classification problems can be found in several fields, such as meteorological prediction [10,11], or medical research [19], among others.

V. Lemaire et al. (Eds.): AALTD 2020, LNAI 12588, pp. 19–29, 2020.
https://doi.org/10.1007/978-3-030-65742-0_2

The value of an idea lies in the using of it.

Thomas A. Edison

6 Additional works

This chapter presents additional works tackled during the development of this Thesis, to satisfy and accomplish several challenges and goals associated with different national projects granted to the research group. In this sense, this chapter is divided into two completely different sections: 1) non-temporal data regression, in which various engineering application problems have been considered, and 2) non-temporal data classification, in which several health-related problems have been solved, both sections using machine learning (ML) techniques.

6.1 Non-temporal data regression: engineering applications

Ultimately, engineering processes are usually designed to collect data from the different operations involved. This data collection is performed aiming to improve operations trying to minimise as much as possible the cost associated to them. Moreover, these operations usually concern several objectives, aiming to optimise all of them simultaneously. In this Section, two real-world problems are solved by applying advanced machine learning (ML) regression techniques. More concretely, multi-task evolutionary artificial neural networks (MTEANNs) have been proposed to solve both of them, according to the excellent ability of this methodology to train a shared structure inferred from the data.

Main publications associated to this section:

- F. Comino, **D. Guijo-Rubio**, M. R. de Adana and C. Hervás-Martínez. "Validation of multitask artificial neural networks to model desiccant wheels activated at low temperature", International Journal of Refrigeration, Vol. 100. 2019, pp. 434 – 442.
JCR (2019): 3.461 Position: 11/61 (Q1).
DOI: 10.1016/j.ijrefrig.2019.02.002

- F.J. Jiménez-Romero, **D. Guijo-Rubio**, F.R. Lara-Raya, A. Ruiz-González y C. Hervás-Martínez. "Validation of artificial neural networks to model the acoustic behaviour of induction motors", Applied Acoustics, Vol. 166, 2020, pp. 107332.
JCR (2019): 2.440 Position: 9/32 (Q2).
DOI: 10.1016/j.apacoust.2020.107332

6.1.1 Validation of multitask artificial neural networks to model desiccant wheels activated at low temperature

In this paper, we present the modelling of desiccant wheels (DWs) activated at low temperature. Food and pharmaceutical industries need to control the internal moisture content, given that an exceed in the indoor air humidity leads to problems of moulds and fungus. Hence, controlling the indoor humidity has a great interest for these industries. Moreover, desiccant wheels provide many advantages over standard systems. Nevertheless, in order to achieve an excellent dehumidification capacity, significant energy consumption is required. In this way, when the DWs are activated at low temperature, the environmental impact is reduced, and, hence, DWs activated at low temperatures are considered for this study.

The main goal of this paper is to apply artificial neural networks (ANNs) to develop an empirical parsimonious model of a DW activated at low temperatures, achieving good accuracy without being computationally intensive. For this, numerous experiments have been performed, covering a wide range of operating conditions of the process and regeneration air flows. Furthermore, the predictive variables are the outlet process air temperature and the outlet process air humidity ratio, being both real-valued variables.

Regarding the architecture of the ANN models, the two most common basis functions (sigmoidal unit (SU) and product unit (PU)) are used in the hidden layer, whereas we only consider linear outputs. Note that another goal of the paper is obtaining simple models, i.e. with a small number of connections, thus, the use of MTEANNs is justified, aiming to obtain simple homogeneous models with high accuracy.

The results obtained demonstrated that ANNs with SUs in the hidden layer are the best method for modelling the DWs, not only in terms of mean squared error (MSE) and standard error of prediction (SEP), but also in terms of simplicity (smallest number of connections). Hence, it can be said that MTEANNs are an effective transfer mechanism due to the fact that they are highly convenient for extracting common features of multiple tasks.

International Journal of Refrigeration 100 (2019) 434–442

Contents lists available at ScienceDirect

International Journal of Refrigeration

journal homepage: www.elsevier.com/locate/ijrefrig

Validation of multitask artificial neural networks to model desiccant wheels activated at low temperature

F. Comino[a,*], D. Guijo-Rubio[b], M. Ruiz de Adana[a], C. Hervás-Martínez[b]

[a] Departamento de Química-Física y Termodinámica Aplicada, Universidad de Córdoba, Córdoba, Spain
[b] Departamento de Informática y Análisis Numérico, Universidad de Córdoba, Córdoba, Spain

ARTICLE INFO

Article history:
Received 27 October 2018
Revised 30 January 2019
Accepted 1 February 2019
Available online 7 February 2019

Keywords:
Artificial neural networks
Sigmoid units
Empirical models for desiccant wheels

ABSTRACT

Desiccant wheels (DW) could be a serious alternative to conventional dehumidification systems based on direct expansion units, which depend on electrical energy. The main objective of this work was to evaluate the use of multitask artificial neural networks (ANNs) as a modelling technique for DWs activated at low temperature with low computational load and good accuracy. Two different ANN models were developed to predict two output variables: outlet process air temperature and humidity ratio. The results show that a sigmoid unit neural network obtained 0.390 and 2.987 for MSE and SEP, respectively. These results outline the effective transfer mechanism of multitask ANNs to extract common features of multiple tasks, being useful for modelling a DW activated at low temperature. On the other hand, moisture removal capacity of the DW and its performance were analysed under several inlet air conditions, showing an increase under process air conditions close to saturation air.

Validation de réseaux de neurones artificiels multitâches pour modéliser des roues déshydratantes activées à basse température

Mots-clés: Réseaux neuronaux artificiels; Unités sigmoïdes; Modèles empiriques de roues déshydratantes

1. Introduction

Dehumidification systems are necessary to maintain the required indoor conditions in buildings with high latent loads. Food and pharmaceutical industries also have a significant interest in controlling the internal moisture content, (De Antonellis et al., 2016; Wang et al., 2018). Excessive indoor air humidity can cause problems related to the indoor air quality of the building owing to moulds and fungus, (Bornehag et al., 2001). Therefore, it is necessary to control the humidity in an indoor environment.

Several techniques for removing moisture from air under ambient pressure conditions have been studied, (Mazzei et al., 2005). A widely used method to dehumidify air is the use of conventional dehumidification systems based on direct expansion units, i.e. DX systems, which operate according to the vapour-compression cycle. However, these units depend mainly on electrical energy. Desiccant dehumidification systems offer a promising alternative to conventional dehumidification units.

A type of desiccant dehumidification system is a desiccant wheel (DW). Many authors have analysed DWs experimentally and numerically (Cao et al., 2014; De Antonellis and Kim, 2018), in particular, focusing on the analysis of parameters influencing outlet process air conditions. The air regeneration temperature is usually used to control the outlet air conditions of DW (Harriman III, 2001), because the higher the regeneration air temperature, the higher is the dehumidification capacity. Nevertheless, significant energy consumption is required to achieve high regeneration temperature. Other studies showed an acceptable dehumidification capacity when the DW was activated at low temperature (Al-Alili et al., 2014; Comino et al., 2016; White et al., 2011), thus reducing the environmental impact associated with air dehumidification. In this study, values below 60 °C were considered as low regeneration temperatures.

* Corresponding author.
E-mail address: francisco.comino@uco.es (F. Comino).

https://doi.org/10.1016/j.ijrefrig.2019.02.002
0140-7007/

6.1.2 Validation of artificial neural networks to model the acoustic behaviour of induction motors

In this paper, we validate the use of ANNs for the modelling of the acoustic behaviour produced by induction motors. Bearing in mind that the population is subject to a wide range of noises, reducing the physical and psychological discomfort associated to any device or activity has a significant impact on our lives. One of these devices widely present in our daily routine are the induction motors, which are used in a wide range of industrial and household applications such as fridges or washing-machines, among others, or in transport vehicles.

The main goal of this paper is to develop a MTEANN able to model both the physical discomfort (typically measured by the equivalent sound pressure level) and the psychological discomfort (generally studied by the loudness, the roughness and the sharpness). Therefore, the use of MTEANNs is justified given that it allows predicting these four outputs simultaneously in a single model, in general, with fewer connections and better performance, given that it benefits from extracting common features of the different tasks.

ANNs have been previously used in this field, however, up-to-the-knowledge of the authors, the development of a single model for predicting both the physical and the psychological discomfort of an electric induction motor has not been studied previously. Besides, a total of 40 inputs has been used to characterise the problem, including harmonic distortions of voltage and intensity, among others. Furthermore, the experimental data included in the study has been obtained from three different modulation techniques, each of them with a different primary objective, aiming to obtain a representative sample of tests and consider a wide variety of values, increasing in this way, the robustness of the models.

Regarding the ANN architectures, the two most common basis functions are used in the hidden layer (concretely, SU and PU), both with linear outputs. Even though the results indicate that both models obtained excellent predictions using a small number of connections, the ANN using PUs in the hidden layer is the one achieving the best performance in terms of both MSE and SEP. This model only considers 10 variables out of the initial 40 inputs. Finally, it could be concluded that MTEANNs are excellent to extract common features of related tasks.

Applied Acoustics 166 (2020) 107332

Contents lists available at ScienceDirect

Applied Acoustics

journal homepage: www.elsevier.com/locate/apacoust

Validation of artificial neural networks to model the acoustic behaviour of induction motors

F.J. Jiménez-Romero [a,*,1], D. Guijo-Rubio [b,1], F.R. Lara-Raya [a], A. Ruiz-González [c], C. Hervás-Martínez [b]

[a] Department of Electrical Engineering, Universidad de Córdoba, Córdoba, Spain
[b] Department of Computer Science and Numerical Analysis, Universidad de Córdoba, Córdoba, Spain
[c] Department of Electrical Engineering, University of Málaga EII, Málaga, Spain

ARTICLE INFO

Article history:
Received 10 January 2020
Received in revised form 6 March 2020
Accepted 12 March 2020

Keywords:
Artificial neural networks
Sigmoid units
Pulse width modulation
Sound quality
Induction motor

ABSTRACT

In the last decade, the sound quality of electric induction motors is a hot topic in the research field. Specially, due to its high number of applications, the population is exposed to physical and psychological discomfort caused by the noise emission. Therefore, it is necessary to minimise its psychological impact on the population. In this way, the main goal of this work is to evaluate the use of multitask artificial neural networks as a modelling technique for simultaneously predicting psychoacoustic parameters of induction motors. Several inputs are used, such as, the electrical magnitudes of the motor power signal and the number of poles, instead of separating the noise of the electric motor from the environmental noise. Two different kind of artificial neural networks are proposed to evaluate the acoustic quality of induction motors, by using the equivalent sound pressure, the loudness, the roughness and the sharpness as outputs. Concretely, two different topologies have been considered: simple models and more complex models. The former are more interpretable, while the later lead to higher accuracy at the cost of hiding the cause-effect relationship. Focusing on the simple interpretable models, product unit neural networks achieved the best results: 38.77 for MSE and 13.11 for SEP. The main benefit of this product unit model is its simplicity, since only 10 inputs variables are used, outlining the effective transfer mechanism of multitask artificial neural networks to extract common features of multiple tasks. Finally, a deep analysis of the acoustic quality of induction motors in done using the best product unit neural networks.

1. Introduction

Electric induction motors are used in a wide range of industrial and household applications, from small electrical devices, to large industrial machinery and transport vehicles.

When an induction motor is designed, it is optimized to work powered by a 50 Hz sinusoidal signal. In these conditions, the motor generates the lowest level of electromagnetic noise. Therefore, the noise increases if the motor is fed by a non-sinusoidal signal, for instance, when a power inverter is used to generate the feed signal, using Pulse Width Modulation (PWM) techniques. This sort of technique is widely used to control the operation of the induction machine, emitting a higher noise [1,2].

Three noise components can be distinguished according to their source: mechanical, aerodynamic and electromagnetic noise. Specifically, the mechanical noise is the result of friction in the shaft bearings, whereas, aerodynamic noise is caused by the flow of air driven by the fan through the machine. On the other hand, the electromagnetic component is originated by the interactions of the electromagnetic fields generated in the stator and rotor.

Abbreviations: ANN, artificial neural network; B, basis function; d/D, number of inputs/dataset; f, frequency; f_c/k_c, control parameter of the HIPWM-FMTC technique; HIPWM-FMTC, harmonics injection pulse width modulation frequency modulated triangular carrier using sinusoidal function; HIPWM-FMTC2, harmonics injection pulse width modulation frequency modulated triangular carrier using linear function; *Ithd*, current distortion harmonic; $I50, \ldots, I2450$, current harmonic of the indicated frequency; k/μ, control parameter of the SLPWM technique; L, loudness; *Laeq*, equivalent sound pressure level; M, modulation index; *MSE*, mean squared error; n, size of the dataset; *Ns*, synchronism speed; p, number of poles; PUNN, product unit neural network; PWM, pulse width modulation; r/R, number of outputs/roughness; *SA*, sharpness; SEP, standard error of prediction; SLPWM, slope pulse width modulation; SUNN, sigmoid unit product unit; *Vthd*, voltage distortion harmonic; $V50, \ldots, V2450$, voltage harmonic of the indicated frequency; w, parameters of the basis function; x, input vector to the ANN; y, real output; $\hat{y}$, predicted output; α, control parameter of the HIPWM-FMTC2 technique; β, coefficients of the model.

* Corresponding author at: Department of Electrical Engineering, Universidad de Córdoba, Campus Rabanales, Edificio Leonardo Da Vinci.
E-mail address: fjjimenez@uco.es (F.J. Jiménez-Romero).

[1] F. J. Jiménez-Romero and D. Guijo-Rubio contributed equally to this work.

https://doi.org/10.1016/j.apacoust.2020.107332
0003-682X/

6.2 Non-temporal data classification: health-related problems

Health is one of the fields taking more advantage of the use of machine learning (ML) techniques, giving that most of the processes involve subjectivity (introduced by medical decisions) and also objectivity (strict mathematical scores). Therefore, developing new techniques bridging the gap between these two extremes has a significant impact on health issues. In this sense, two main health-based applications are presented in this Thesis: 1) human immunodeficiency virus (HIV)/hepatitis C virus (HCV) infection for which determining the typology of patients to be treated with antivirals is of interest, given the huge number of patients requiring the treatment; and 2) liver transplantation, in which we can tackle the problem by performing a survival analysis or by developing a rule-based system for the management of the waiting list, among others.

Main publications associated to this section:

- A. Rivero-Juárez, **D. Guijo-Rubio**, F. Téllez, R. Palacios, D. Merino, J. Macías, J.C. Fernández, P.A. Gutiérrez, A. Rivero and C. Hervás-Martínez. "Using machine learning methods to determine a typology of patients with HIV-HCV infection to be treated with antivirals", PLoS One, Vol. 15(1). 2020, pp. e0227188.
JCR (2019): 2.740 Position: 27/71 (Q2).
DOI: 10.1371/journal.pone.0227188

- **D. Guijo-Rubio**, P.J. Villalón-Vaquero, P.A. Gutiérrez, M.D. Ayllón, J. Briceño y C. Hervás-Martínez. "Modelling survival by machine learning methods in liver transplantation: application to the UNOS dataset". 20th International Conference on Intelligent Data Engineering and Automated Learning (IDEAL 2019). 2019. LNCS, Vol. 11872, pp. 97 – 104.
DOI: 10.1007/978-3-030-33617-2_11

- **D. Guijo-Rubio**, J. Briceño, P. A. Gutiérrez, M.D. Ayllón, R. Ciria, C. Hervás Martínez. "Comparison of statistical methods and machine learning techniques for donor-recipient matching in liver transplantation". PLos One, 2021.
JCR (2019): 2.740 Position: 27/71 (Q2). Accepted.

Other publications associated to this section:

- **D. Guijo-Rubio**, P.A. Gutiérrez y C. Hervás-Martínez. "Machine learning methods in organ transplantation", Current Opinion in Organ Transplantation, Vol. 25(4), 2020, pp. 399 – 405.
JCR (2019): 2.571 Position: 13/24 (Q3).
DOI: 10.1097/MOT.0000000000000774

6.2.1 Using machine learning methods to determine a typology of patients with HIV/HCV infection to be treated with antivirals

In this paper, artificial neural networks (ANNs) are used to identify those factors, for HIV/HCV co-infected patients, that were not included among the prioritisation criteria considered before treatment uptake. In the last few years, direct-acting antiviral drugs have been considered for treating HCV infection with high cure rates. Nevertheless, it can not be universally provided to all the HIV/HCV co-infected patients, due to the high number of people waiting for it. Hence, prioritisation criteria have been established by the competent health authorities with the main goal of achieving the highest survival rates and maximising the benefits for the co-infected patients. However, this strategy has not been assessed yet.

Apart from this lack of evaluation, identifying those variables limiting the treatment uptake for HIV/HCV co-infected patients is of significant interest. For this, data from the Spanish HERACLES cohort has been used. It is worthy of mention that Spain provides universal health care access, and that Spanish health authorities elaborated a national strategy for initialising and prioritising HCV treatment. Nevertheless, clinical, epidemiological and geographic factors associated with low probabilities for having access to the treatment have not been evaluated.

The goals of this paper are twofold: 1) to develop a classification model based on ANNs maximising the global performance (known as correct classification rate (CCR)), and achieving the highest accuracy for the minority class (untreated patients), known as minimum sensitivity (MS); and 2) to analyse the best model obtained in order to study which characteristics present in the patients influence more on the probability to be treated. For this, our study is focused on using ANNs with different basis functions in the hidden layer. In this sense, we have used three different basis functions: sigmoidal units (SUs), product units (PUs) and radial basis functions (RBFs). Regarding the characteristics of the population used in the study, 17 patient variables have been used, such as the age, if they had been in jail or the HCV genotype, among others.

The results obtained have shown that the ANNs using RBFs in the hidden layer have achieved an excellent performance, using only 8 connections. Moreover, this best model only considers six input variables, making the model easy to be interpreted and implemented, leading to a decrease in the quantity of information to be known about the patient, and, therefore, avoiding information errors. Furthermore, one of the variables included by the best model is the recent people who inject drugs (PWID), which is object of discussion. However, its non-inclusion in the best model has reduced significantly the performance, leading to a trivial classifier, classifying almost all the patterns in one class. Therefore, its inclusion is mandatory given its denoted importance.

PLOS ONE

RESEARCH ARTICLE

Using machine learning methods to determine a typology of patients with HIV-HCV infection to be treated with antivirals

Antonio Rivero-Juárez[1☯], David Guijo-Rubio[2☯]*, Francisco Tellez[3], Rosario Palacios[4], Dolores Merino[5], Juan Macías[6], Juan Carlos Fernández[2], Pedro Antonio Gutiérrez[2], Antonio Rivero[1], César Hervás-Martínez[2]

1 Unidad de Enfermedades Infecciosas, Hospital Universitario Reina Sofía de Córdoba, Instituto Maimónides de Investigación Biomédica de Córdoba, Universidad de Córdoba, Córdoba, España, **2** Departamento de Informática y Análisis Numérico, Universidad de Córdoba, Córdoba, España, **3** Unidad de Enfermedades Infecciosas, Hospital Universitario de Puerto Real, Cádiz, España, **4** Unidad de Enfermedades Infecciosas, Hospital Juan Ramón Jiménez e Infanta Elena de Huelva, Huelva, España, **5** Unidad de Enfermedades Infecciosas, Hospital Universitario Virgen de la Victoria, Complejo Hospitalario Provincial de Málaga, Málaga, España, **6** Unidad de Enfermedades Infecciosas, Hospital Universitario de Valme, Instituto de Biomedicina de Sevilla, Sevilla, España

☯ These authors contributed equally to this work.
* dguijo@uco.es

OPEN ACCESS

Citation: Rivero-Juárez A, Guijo-Rubio D, Tellez F, Palacios R, Merino D, Macías J, et al. (2020) Using machine learning methods to determine a typology of patients with HIV-HCV infection to be treated with antivirals. PLoS ONE 15(1): e0227188. https://doi.org/10.1371/journal.pone.0227188

Editor: Yury E. Khudyakov, Centers for Disease Control and Prevention, UNITED STATES

Received: April 24, 2019

Accepted: December 13, 2019

Published: January 10, 2020

Data Availability Statement: An anonymized version of the dataset has been uploaded to the National Addiction and HIV Data Program (NAHDAP) at https://www.icpsr.umich.edu/icpsrweb/NAHDAP/; The ID is NAHDAP-116804.

Funding: DGR, JCF, PAG and CHM were supported by TIN2017-85887-C2-1-P - Spanish Ministry of Economy and Competitiveness (MINECO) and FEDER funds - NO DGR - FPU16/02128 - Spanish Ministry of Education and Science - NO DGR - PI15/01570 - Fundación de Investigación

Abstract

Several European countries have established criteria for prioritising initiation of treatment in patients infected with the hepatitis C virus (HCV) by grouping patients according to clinical characteristics. Based on neural network techniques, our objective was to identify those factors for HIV/HCV co-infected patients (to which clinicians have given careful consideration before treatment uptake) that have not being included among the prioritisation criteria. This study was based on the Spanish HERACLES cohort (NCT02511496) (April-September 2015, 2940 patients) and involved application of different neural network models with different basis functions (product-unit, sigmoid unit and radial basis function neural networks) for automatic classification of patients for treatment. An evolutionary algorithm was used to determine the architecture and estimate the coefficients of the model. This machine learning methodology found that radial basis neural networks provided a very simple model in terms of the number of patient characteristics to be considered by the classifier (in this case, six), returning a good overall classification accuracy of 0.767 and a minimum sensitivity (for the classification of the minority class, *untreated* patients) of 0.550. Finally, the area under the *ROC* curve was 0.802, which proved to be exceptional. The parsimony of the model makes it especially attractive, using just eight connections. The independent variable "recent PWID" is compulsory due to its importance. The simplicity of the model means that it is possible to analyse the relationship between patient characteristics and the probability of belonging to the *treated* group.

6.2.2 Modelling survival by machine learning methods in liver transplantation: application to the UNOS dataset

In this paper, survival analysis (SA) is applied to the liver transplantation (LT) problem, which is an accepted treatment for patients with end-stage chronic liver disease. The united network for organ sharing (UNOS) organisation provided us with the largest dataset regarding transplants and organ sharing with more than 9 kinds of transplants. One of these is liver transplant, which represents almost a 22% of the transplants made in the USA. This database contains more than $200,000$ records and over 380 variables, including variables from the donor, the recipient and from the transplant procedure.

The aim of this paper is to apply SA techniques to model the survival in LT, using the largest dataset available. Note that working with databases including a huge number of transplants is of significant interest, given that it allows us to ensure the applicability of these techniques in predicting LT survival worldwide. More concretely, in this work, the SA techniques used are based on ML algorithms, given that they have been successfully applied to handle survival data in other fields.

SA is a branch of the statistics whose main objective is to model data where the outcome is the time until an event of interest occurs. Note that one of the main characteristics of this kind of analysis is the presence of censored instances, i.e. the event of interest is not registered (in this case the liver graft failure), which can not be handled by standard ML techniques. Therefore, in this paper, the use of complex ML techniques adapted to the SA is needed, given the difficulty of the problem caused not only by the existence of a 74% of censored patterns, but also because the mean censored time is 3.95 years.

The following three groups of SA methods are applied: Cox's-regression-based models, models based on gradient boosting and adaptations of support vector machines (SVMs) paradigm to SA. Results denote that gradient boosting-based models stand out as the best methods regarding the most popular measure in SA, the concordance index (ipcw). On the other hand, Cox's-regression-based models are able to obtain more balanced results.

Modelling Survival by Machine Learning Methods in Liver Transplantation: Application to the UNOS Dataset

David Guijo-Rubio[1]( ), Pedro J. Villalón-Vaquero[1], Pedro A. Gutiérrez[1], Maria Dolores Ayllón[2], Javier Briceño[2], and César Hervás-Martínez[1]

[1] Department of Computer Sciences, Universidad de Córdoba, Córdoba, Spain
{dguijo,i52vivap,pagutierrez,chervas}@uco.es

[2] Unit of Hepatobiliary Surgery and Liver Transplantation, Córdoba, Spain
lolesat83@hotmail.com, javibriceno@hotmail.com

Abstract. The aim of this study is to develop and validate a machine learning (ML) model for predicting survival after liver transplantation based on pre-transplant donor and recipient characteristics. For this purpose, we consider a database from the United Network for Organ Sharing (UNOS), containing 29 variables and 39,095 donor-recipient pairs, describing liver transplantations performed in the United States of America from November 2004 until June 2015. The dataset contains more than a 74% of censoring, being a challenging and difficult problem. Several methods including proportional-hazards regression models and ML methods such as Gradient Boosting were applied, using 10 donor characteristics, 15 recipient characteristics and 4 shared variables associated with the donor-recipient pair. In order to measure the performance of the seven state-of-the-art methodologies, three different evaluation metrics are used, being the concordance index (*ipcw*) the most suitable for this problem. The results achieved show that, for each measure, a different technique obtains the highest value, performing almost the same, but, if we focus on *ipcw*, Gradient Boosting outperforms the rest of the methods.

Keywords: United Network for Organ Sharing · Liver transplant · Survival analysis · Machine learning

1 Introduction

The Survival Analysis (SA) is a field traditionally tackled by statistical methods, aiming to model the data where the outcome is the time until the occurrence of an event of interest. One important characteristic of SA is that, for some of the instances, the outcomes are unobservable since these are no longer monitored or the study has finished previous to the occurrence of the event of interest; these instances are known as censored instances. Note that most of the SA books introduce the topic from a pure statistical point of view [1,2].

H. Yin et al. (Eds.): IDEAL 2019, LNCS 11872, pp. 97–104, 2019.
https://doi.org/10.1007/978-3-030-33617-2_11

6.2.3 Comparison of statistical methods and machine learning techniques for donor-recipient matching in liver transplantation

In this paper, we tackle the problem of donor-recipient matching in LT. Donor-recipient matching is one of the most challenging problems nowadays in the heath care system, given the increasing number of recipients and the decreasing number of donors. In this study, we face this problem by using the largest dataset containing donor-recipient pairs in LT, the dataset provided by the UNOS. There are several models in the literature aiming to support donor-recipient matching, such as the model for end-stage liver disease (MELD). However, these models are a subject of discussion, given that some of them focus on reducing the mortality in the waiting list, disregarding the result of the transplant.

The main goal of this study is to analyse the behaviour of different predictive methods in this field. For this, the dataset considered is similar to the one considered in the work presented in previous Subsection 6.2.2. Another important goal of this work is to obtain an efficient and accurate approach combining information from donors, from recipients, and from the pre- and post- transplant characteristics. Furthermore, four different end-points, periods of time for controlling the graft-loss, are considered (3 months, 1, 2 and 5 years).

Regarding the survival prediction methods applied to the donor-recipient matching problem, two groups are considered: 1) classical statistical methods, such as logistic regression (LR), and 2) standard ML techniques, such as SVM, gradient boosting (GB) or multilayer perceptron (MLP), among others. Apart from these predictive models, standard widely used scores are used for comparison purposes. The results obtained have outlined that all the techniques achieve similar performances except LR, which achieves the best performance. Furthermore, focusing on the 5 years end-point, results have demonstrated that LR outperforms also the state-of-the-art scores for donor-recipient matching in LT.

In addition, given that LR is the method achieving the best results for the longest end-point, a rule-based system is developed in order to bridge the gap between the subjective medical decision and the strictly objective mathematical scores. This rule-based system is objective, not including human subjectivity, it is also optimal, in the sense that it is able to increase the post-transplant survival rates, and it is fair, because, without significant differences between two recipients, the donor-recipient matching is done by using the MELD score. Some other conclusions can be drawn from this study, such as the fact that using multicentre datasets is a subject of controversy, given that incongruities can be found in the procedures applied by the different health care systems. Some of these procedures include the way missing data is imputed or categorising a same situation contradictorily.

Statistical methods versus machine learning techniques for donor-recipient matching in liver transplantation

David Guijo-Rubio[1,*], Javier Briceño[2], Pedro Antonio Gutiérrez[1], Maria Dolores Ayllón[2], Rubén Ciria[2], César Hervás-Martínez[1]

1 Department of Computer Sciences and Numerical Analysis, University of Córdoba, Spain.
2 Unit of Hepatobiliary Surgery and Liver Transplantation, Hospital Universitario Reina Sofía, IMIBIC, Córdoba, Spain.

* Corresponding author at: Department of Computer Sciences and Numerical Analysis, Rabanales Campus, 3rd floor Albert Einstein Building, 14071 University of Córdoba, Spain. dguijo@uco.es (D. Guijo-Rubio)

Abstract

Donor-Recipient (D-R) matching is one of the main challenges to be fulfilled nowadays. Due to the increasing number of recipients and the small amount of donors in liver transplantation, the allocation method is crucial. In this paper, to establish a fair comparison, the United Network for Organ Sharing database was used with 4 different end-points (3 months, and 1, 2 and 5 years), with a total of $39,189$ D-R pairs and 28 donor and recipient variables. Modelling techniques were divided into two groups: 1) classical statistical methods, including Logistic Regression (LR) and Naïve Bayes (NB), and 2) standard machine learning techniques, including Multilayer Perceptron (MLP), Random Forest (RF), Gradient Boosting (GB) or Support Vector Machines (SVM), among others. The methods were compared with standard scores, MELD, SOFT and BAR. For the 5-years end-point, LR (AUC = 0.654) outperformed several machine learning techniques, such as MLP (AUC = 0.599), GB (AUC = 0.600), SVM (AUC = 0.624) or RF (AUC = 0.644), among others. Moreover, LR also outperformed standard scores. The same pattern was reproduced for the others 3 end-points. Complex machine learning methods were not able to improve the performance of liver allocation, probably due to the implicit limitations associated to the collection process of the database.

Introduction

Donor-Recipient (D-R) matching is one of the most challenging topics in Liver Transplantation (LT). Considering the increasing number of candidates for LT and the scarce number of available donors, the rationale for assignment of a given donor to potential candidates on a waiting list is a matter of controversy. For this purpose, some scores have been designed, whose implementation in practice has its supporters and detractors. Model for End-Stage Liver Disease (MELD) [1], Survival Following Liver Transplantation score (SOFT) [2] or Balance of Risk (BAR) [3] are examples of the intention to match donors and recipients to obtain the best post-transplant result. However, this result is also a subject of discussion. For some of these scores, the main objective is to decrease the mortality in the waiting list without affecting the result of the transplant. This is the case of MELD, the most widespread prioritization system

We can only see a short distance ahead, but we can see plenty there that needs to be done.

Alan Turing

7 Discussion and conclusions

In this last chapter, the main conclusions extracted from the previous research are described, as well as the research lines to be explored in future works.

7.1 Conclusions

In this Thesis, we mainly focus on three research lines regarding time series and on an additional chapter including some complementary works carried out for covering the challenges of several national projects. The three main working lines regarding time series are: preprocessing and clustering, prediction and classification of time series. In the following subsections, the main contributions to this Thesis are summarised.

7.1.1 Time series preprocessing

Preprocessing is an important area receiving a lot of attention in the last years given its importance for subsequent tasks, such as prediction or classification, among others. Concretely, in this Thesis, we present a novel approach to time series clustering, which consists in grouping time series based on their similarity with the main goal of discovering significant patterns in the dataset. More concretely, our proposal is divided into two stages. The first stage is applied to the time series individually, and it is focused on simplifying the time series as much as possible, keeping the highest quantity of information. The second stage is applied globally to all the time series of the dataset and consists in using a clustering algorithm for grouping the time series.

More concretely, the first stage is divided into three sub-stages: 1) the first sub-stage consists in the application of a segmentation procedure to divide the time series into segments. This segmentation is carried out by means of a growing window, which introduces points until exceeding a maximum error computed from the corresponding least-squares polynomial approximation of the segment. 2) The second sub-stage consists in projecting all the obtained segments into a fixed-size vector of statistical characteristics, in order to generate a common structure for all sorts of segments. 3) The last sub-stage is the clustering of these statistical characteristics vectors representing segments.

Regarding the second stage, once the time series have been reduced to equal-length vectors of statistical characteristics, a common structure for the time series is built by including information of the centroids and of the segments with the highest variance. In this way, both common and specific information of the segments is included, apart from meta-information of the clustering process, such as the number of segments and the error difference between the most distant segments. Finally, this stage finishes with a hierarchical clustering stage, grouping this novel time series representation by their similarity.

This approach is compared against 3 state-of-the-art techniques over 84 datasets from the University of East Anglia and University of California Riverside (UEA/UCR) time series classification (TSC) repository. The results achieved demonstrate that our proposal outperforms the rest of the approaches when dealing with large datasets including long time series. Furthermore, the computational load of our approach is competitive against the rest, being more computationally intensive during the first main stage (segmentation and vector projection procedures).

According to the objectives established in Chapter 2, Chapter 3 satisfies objectives 3 and 4. Moreover, this topic is supported on 1 JCR-indexed journal paper and 2 national/international conferences.

7.1.2 Time series prediction

Prediction is the most important field in time series data mining, mainly given that accurately predicting what is going to happen in the future is a challenging task, increasing its difficulty for larger time prediction horizons. Traditionally, it has been tackled in many several ways, from using standard autoregressive models (ARs) or moving average models (MAs) to more complex approaches based on deep learning (DL), such as long-short-time memory (LSTM) or convolutional neural networks (CNNs). These previous models tackle the time series prediction task using real values. Nevertheless, in this Thesis, we propose several strategies using different types of representation.

On the one hand, we present the idea of transforming the prediction problem into an ordinal classification (OC) task. This idea has been applied in two different approaches: 1) for predicting low-visibility events due to fog, and 2) for predicting convective cloud formation. The first approach is carried out in the Valladolid airport (Spain), in which the presence of fog is frequent. In this paper, the runway visual range (RVR) is used for characterising the quantity of fog for a given day, which is categorised into three categories: CLEAR, MIST and FOG, according to the visibility conditions. Note that there is an ordinal nature among the labels, therefore, the use of ordinal classifiers is interesting for taking advantage of the ordinal information. Hence, we propose the use of three different windows based on AR models: 1) fixed window resembling the behaviour of standard AR models, 2) dynamic window adding values until there is a change in the label, and 3) dynamic window adding values until there is a change in terms of variance. Finally, to achieve this prediction, several ordinal classifiers have been applied to the transformed dataset, comparing them against several state-of-the-art techniques, such as support vector regressor (SVR) or the persistence model, the latter being based on the prediction rule $Y_t = Y_{t-1}$. The idea proposed outperforms the rest of the methodologies in terms of both average mean absolute error (AMAE) and minimum sensitivity (MS).

Regarding the second paper, following the idea of using OC for transforming the original prediction task, we carry out the prediction of convective situations. In this field, there are four different kinds of situations: CLEAR, in which there are no clouds sighted, TCU, which represents cumulus congestus situations, CB, which means that cumulonimbus are sighted, and TS, which represents thunderstorm situations. Moreover, regarding the input features, these events are characterised by means of atmospheric and meteorological data collected from two sources of information: airport's radiosonde station and reanalysis data. As in the previous approach, in this paper, the output is also ordinal, and therefore, the use of ordinal classifiers is justified. Due to the imbalance degree of the dataset, i.e. the amount of days with convective situations is tiny in comparison with the number of clear days, undersampling and ordinal oversampling methods are applied in order to balance the dataset. After that, 13 different ordinal methods are compared against several state-of-the-art techniques, demonstrating that ordinal approaches perform better. Another comparison against terminal aerodrome forecasts (TAFs) is also carried out, denoting that our approach is better for detecting any situation involving convective clouds (CB, TCU and TS), whereas TAFs are better at avoiding false alarms.

On the other hand, the second group of approaches follow the idea of time series forecasting. However, although they belong to the same group, three different perspectives have been applied. In the first paper, we tackle the problem of convective cloud formation from the multi-objective point of view. As previously specified, the dataset is enormously imbalanced, thus, the multi-objective paradigm fit perfectly with this sort of problems. In this sense, we apply multi-objective evolutionary artificial neural networks (MOEANNs) optimising both metrics, the correct classification rate (CCR) and the MS. Concretely, the first one aims to achieve a good global performance, whereas the second one is designed to improve the accuracy of the minority classes. The results achieved demonstrate that the use of the multi-objective methodology is appropriate, given that the best performance in terms of both previous metrics is achieved by a MOEANN using the MS extreme of the Pareto front. This methodology is compared against TAFs, concluding that our proposal improves the results of standard airport mechanisms.

The second paper in this group of approaches consists in solving the challenging solar radiation prediction problem, which is raising more and more attention given the interest in renewable energies. The solar radiation prediction is carried out with a 1h time prediction horizon. Moreover, the predicted variables are obtained from integrating Meteosat satellites with Copernicus atmosphere monitoring service (CAMS) and *Solar-GIS* model. Apart from integrating several sources of information, we propose different configurations for the dataset, with the aim of increasing the understanding about the predictive variables importance. Furthermore, in this paper, we propose the use of several mixtures of basis functions for the hidden and the output layers. Concretely, in this paper we present the mixture of sigmoidal units (SUs) in hidden layer with product units (PUs) in the output layer, resulting in an excellent performance for the hourly prediction of solar energy. Our proposal is compared against state-of-the-art techniques such as extreme learning machines (ELMs), multilayer perceptrons (MLPs) or SVRs, among others.

Finally, the last paper in this group of approaches is based on the application of multi-task evolutionary artificial neural networks (MTEANNs) to the energy flux prediction. The main goal behind this paper is to develop a methodology able to accurately predict the energy flux with the aim of stabilising this source of energy, reducing, in this sense, as much as possible, the influence of the tides and waves stochastic nature. For this, we select four different time prediction horizons (6h, 12h, 24h and 48h). Moreover, trying to avoid using information collected by measurement instruments, all the input data used in this paper is obtained from reanalysis sources. This problem is tackled by means of MTEANNs using several basis functions in the hidden layer (SUs, PUs and radial basis functions (RBFs)), in combination with linear models and PUs for the output layer. Our proposal is compared against state-of-the-art methods such as SVRs and ELMs, concluding that our approach is an excellent technique for both short- and long-term energy flux prediction.

According to the objectives established in Chapter 2, Chapter 4 partially satisfies objectives 1, 2, 3, and 7. Regarding publications related with the prediction of time series, 5 JCR-indexed journal papers and 3 national/international conferences have been published.

7.1.3 Time series classification

TSC is the most popular field of time series data mining in the literature due to the huge interest developed in the last few years. One of the main reasons behind this success is the existence of a global repository with more than 150 time series datasets increasing more and more, known as the UEA/UCR TSC repository. Currently, hierarchical vote system collective of transformation-based ensemble (HIVE-COTE) [120] is one of the approaches achieving the best performance for most of the datasets, but it is computationally intensive, due to the fact that it consists in an ensemble embedding five different approaches. Therefore, one of the main challenges being tackled nowadays is improving the performance of HIVE-COTE (or at least obtaining a non-significantly worse performance) while decreasing significantly the computational load. In this sense, several approaches have been presented to the literature. In this Thesis, we propose a hybrid approach for the shapelet transform (ST) methodology, in order to achieve a competitive performance with a significant decrease in the computational time.

The method presented to the literature consists in developing a hybrid model between the standard ST and the learned shapelets (LS) [76] methodology. The main idea behind this method is taking advantage from the data driven shapelet search of the ST and from searching for shapelets in the entire shapelet space through stochastic gradient descent (SGD) (the case of LS). For this, the main proposal lays in hybridising these two methodologies, including a time constraint of 1 hour data driven search, in which the best k shapelets are extracted from the training set, and then, using these k best shapelets in the first layer of a CNN to optimise them. This layer is divided into: 1) a feature extraction step computing the distances between the shapelet and the original time series, and 2) a pooling step keeping the minimum of all these distances. Moreover, the optimisation procedure consists in tuning both the shapelet values and the weights of the logistic regression (LR) using SGD. Once the shapelets are optimised, they are extracted in order to build the transformed dataset, and after that, the rotation forest (RF) classifier is applied.

The results achieved by this hybrid method are significantly better than either approach in isolation. Moreover, the results of our proposal are highly competitive with respect to the full shapelet search included in the ST embedded in the HIVE-COTE, being the computational load much lower.

Moreover, a second line of research is carried out in this field. From the whole set of time series datasets of the UEA/UCR TSC repository, there are several time series datasets including ordinal information among the labels. The classification of these time series datasets is proposed to be known as time series ordinal classification (TSOC). Up-to-the-knowledge of the authors, this field is unexplored in a general way, no approaches having been published in the literature. Therefore, in this Thesis, we present a first approach based on the ST, given its importance among the TSC field and a comparison baseline against the main approaches in the state-of-the-art. The ST pseudocode is composed of several steps in which ordinal information could be taken into account. In this way, one of these steps is the quality measurement of the shapelets. For this, we propose up to three different shapelet quality measures considering ordinal information. They are based on adaptations to the ordinal paradigm of traditional indices such as the Fisher score or the Pearson's and Spearman's determination coefficients, instead of using the information gain (IG), which is the standard shapelet quality measure. The results demonstrate that the Pearson's determination coefficient adaptation for shapelet quality measure outperforms the rest of the scores, independently of the subsequent classifier used. Furthermore, regarding the classifiers applied to the transform, the most popular ordinal classifiers (proportional odds model (POM) and support vector for ordinal regression with implicit constraints (SVORIM)) and other standard classifiers in the literature are selected, in order to demonstrate that ordinal techniques are able to achieve better results. The results achieved by SVORIM are better than those obtained by POM or SVR, among others. Hence, the first paper concludes that the ST using the Pearson's determination coefficient along with the use of the SVORIM ordinal classifier obtains the best results, this difference being statistically significant.

Moreover, this proposed methodology is compared, in a second paper, against the main state-of-the-art techniques in TSC, which are HIVE-COTE, time series combination of heterogeneous and integrated embedding forest (TS-CHIEF) [168] and inceptionTime [63]. These techniques are highly competitive in terms of accuracy, and even though HIVE-COTE is the best of them, TS-CHIEF and inceptionTime stand out for their scalability and efficiency. However, when dealing with ordinal datasets, all of them are outperformed by the ST version adapted to OC, using the Pearson's determination coefficient as shapelet quality measure and SVORIM as the final classifier. Also, it is worthy of mention that nominal classifiers benefit from the ordinal information induced by the ST using the Pearson's determination coefficient and achieve better results than the three previous ensemble approaches.

According to the objectives established in Chapter 2, Chapter 5 partially satisfies objectives 3, 5, and 6. Moreover, the works related to this topic have been done in collaboration with the University of East Anglia (UEA), in which two international research stays were done (3 months each). The research on this topic is based on 4 international conferences.

7.1.4 Additional works

Although this Thesis is based on time series data mining, given that the research group of the author is immersed in several regional and national projects, the following works are proposed for solving real-world problems regarding many different fields. Concretely, two main areas have been considered: 1) engineering applications, in which modelling of desiccant wheels (DWs) and modelling of the acoustic behaviour of induction motors have been tackled, and 2) health-related problems, in which the work has been focused on the human immunodeficiency virus (HIV)/hepatitis C virus (HCV) disease and in liver transplantation (LT).

On the one hand, regarding the engineering applications, two main real-world problems have been tackled. These engineering applications typically concern more than one objective, trying to optimise them all simultaneously. Hence, in this Thesis, we carry out the modelling of DWs and the induction motors acoustic behaviour by using MTEANNs, which are able to exploit the shared information for related tasks. More concretely, in the first problem considered we deal with DWs used to control the internal moisture content of buildings, having a significant impact on some industries, given that it avoids the germination of fungus, among other advantages. For this, a wide variety of operating conditions for DWs are used for developing a mathematical model able to predict the outlet process air temperature and the outlet process air humidity ratio. This mathematical model consists in a MTEANN using SUs in the hidden layer, which is able to achieve an excellent performance for both outputs. Regarding the second engineering application, MTEANNs are also proposed for modelling the acoustic behaviour of induction motors. Noises produce huge physical and psychological discomfort to the population, hence, aiming to reduce it can have a significant impact on our daily lives. The physical and psychological discomfort is measured by means of several indices, thus, we propose the use of MTEANNs to extract common features from these indices. The best results are achieved by a MTEANN using PUs in the hidden layer. Furthermore, it is worthy of mention that this best model is very simple and only uses a quarter of the input variables.

On the other hand, regarding health-based problems, other two applications are considered: firstly we tackle the HIV/HCV disease, for which determining the typology of co-infected patient to be treated is of enormous interest, given that the number of patients requiring the treatment is high, and therefore, prioritisation criteria must be established. Nevertheless, this strategy has not been previously evaluated, paying, in this sense, low attention to the variables limiting the treatment uptake. In this Thesis, we propose the use of evolutionary artificial neural networks (EANNs) to develop the simplest possible model achieving a good performance, in order to analyse which characteristics influence more on the probability to be treated. The best results are achieved by an artificial neural network (ANN) with RBFs in the hidden layer. This model only considers 6 inputs, being easy to be interpreted and implemented. Regarding the second health-based problem, the LT problem, we present two different ideas for approaching it using the united network for organ sharing (UNOS) database. This database is composed of more than $200,000$ records including information from all the parts involved in the LT process: donor, recipient and transplant procedure. In the first work, survival analysis (SA) techniques are applied in order to model the survival in LT. For this, several approaches based on machine learning (ML) techniques are proposed, the gradient boosting (GB) technique applied to SA being the one achieving the best results. Furthermore, in this Thesis, we also tackle the problem of donor-recipient matching in LT by means of ML classification tasks. Traditionally, this matching has been performed following scores such as model for end-stage liver disease (MELD), but this disregards the result of the transplant. In this sense, we propose the use of ML techniques and compare them against traditional statistical methods and some other scores apart from MELD. Besides, we also present a rule-based system halfway between strict objectivity, included by ML methods, and subjectivity, induced by the medical decision.

According to the objectives established in Chapter 2, Chapter 6 partially satisfies objectives 1, 2, and 7. Regarding publications, 4 JCR-indexed journal papers (another one is currently under review) and 2 national/international conferences have been published.

7.2 General discussion and future work

In this Thesis, three different topics are covered concerning time series data mining and one extra topic concerning additional works. All of them are solved using a diverse range of machine learning (ML) approaches. The main proposals are presented in 11 international journal papers and 14 national/international conference papers. More concretely, the topics of this Thesis include: 1) preprocessing of time series, where a time series clustering algorithm based on time series segmentation is proposed; 2) prediction of time series, in which several approaches are followed to solve this challenging task; 3) classification of time series, in which both time series classification (TSC) and time series ordinal classification (TSOC) are approached; and 4) additional works related to several real-world problems which are solved using a wide variety of ML techniques.

On the first topic, we approach the time series preprocessing area by developing a novel technique for time series clustering based on time series segmentation and statistical techniques. On the second topic, we present several ways for tackling the prediction of time series, considering both nominal/ordinal classification and regression tasks, using a wide variety of paradigms. On the third topic, we propose a hybrid method for TSC based on shapelet transform (ST) and a first approach to a novel area, the TSOC field. Finally, on the forth topic, we use different ML techniques for solving additional works helping to tackle the challenges of different regional and national projects.

All of these contributions are applied to a wide range of real-world problems, which include fog prediction and detection of convective situations, wave height and solar radiation forecasting, donor-recipient matching in liver transplantation (LT), the identification of human immunodeficiency virus (HIV)/hepatitis C virus (HCV) co-infected patient typology, and the modelling of engineering processes, such as desiccant wheels (DWs) or the acoustic behaviour of induction motors. Therefore, it can be concluded that all the challenges and objectives presented in Chapter 2 are successfully fulfilled.

Furthermore, as future work, the following ideas can be considered. Firstly, regarding the time series clustering approach presented in this Thesis, instead of using a segmentation procedure based on growing windows, the use of more complex time series segmentation techniques based on genetic algorithms could improve the performance. Moreover, the idea of simplifying the time series using statistical characteristics can be used as previous step for prediction or classification (TSC or TSOC) tasks.

Regarding the time series prediction topic, the prediction of convective situations can be solved by means of label distribution learning techniques [71], given the nature of the original dataset. This novel learning paradigm consists in predicting the degree to which each label describes the instance.

Finally, transforming 1D time series to 2D image-like representation is a recent research line being tackled at the moment of writing this Thesis, which opens a wide research field in time series data mining. Given the outstanding performance achieved by deep learning (DL) techniques, the use of 2D image-like time series representation has been shown to be an interesting idea for solving challenging tasks, such as TSC and TSOC, or other unsupervised tasks, such as time series clustering.

More concretely, the main advantage of this transformation is enabling the subsequent application of a convolutional neural network (CNN) model, which has been proved to be excellent when applied to images. First approaches following this idea have been recently proposed in the literature [38, 157], demonstrating its potential. The way time series are converted to images includes several steps: first of all, time series are simplified by means of a time series segmentation approach, then, the segments obtained are projected into vectors of statistical features which are converted to images by applying a given transformation. After that, once the images are obtained, a CNN is considered. Note that there are a lot of possibilities, from modifying the technique for reducing the dimensionality of the time series or the way images are created, to the application of different CNN models.

References

[1] S. Aghabozorgi, A. S. Shirkhorshidi, and T. Y. Wah. Time-series clustering–a decade review. *Information Systems*, 53:16–38, 2015.

[2] S. Aghabozorgi and Y. W. Teh. Stock market co-movement assessment using a three-phase clustering method. *Expert Systems with Applications*, 41(4):1301–1314, 2014.

[3] R. Agrawal, J. Gehrke, D. Gunopulos, and P. Raghavan. Automatic subspace clustering of high dimensional data for data mining applications. In *Proceedings of the 1998 ACM SIGMOD international conference on Management of data*, pages 94–105, 1998.

[4] A. Agresti. *Analysis of ordinal categorical data*, volume 656. John Wiley & Sons, 2010.

[5] R. Ahmed, S. Dey, and M. Mohan. A study to improve night time fog detection in the indo-gangetic basin using satellite data and to investigate the connection to aerosols. *Meteorological Applications*, 22(4):689–693, 2015.

[6] H. Akaike. A new look at the statistical model identification. *IEEE transactions on automatic control*, 19(6):716–723, 1974.

[7] M. Alharbi. Daily global solar radiation forecasting using ann and extreme learning machine: A case study in saudi arabia. Master's thesis, Dalhousie University, Halifax, Nova Scotia, 2013.

[8] A. Angstrom. Solar and terrestrial radiation. report to the international commission for solar research on actinometric investigations of solar and atmospheric radiation. *Quarterly Journal of the Royal Meteorological Society*, 50(210):121–126, 1924.

[9] M. Ankerst, M. M. Breunig, H.-P. Kriegel, and J. Sander. Optics: ordering points to identify the clustering structure. *ACM Sigmod record*, 28(2):49–60, 1999.

[10] O. Arbelaitz, I. Gurrutxaga, J. Muguerza, J. M. PéRez, and I. Perona. An extensive comparative study of cluster validity indices. *Pattern Recognition*, 46(1):243–256, 2013.

[11] N. Asadi, A. Mirzaei, and E. Haghshenas. Creating discriminative models for time series classification and clustering by HMM ensembles. *IEEE transactions on cybernetics*, 46(12):2899–2910, 2016.

[12] M. D. Ayllón, R. Ciria, M. Cruz-Ramírez, M. Pérez-Ortiz, I. Gómez, R. Valente, J. O'Grady, M. de la Mata, C. Hervás-Martínez, N. D. Heaton, et al. Validation of artificial neural networks as a methodology for donor-recipient matching for liver transplantation. *Liver Transplantation*, 24(2):192–203, 2018.

[13] S. Baccianella, A. Esuli, and F. Sebastiani. Evaluation measures for ordinal regression. In *2009 Ninth international conference on intelligent systems design and applications*, pages 283–287. IEEE, 2009.

[14] T. Back. *Evolutionary algorithms in theory and practice: evolution strategies, evolutionary programming, genetic algorithms*. Oxford university press, 1996.

[15] A. Bagnall, M. Flynn, J. Large, J. Line, A. Bostrom, and G. Cawley. Is rotation forest the best classifier for problems with continuous features? *arXiv preprint arXiv:1809.06705*, 2018.

[16] A. Bagnall and G. Janacek. A run length transformation for discriminating between auto regressive time series. *Journal of classification*, 31(2):154–178, 2014.

[17] A. Bagnall, J. Lines, A. Bostrom, J. Large, and E. Keogh. The great time series classification bake off: a review and experimental evaluation of recent algorithmic advances. *Data Mining and Knowledge Discovery*, 31(3):606–660, 2017.

[18] A. Bagnall, J. Lines, J. Hills, and A. Bostrom. Time-series classification with cote: the collective of transformation-based ensembles. *IEEE Transactions on Knowledge and Data Engineering*, 27(9):2522–2535, 2015.

[19] A. Baraldi and P. Blonda. A survey of fuzzy clustering algorithms for pattern recognition. i. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 29(6):778–785, 1999.

[20] V. Batagelj and M. Bren. Comparing resemblance measures. *Journal of classification*, 12(1):73–90, 1995.

[21] M. G. Baydogan, G. Runger, and E. Tuv. A bag-of-features framework to classify time series. *IEEE transactions on pattern analysis and machine intelligence*, 35(11):2796–2802, 2013.

[22] T. Bergot, E. Terradellas, J. Cuxart, A. Mira, O. Liechti, M. Mueller, and N. W. Nielsen. Intercomparison of single-column numerical models for the prediction of radiation fog. *Journal of Applied Meteorology and Climatology*, 46(4):504–521, 2007.

[23] S. A. Billings and G. L. Zheng. Radial basis function network configuration using genetic algorithms. *Neural Networks*, 8(6):877–890, 1995.

[24] A. Binojkumar, B. Saritha, and G. Narayanan. Acoustic noise characterization of space-vector modulated induction motor drives—an experimental approach. *IEEE Transactions on Industrial Electronics*, 62(6):3362–3371, 2014.

[25] F. Birol. Key world energy statistics. *International Energy Agency*, 2017.

[26] C. M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.

[27] A. Bostrom and A. Bagnall. Binary shapelet transform for multiclass time series classification. In *Transactions on Large-Scale Data-and Knowledge-Centered Systems XXXII*, pages 24–46. Springer, 2017.

[28] J. B. Bremnes and S. C. Michaelides. Probabilistic visibility forecasting using neural networks. In *Fog and Boundary Layer Clouds: Fog Visibility and Forecasting*, pages 1365–1381. Springer, 2007.

[29] J. Briceño, M. Cruz-Ramírez, M. Prieto, M. Navasa, J. O. De Urbina, R. Orti, M.-Á. Gómez-Bravo, A. Otero, E. Varo, S. Tomé, et al. Use of artificial intelligence as an innovative donor-recipient matching model for liver transplantation: results from a multicenter spanish study. *Journal of hepatology*, 61(5):1020–1028, 2014.

[30] N. Burgess. A constructive algorithm that converges for real-valued input patterns. *International Journal of Neural Systems*, 5(01):59–66, 1994.

[31] T. Caliński and J. Harabasz. A dendrite method for cluster analysis. *Communications in Statistics-theory and Methods*, 3(1):1–27, 1974.

[32] Z. Cao and H. Cai. Identification of forcing mechanisms of convective initiation over mountains through high-resolution numerical simulations. *Advances in Atmospheric Sciences*, 33(10):1104, 2016.

[33] M. E. Celebi. *Partitional clustering algorithms*. Springer, 2014.

[34] V. Chandola, A. Banerjee, and V. Kumar. Anomaly detection: A survey. *ACM computing surveys (CSUR)*, 41(3):1–58, 2009.

[35] K.-Y. Chang, C.-S. Chen, and Y.-P. Hung. Ordinal hyperplanes ranker with cost sensitivities for age estimation. In *CVPR 2011*, pages 585–592. IEEE, 2011.

[36] C. Chatfield. *Time-series forecasting*. CRC press, 2000.

[37] H. Chen, F. Tang, P. Tino, and X. Yao. Model-based kernel for efficient time series analysis. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 392–400, 2013.

[38] W. Chen and K. Shi. A deep learning framework for time series classification using relative position matrix and convolutional neural network. *Neurocomputing*, 359:384–394, 2019.

[39] W. Chu and S. S. Keerthi. Support vector ordinal regression. *Neural computation*, 19(3):792–815, 2007.

[40] C. A. C. Coello, G. B. Lamont, D. A. Van Veldhuizen, et al. *Evolutionary algorithms for solving multi-objective problems*, volume 5. Springer, 2007.

[41] R. d. O. Colabone, A. L. Ferrari, F. A. d. S. Vecchia, and A. R. B. Tech. Application of artificial neural networks for fog forecast. *Journal of Aerospace Technology and Management*, 7(2):240–246, 2015.

[42] F. Comino, M. R. de Adana, and F. Peci. First and second order simplified models for the performance evaluation of low temperature activated desiccant wheels. *Energy and Buildings*, 116:574–582, 2016.

[43] F. Comino, D. Guijo-Rubio, M. R. de Adana, and C. Hervás-Martínez. Validation of multitask artificial neural networks to model desiccant wheels activated at low temperature. *International Journal of Refrigeration*, 100:434–442, 2019.

[44] L. Cornejo-Bueno, J. Nieto-Borge, P. García-Díaz, G. Rodríguez, and S. Salcedo-Sanz. Significant wave height and energy flux prediction for marine energy applications: A grouping genetic algorithm–extreme learning machine approach. *Renewable Energy*, 97:380–389, 2016.

[45] C. Cortes and V. Vapnik. Support vector machine. *Machine learning*, 20(3):273–297, 1995.

[46] M. Costa. Probabilistic interpretation of feedforward network outputs, with relationships to statistical prediction of ordinal quantities. *International journal of neural systems*, 7(05):627–637, 1996.

[47] L. Cuadra, S. Salcedo-Sanz, J. Nieto-Borge, E. Alexandre, and G. Rodríguez. Computational intelligence in wave energy: Comprehensive review and case study. *Renewable and Sustainable Energy Reviews*, 58:1223–1246, 2016.

[48] G. Cybenko. Approximation by superpositions of a sigmoidal function. *Mathematics of control, signals and systems*, 2(4):303–314, 1989.

[49] D. L. Davies and D. W. Bouldin. A cluster separation measure. *IEEE transactions on pattern analysis and machine intelligence*, PAMI-1(2):224–227, 1979.

[50] K. Deb. *Multi-objective optimization using evolutionary algorithms*, volume 16. John Wiley & Sons, 2001.

[51] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, 6(2):182–197, April 2002.

[52] H. Deng, G. Runger, E. Tuv, and M. Vladimir. A time series forest for classification and feature extraction. *Information Sciences*, 239:142–153, 2013.

[53] S. Dey. On the theoretical aspects of improved fog detection and prediction in India. *Atmospheric Research*, 202:77–80, 2018.

[54] D. L. Donoho and I. M. Johnstone. Projection-based approximation and a duality with kernel methods. *The Annals of Statistics*, pages 58–106, 1989.

[55] J. C. Dunn. A fuzzy relative of the isodata process and its use in detecting compact well-separated clusters. *Journal of Cybernetics*, 3:32–57, 1973.

[56] R. Durbin and D. E. Rumelhart. Product units: A computationally powerful and biologically plausible extension to backpropagation networks. *Neural computation*, 1(1):133–142, 1989.

[57] D. Dutta and S. Chaudhuri. Nowcasting visibility during wintertime fog over the airport of a metropolis of India: decision tree algorithm and artificial neural network approach. *Natural Hazards*, 75:1349–1368, 2015.

[58] M. Ester, H.-P. Kriegel, J. Sander, X. Xu, et al. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Kdd*, volume 96 (34), pages 226–231, 1996.

[59] V. Estivill-Castro. Why so many clustering algorithms: a position paper. *ACM SIGKDD explorations newsletter*, 4(1):65–75, 2002.

[60] D. Fabbian, R. de Dear, and S. Lellyett. Application of artificial neural network forecasts to predict fog at Canberra international airport. *Weather and Forecasting*, 22(2):372–381, 2007.

[61] Y. Fang, H. Chen, and T. Zhang. Contribution of acoustic harmonics to sound quality of pure electric powertrains. *IET Electric Power Applications*, 12(6):808–814, 2018.

[62] H. I. Fawaz, G. Forestier, J. Weber, L. Idoumghar, and P.-A. Muller. Deep learning for time series classification: a review. *Data Mining and Knowledge Discovery*, 33(4):917–963, 2019.

[63] H. I. Fawaz, B. Lucas, G. Forestier, C. Pelletier, D. F. Schmidt, J. Weber, G. I. Webb, L. Idoumghar, P.-A. Muller, and F. Petitjean. Inceptiontime: Finding alexnet for time series classification. *Data Mining and Knowledge Discovery*, 34(6):1936–1962, 2020.

[64] J. C. Fernández, F. J. Martínez, C. Hervás, and P. A. Gutiérrez. Sensitivity versus accuracy in multiclass problems using memetic pareto evolutionary neural networks. *IEEE Transactions on Neural Networks*, 21(5):750–770, 2010.

[65] J. C. Fernández, S. Salcedo-Sanz, P. A. Gutiérrez, E. Alexandre, and C. Hervás-Martínez. Significant wave height and energy flux range forecast with machine learning classifiers. *Engineering Applications of Artificial Intelligence*, 43:44–53, 2015.

[66] E. Frank and M. Hall. A simple approach to ordinal classification. In *European Conference on Machine Learning*, pages 145–156. Springer, 2001.

[67] T.-C. Fu. A review on time series data mining. *Engineering Applications of Artificial Intelligence*, 24(1):164 – 181, 2011.

[68] E. Fuchs, T. Gruber, J. Nitschke, and B. Sick. Online segmentation of time series based on polynomial least-squares approximations. *IEEE Trans. Pattern Anal. Mach. Intell.*, 32(12):2232–2245, 2010.

[69] J. G. Galway. The lifted index as a predictor of latent instability. *Bulletin of the American Meteorological Society*, 37(10):528–529, 1956.

[70] G. Gan, C. Ma, and J. Wu. *Data clustering: theory, algorithms, and applications*. SIAM, 2007.

[71] X. Geng. Label distribution learning. *IEEE Transactions on Knowledge and Data Engineering*, 28(7):1734–1748, 2016.

[72] S. Ghimire, R. C. Deo, N. Raj, and J. Mi. Deep solar radiation forecasting with convolutional neural network and long short-term memory network algorithms. *Applied Energy*, 253:113541, 2019.

[73] S. Ghimire, R. C. Deo, N. Raj, and J. Mi. Wavelet-based 3-phase hybrid svr model trained with satellite-derived predictors, particle swarm optimization and maximum overlap discrete wavelet transform for solar radiation prediction. *Renewable and Sustainable Energy Reviews*, 113:109247, 2019.

[74] R. Ghorbani and R. Ghousi. Predictive data mining approaches in medical diagnosis: A review of some diseases prediction. *International Journal of Data and Network Science*, 3(2):47–70, 2019.

[75] F. Glover and M. Laguna. Tabu search. In *Handbook of combinatorial optimization*, pages 2093–2229. Springer, 1998.

[76] J. Grabocka, N. Schilling, M. Wistuba, and L. Schmidt-Thieme. Learning time-series shapelets. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 392–401, 2014.

[77] D. Graves and W. Pedrycz. Proximity fuzzy clustering and its application to time series clustering and prediction. In *Intelligent Systems Design and Applications (ISDA), 2010 10th International Conference on*, pages 49–54. IEEE, 2010.

[78] S. Guha, R. Rastogi, and K. Shim. Cure: an efficient clustering algorithm for large databases. *ACM Sigmod record*, 27(2):73–84, 1998.

[79] D. Guijo-Rubio, A. Durán-Rosal, P. Gutiérrez, A. Gómez-Orellana, C. Casanova-Mateo, J. Sanz-Justo, S. Salcedo-Sanz, and C. Hervás-Martínez. Evolutionary artificial neural networks for accurate solar radiation prediction. *Energy*, 210:118374, 2020.

[80] D. Guijo-Rubio, P. A. Gutiérrez, C. Casanova-Mateo, J. C. Fernández, A. M. Gómez-Orellana, P. Salvador-González, S. Salcedo-Sanz, and C. Hervás-Martínez. Prediction of convective clouds formation using evolutionary neural computation techniques. *Neural Computing and Applications*, pages 1–13, 2020.

[81] P. A. Gutiérrez, M. Perez-Ortiz, J. Sanchez-Monedero, F. Fernandez-Navarro, and C. Hervas-Martinez. Ordinal regression methods: survey and experimental study. *IEEE Transactions on Knowledge and Data Engineering*, 28(1):127–146, 2016.

[82] I. Guyon and A. Elisseeff. An introduction to variable and feature selection. *Journal of machine learning research*, 3(Mar):1157–1182, 2003.

[83] I. Guyon, S. Gunn, M. Nikravesh, and L. A. Zadeh. *Feature extraction: foundations and applications*, volume 207. Springer, 2008.

[84] E. F. Harrington. Online ranking/collaborative filtering using the perceptron algorithm. In *Proceedings of the 20th International Conference on Machine Learning (ICML-03)*, pages 250–257, 2003.

[85] N. Hatami, Y. Gavet, and J. Debayle. Classification of time-series images using deep convolutional neural networks. In *Tenth international conference on machine vision (ICMV 2017)*, volume 10696, page 106960Y. International Society for Optics and Photonics, 2018.

[86] V. Hautamaki, P. Nykanen, and P. Franti. Time-series clustering by approximate prototypes. In *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*, pages 1–4. IEEE, 2008.

[87] W. He, G. Feng, Q. Wu, T. He, S. Wan, and J. Chou. A new method for abrupt dynamic change detection of correlated time series. *International Journal of climatology*, 32(10):1604–1614, 2012.

[88] C. Hervás, P. A. Gutierrez, M. Silva, and J. M. Serrano. Combining classification and regression approaches for the quantification of highly overlapping capillary electrophoresis peaks by using evolutionary sigmoidal and product unit neural networks. *Journal of Chemometrics: A Journal of the Chemometrics Society*, 21(12):567–577, 2007.

[89] J. Hills, J. Lines, E. Baranauskas, J. Mapp, and A. Bagnall. Classification of time series by shapelet transformation. *Data Mining and Knowledge Discovery*, 28(4):851–881, 2014.

[90] G. Hinton and T. Sejnowski. *Unsupervised Learning: Foundations of Neural Computation*. Computational Neuroscience. Mit Press, 1999.

[91] A. E. Hoerl, R. W. Kannard, and K. F. Baldwin. Ridge regression: some simulations. *Communications in Statistics-Theory and Methods*, 4(2):105–123, 1975.

[92] K. Hornik, M. Stinchcombe, H. White, et al. Multilayer feedforward networks are universal approximators. *Neural networks*, 2(5):359–366, 1989.

[93] H. B. Huang, X. R. Huang, R. X. Li, T. C. Lim, and W. P. Ding. Sound quality prediction of vehicle interior noise using deep belief networks. *Applied Acoustics*, 113:149–161, 2016.

[94] G. Ibarra-Berastegi, J. Saénz, G. Esnaola, A. Ezcurra, and A. Ulazia. Short-term forecasting of the wave energy flux: Analogues, random forests, and physics-based models. *Ocean Engineering*, 104:530–539, 2015.

[95] A. ICAO. 3: Annex 3 to the convention on international civil aviation: Meteorological service for international air navigation, 2004.

[96] C. Igel and M. Hüsken. Improving the rprop learning algorithm. In *Proceedings of the second international ICSC symposium on neural computation (NC 2000)*, volume 2000, pages 115–121. Citeseer, 2000.

[97] M. Iqbal. *An introduction to solar radiation*. Elsevier, 2012.

[98] F. Itakura. Minimum prediction residual principle applied to speech recognition. *IEEE Transactions on acoustics, speech, and signal processing*, 23(1):67–72, 1975.

[99] D. Jani, M. Mishra, and P. Sahoo. Performance prediction of rotary solid desiccant dehumidifier in hybrid air-conditioning system using artificial neural network. *Applied Thermal Engineering*, 98:1091–1103, 2016.

[100] Y.-S. Jeong, M. K. Jeong, and O. A. Omitaomu. Weighted dynamic time warping for time series classification. *Pattern recognition*, 44(9):2231–2240, 2011.

[101] Y. Jin. *Multi-objective machine learning*, volume 16. Springer Science & Business Media, 2006.

[102] P. S. Kamath and W. R. Kim. The model for end-stage liver disease (meld). *Hepatology*, 45(3):797–805, 2007.

[103] M. S. G. Karypis, V. Kumar, and M. Steinbach. A comparison of document clustering techniques. In *TextMining Workshop at KDD2000 (May 2000)*, 2000.

[104] L. Kaufman and P. J. Rousseeuw. *Finding groups in data: an introduction to cluster analysis*, volume 344. John Wiley & Sons, 2009.

[105] L. Kaufmann. Clustering by means of medoids. In *Proc. Statistical Data Analysis Based on the L1 Norm Conference, Neuchatel, 1987*, pages 405–416, 1987.

[106] E. Keogh, S. Chu, D. Hart, and M. Pazzani. Segmenting time series: A survey and novel approach. In *Data mining in time series databases*, pages 1–21. World Scientific, 2004.

[107] E. Keogh and S. Kasetty. On the need for time series data mining benchmarks: a survey and empirical demonstration. *Data Mining and knowledge discovery*, 7(4):349–371, 2003.

[108] S. B. Kotsiantis and P. E. Pintelas. A cost sensitive technique for ordinal classification problems. In *Hellenic Conference on Artificial Intelligence*, pages 220–229. Springer, 2004.

[109] S. Kramer, G. Widmer, B. Pfahringer, and M. De Groeve. Prediction of ordinal classes using regression trees. *Fundamenta Informaticae*, 47(1-2):1–13, 2001.

[110] H.-P. Kriegel, P. Kröger, J. Sander, and A. Zimek. Density-based clustering. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 1(3):231–240, 2011.

[111] N. K. Kumar, R. Savitha, and A. Al Mamun. Ocean wave height prediction using ensemble of extreme learning machine. *Neurocomputing*, 277:12–20, 2018.

[112] M. Leng, X. Lai, G. Tan, and X. Xu. Time series representation for anomaly detection. In *Computer Science and Information Technology, 2009. ICCSIT 2009. 2nd IEEE International Conference on*, pages 628–632. IEEE, 2009.

[113] T. W. Liao. Clustering of time series data—a survey. *Pattern Recognition*, 38(11):1857–1874, 2005.

[114] H.-T. Lin and L. Li. Reduction from cost-sensitive ordinal ranking to weighted binary classification. *Neural Computation*, 24(5):1329–1367, 2012.

[115] J. Lin, E. Keogh, S. Lonardi, J. P. Lankford, and D. M. Nystrom. Visually mining and monitoring massive time series. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 460–469, 2004.

[116] J. Lin, E. Keogh, L. Wei, and S. Lonardi. Experiencing sax: a novel symbolic representation of time series. *Data Mining and knowledge discovery*, 15(2):107–144, 2007.

[117] J. Lin, R. Khade, and Y. Li. Rotation-invariant similarity in time series using bag-of-patterns representation. *Journal of Intelligent Information Systems*, 39(2):287–315, 2012.

[118] Y. Lin, S. Dong, and S. Tao. Modelling long-term joint distribution of significant wave height and mean zero-crossing wave period using a copula mixture. *Ocean Engineering*, 197:106856, 2020.

[119] J. Lines and A. Bagnall. Time series classification with ensembles of elastic distance measures. *Data Mining and Knowledge Discovery*, 29(3):565–592, 2015.

[120] J. Lines, S. Taylor, and A. Bagnall. Time series classification with hive-cote: The hierarchical vote collective of transformation-based ensembles. *ACM Transactions on Knowledge Discovery from Data*, 12(5), 2018.

[121] R. P. Lippmann. Pattern classification using neural networks. *IEEE communications magazine*, 27(11):47–50, 1989.

[122] J. MacQueen et al. Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, volume 1 (14), pages 281–297. Oakland, CA, USA, 1967.

[123] I. Maqsood, M. R. Khan, and A. Abraham. An ensemble of neural networks for weather forecasting. *Neural Computing & Applications*, 13(2):112–122, 2004.

[124] F. J. Martínez-Estudillo, C. Hervás-Martínez, P. A. Gutiérrez, and A. C. Martínez-Estudillo. Evolutionary product-unit neural networks classifiers. *Neurocomputing*, 72(1-3):548–561, 2008.

[125] P. McCullagh. Regression models for ordinal data. *Journal of the Royal Statistical Society: Series B (Methodological)*, 42(2):109–127, 1980.

[126] W. S. McCulloch and W. Pitts. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4):115–133, 1943.

[127] R. C. Miller. *Notes on analysis and severe-storm forecasting procedures of the Air Force Global Weather Central*, volume 200. AWS, 1975.

[128] J. Mo, Y. Wang, and M. Ye. Network traffic outlier detection based on full convolutional network. In *2019 15th International Conference on Computational Intelligence and Security (CIS)*, pages 366–370. IEEE, 2019.

[129] M. Mohri, A. Rostamizadeh, and A. Talwalkar. *Foundations of machine learning*. MIT press, 2018.

[130] M. Moncrieff. A theory of organized steady convection and its transport properties. *Quarterly Journal of the Royal Meteorological Society*, 107(451):29–50, 1981.

[131] D. C. Montgomery, E. A. Peck, and G. G. Vining. *Introduction to linear regression analysis*, volume 821. John Wiley & Sons, 2012.

[132] F. Murtagh and P. Contreras. Algorithms for hierarchical clustering: an overview. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 2(1):86–97, 2012.

[133] S. Nitsure, S. Londhe, and K. Khare. Wave forecasts using wind information and genetic programming. *Ocean Engineering*, 54:61–69, 2012.

[134] Z. Niu, M. Zhou, L. Wang, X. Gao, and G. Hua. Ordinal regression with multiple output cnn for age estimation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4920–4928, 2016.

[135] J. Olivier and C. Aldrich. Dynamic monitoring of grinding circuits by use of global recurrence plots and convolutional neural networks. *Minerals*, 10(11):958, 2020.

[136] I. C. A. Organization. Annex 3 to the Convention on International Civil Aviation: Meteorological Service for International Air Navigation, Eighteenth Edition 2013.

[137] J. Paparrizos and L. Gravano. k-shape: Efficient and accurate clustering of time series. In *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data*, pages 1855–1870. ACM, 2015.

[138] J. Park and I. W. Sandberg. Universal approximation using radial-basis-function networks. *Neural computation*, 3(2):246–257, 1991.

[139] R. Perez, E. Lorenz, S. Pelland, M. Beauharnois, G. Van Knowe, K. Hemker Jr, D. Heinemann, J. Remund, S. C. Müller, W. Traunmüller, et al. Comparison of numerical weather prediction solar irradiance forecasts in the us, canada and europe. *Solar Energy*, 94:305–326, 2013.

[140] M. Pérez-Ortiz, A. M. Durán-Rosal, P. A. Gutiérrez, J. Sánchez-Monedero, A. Nikolaou, F. Fernández-Navarro, and C. Hervás-Martínez. On the use of evolutionary time series analysis for segmenting paleoclimate data. *Neurocomputing*, 326:3–14, 2019.

[141] T. Púčik, P. Groenemeijer, D. Rỳva, and M. Kolář. Proximity soundings of severe and nonsevere thunderstorms in central europe. *Monthly Weather Review*, 143(12):4805–4821, 2015.

[142] J. R. Quinlan. Induction of decision trees. *Machine learning*, 1(1):81–106, 1986.

[143] T. Rakthanmanon and E. Keogh. Fast shapelets: A scalable algorithm for discovering time series shapelets. In *proceedings of the 2013 SIAM International Conference on Data Mining*, pages 668–676. SIAM, 2013.

[144] W. M. Rand. Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical association*, 66(336):846–850, 1971.

[145] S. Rao and S. Mandal. Hindcasting of storm waves using neural networks. *Ocean Engineering*, 32:667–684, 2005.

[146] T. Räsänen and M. Kolehmainen. *Feature-Based Clustering for Electricity Use Time Series Data*, pages 401–412. Springer Berlin Heidelberg, Berlin, Heidelberg, 2009.

[147] C. E. Rasmussen. Gaussian processes in machine learning. In *Summer school on machine learning*, pages 63–71. Springer, 2003.

[148] R. Reed. Pruning algorithms-a survey. *IEEE transactions on Neural Networks*, 4(5):740–747, 1993.

[149] E. Rendón, I. Abundez, A. Arizmendi, and E. M. Quiroz. Internal versus external cluster validation indexes. *International Journal of computers and communications*, 5(1):27–34, 2011.

[150] S. Resino, J. A. Seoane, J. M. Bellón, J. Dorado, F. Martin-Sanchez, E. Álvarez, J. Cosín, J. C. López, G. Lopéz, P. Miralles, et al. An artificial neural network improves the non-invasive diagnosis of significant fibrosis in HIV/HCV coinfected patients. *Journal of Infection*, 62(1):77–86, 2011.

[151] A. Rivero-Juárez, D. Guijo-Rubio, F. Tellez, R. Palacios, D. Merino, J. Macías, J. C. Fernández, P. A. Gutiérrez, A. Rivero, and C. Hervás-Martínez. Using machine learning methods to determine a typology of patients with hiv-hcv infection to be treated with antivirals. *Plos one*, 15(1):e0227188, 2020.

[152] L. Rokach and O. Maimon. Clustering methods. In *Data mining and knowledge discovery handbook*, pages 321–352. Springer, 2005.

[153] P. J. Rousseeuw. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, 20:53–65, 1987.

[154] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. Learning representations by back-propagating errors. *Nature*, 323(6088):533, 1986.

[155] S. Russell and P. Norvig. *Artificial intelligence: a modern approach*. Prentice-Hall, 2020.

[156] M. Şahin, Y. Kaya, M. Uyar, and S. Yıldırım. Application of extreme learning machine for estimating solar radiation from satellite data. *International Journal of Energy Research*, 38(2):205–212, 2014.

[157] A. B. Said and A. Erradi. Deep-gap: A deep learning framework for forecasting crowdsourcing supply-demand gap based on imaging time series and residual learning. *arXiv preprint arXiv:1911.07625*, 2019.

[158] J. Sánchez, L. López, C. Bustos, J. Marcos, and E. García-Ortega. Short-term forecast of thunderstorms in argentina. *Atmospheric research*, 88(1):36–45, 2008.

[159] J. L. Sánchez, J. L. Marcos, J. Dessens, L. López, C. Bustos, and E. García-Ortega. Assessing sounding-derived parameters as storm predictors in different latitudes. *Atmospheric Research*, 93(1-3):446–456, 2009.

[160] P. Satrio, T. M. I. Mahlia, N. Giannetti, K. Saito, et al. Optimization of hvac system energy consumption in a building using artificial neural network and multi-objective genetic algorithm. *Sustainable Energy Technologies and Assessments*, 35:48–57, 2019.

[161] P. Schäfer. The boss is concerned with time series classification in the presence of noise. *Data Mining and Knowledge Discovery*, 29(6):1505–1530, 2015.

[162] M. Schmitt. On the complexity of computing and learning with multiplicative neural networks. *Neural Computation*, 14:241–301, 2002.

[163] M. Schmitt. On the complexity of computing and learning with multiplicative neural networks. *Neural Computation*, 14(2):241–301, 2002.

[164] G. Schwarz et al. Estimating the dimension of a model. *The annals of statistics*, 6(2):461–464, 1978.

[165] Z. Sen. *Solar energy fundamentals and modeling techniques: atmosphere, environment, climate change and renewable energy*. Springer Science & Business Media, 2008.

[166] O. B. Sezer and A. M. Ozbayoglu. Algorithmic financial trading with deep convolutional neural networks: Time series to image conversion approach. *Applied Soft Computing*, 70:525–538, 2018.

[167] F. Shadabi, R. J. Cox, D. Sharma, and N. Petrovsky. A hybrid decision tree–artificial neural networks ensemble approach for kidney transplantation outcomes prediction. In *International Conference on Knowledge-Based and Intelligent Information and Engineering Systems*, pages 116–122. Springer, 2005.

[168] A. Shifaz, C. Pelletier, F. Petitjean, and G. I. Webb. Ts-chief: A scalable and accurate forest algorithm for time series classification. *Data Mining and Knowledge Discovery*, pages 1–34, 2020.

[169] P. Smyth. Clustering sequences with hidden markov models. In *Advances in neural information processing systems*, pages 648–654, 1997.

[170] B.-Y. Sun, J. Li, D. D. Wu, X.-M. Zhang, and W.-B. Li. Kernel discriminant learning for ordinal regression. *IEEE Transactions on Knowledge and Data Engineering*, 22(6):906–910, 2009.

[171] R. S. Sutton and A. G. Barto. *Reinforcement learning: An introduction*. MIT press, 2018.

[172] R. A. Tenenbaum, F. O. Taminato, and V. S. Melo. Fast auralization using radial basis functions type of artificial neural network techniques. *Applied Acoustics*, 157:106993, 2020.

[173] R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288, 1996.

[174] V. Torra, J. Domingo-Ferrer, J. M. Mateo-Sanz, and M. Ng. Regression for ordinal variables without underlying continuous variables. *Information Sciences*, 176(4):465–474, 2006.

[175] T. N. Truong, D. Laureillard, K. Lacombe, H. D. Thi, P. P. T. Hanh, L. T. T. Xuan, N. C. Thi, A. L. Que, V. V. Hai, N. Nagot, et al. High proportion of hiv-hcv coinfected patients with advanced liver fibrosis requiring hepatitis c treatment in haiphong, northern vietnam (anrs 12262). *PloS one*, 11(5):e0153744, 2016.

[176] H.-H. Tu and H.-T. Lin. One-sided support vector regression for multiclass cost-sensitive classification. In *ICML*, 2010.

[177] P. J. Van Laarhoven and E. H. Aarts. Simulated annealing. In *Simulated annealing: Theory and applications*, pages 7–15. Springer, 1987.

[178] H. L. Van Trees. *Detection, estimation, and modulation theory, part I: detection, estimation, and linear modulation theory*. John Wiley & Sons, 2004.

[179] N. X. Vinh, J. Epps, and J. Bailey. Information theoretic measures for clusterings comparison: Variants, properties, normalization and correction for chance. *The Journal of Machine Learning Research*, 11:2837–2854, 2010.

[180] C. Voyant, M. Muselli, C. Paoli, and M.-L. Nivet. Numerical weather prediction (nwp) and hybrid arma/ann model to predict global radiation. *Energy*, 39(1):341–355, 2012.

[181] D. Wang, B. Larder, A. Revell, J. Montaner, R. Harrigan, F. De Wolf, J. Lange, S. Wegner, L. Ruiz, M. J. Pérez-Elías, et al. A comparison of three computational modelling methods for the prediction of virological response to combination HIV therapy. *Artificial Intelligence in Medicine*, 47(1):63–74, 2009.

[182] W. Wang, J. Yang, R. Muntz, et al. Sting: A statistical information grid approach to spatial data mining. In *VLDB*, volume 97, pages 186–195, 1997.

[183] X. Wang, A. Mueen, H. Ding, G. Trajcevski, P. Scheuermann, and E. Keogh. Experimental comparison of representation methods and distance measures for time series data. *Data Mining and Knowledge Discovery*, 26(2):275–309, 2013.

[184] X. Wang, F. Yu, W. Pedrycz, and J. Wang. Hierarchical clustering of unequal-length time series with area-based shape distance. *Soft Computing*, pages 1–13, 2018.

[185] Z. Wang, W. Yan, and T. Oates. Time series classification from scratch with deep neural networks: A strong baseline. In *2017 International joint conference on neural networks (IJCNN)*, pages 1578–1585. IEEE, 2017.

[186] A. S. Weigend. *Time series prediction: forecasting the future and understanding the past*. Routledge, 2018.

[187] WHO. Global hepatitis report. Technical report, World Health Organization, 2017.

[188] L. Wiessing, M. Ferri, B. Grady, M. Kantzanou, I. Sperle, K. J. Cullen, A. Hatzakis, M. Prins, P. Vickerman, J. V. Lazarus, et al. Hepatitis c virus infection epidemiology among people who inject drugs in europe: a systematic review of data for scaling up treatment and prevention. *PloS one*, 9(7):e103345, 2014.

[189] C. Winship and R. D. Mare. Regression models with ordinal variables. *American sociological review*, pages 512–525, 1984.

[190] WMO. Manual on codes, international codes. Technical report, World Meteorological Organization, 2011.

[191] W. Wu, H. Xiong, and S. Shekhar. *Clustering and information retrieval*, volume 11. Springer Science & Business Media, 2013.

[192] D. Xu and Y. Tian. A comprehensive survey of clustering algorithms. *Annals of Data Science*, 2(2):165–193, 2015.

[193] K. Xu, M. Zhou, D. Yang, Y. Ling, K. Liu, T. Bai, Z. Cheng, and J. Li. Application of ordinal logistic regression analysis to identify the determinants of illness severity of covid-19 in china. *Epidemiology & Infection*, 148, 2020.

[194] C.-L. Yang, Z.-X. Chen, and C.-Y. Yang. Sensor classification using convolutional neural network by encoding multivariate time series as two-dimensional colored images. *Sensors*, 20(1):168, 2020.

[195] Y. Yang and K. Chen. Time series clustering via RPCL network ensemble with different representations. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 41(2):190–199, 2011.

[196] Y. Yang and J. Jiang. Adaptive bi-weighting toward automatic initialization and model selection for hmm-based hybrid meta-clustering ensembles. *IEEE transactions on cybernetics*, 49(5):1657–1668, 2018.

[197] X. Yao. Evolving artificial neural networks. *Proceedings of the IEEE*, 87(9):1423–1447, 1999.

[198] L. Ye and E. Keogh. Time series shapelets: a new primitive for data mining. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 947–956, 2009.

[199] L. Ye and E. Keogh. Time series shapelets: a novel technique that allows accurate, interpretable and fast classification. *Data mining and knowledge discovery*, 22(1-2):149–182, 2011.

[200] Z. Zainuddin, N. Mahat, and Y. A. Hassan. Improving the convergence of the backpropagation algorithm using local adaptive techniques. *International Journal of Computer and Information Engineering*, 1(1), 2005.

[201] L. F. Zarzalejo, L. Ramirez, and J. Polo. Artificial intelligence techniques applied to hourly global irradiance estimation from satellite-derived cloud index. *Energy*, 30(9):1685–1697, 2005.

[202] A. Zendehboudi and X. Li. Desiccant-wheel optimization via response surface methodology and multi-objective genetic algorithm. *Energy Conversion and Management*, 174:649–660, 2018.

[203] T. Zhang, R. Ramakrishnan, and M. Livny. Birch: an efficient data clustering method for very large databases. *ACM sigmod record*, 25(2):103–114, 1996.

[204] X. J. Zhu. Semi-supervised learning literature survey. Technical report, University of Wisconsin-Madison Department of Computer Sciences, 2005.

[205] X. Zhuge and X. Zou. Summertime convective initiation nowcasting over southeastern china based on advanced himawari imager observations. *Journal of the Meteorological Society of Japan. Ser. II*, 2018.