Making decisions for frost prediction in agricultural crops in a soft computing framework

J.M. Cadenas[a,*], M.C. Garrido[a], R. Martínez-España[b], M.A. Guillén-Navarro[b]

[a]*Department of Information and Communications Engineering, University of Murcia, Murcia, Spain*
[b]*Department of Computer Engineering, Catholic University of Murcia, Murcia, Spain*

Abstract

Nowadays, there are many areas of daily life that can obtain benefit from technological advances and the large amounts of information stored. One of these areas is agriculture, giving place to precision agriculture. Frosts in crops are among the problems that precision agriculture tries to solve because produce great economic losses to farmers. The problem of early detection of frost is a process that involves a large amount of wheather data. However, the use of these data, both for the classification and regression task, must be carried out in an adequate way to obtain an inference with quality. A preprocessing of them is carried out in order to obtain a dataset grouping attributes that refer to the same measure in a single attribute expressed by a fuzzy value. From these fuzzy time series data we must use techniques for data analysis that are capable of manipulating them. Therefore, first a regression technique based on *k*-nearest neighbors in a Soft Computing framework is proposed that can deal with fuzzy data, and second, this technique and others to classification are used for the early detection of a frost from data obtained from different weather stations in the Region of Murcia (south-east Spain) with the aim of decrease the damages that these frosts can cause in crops. From the models obtained, an interpretation of the provided information is performed and the most relevant set of attributes is obtained for the anticipated prediction of a frost and of the temperature value. Several experiments are carried out on the datasets to obtain the models with the best performance in the prediction validating the results by means of a statistical analysis.

Keywords: Precision agriculture, crop frost, data analysis, fuzzy data, fuzzy classification and regression, fuzzy *k*-nearest neighbors, fuzzy decision tree

1. Introduction

Agriculture plays a very important role in the economy of a country. Precision agriculture, now also called digital agriculture, and the development of technologies applied to it have emerged as fields that use data-intensive approaches to control productivity while minimizing its environmental impact. The data generated, directly and indirectly, in the environment of agricultural crops are provided by different sensors. These sensors obtain information on crop products, soil, climatic conditions, etc. allowing a better understanding of the environment and the operation itself. All this will lead to faster, more efficient and effective decision-making systems, (Liakos et al., 2018).

*Principal corresponding author
Email address: `jcadenas@um.es` (J.M. Cadenas)

Among many other factors, frosts are meteorological phenomena that cause relatively frequent major damage to the proper development of an agricultural crop. Sometimes these crop frosts are severe and constitute a real potential threat (García-Pedraza and García-Vega, 1991). Frost is influenced by climatic conditions and, among other, by local factors such as topography and terrain orientation, soil types, etc. The study carried out in (García-Pedraza and García-Vega, 1991) highlights that local factors (type of soil, orientation of slopes to the North (shady) or to the South (sunny)) influence frost. Also the damage that a frost can cause to crops depends on the intensity and duration of the frost. In (García-Pedraza and García-Vega, 1991; Lee et al., 2016) frosts are classify as "Advective frosts (black frosts)", "Evaporation frosts" and "Irradiation frosts". In Spain the latter are the most frequent, covering the period autumn-winter-spring. Late spring frosts are usually the most dangerous for crops.

On the "FreshPlaza" website (an independent source of news for companies operating on a global scale in the agricultural sector, particularly fruit and vegetables) there are articles and news related to the sector. Among others, news related to economic losses caused by frost can be found, (Fresh Plaza, 2018, 2019).

In (Fresh Plaza, 2018), the concern in southern Europe about frost damage to early stone fruit is again described for 2018. In March spring begins in the weather calendar however, the cold continues to occur. The strong east wind causes temperatures to fall well below freezing in many parts of Europe. Cold and snow disrupt life in several European countries with serious market consequences. In particular, the consequences in southern and eastern Europe (such as Spain, Italy, northern France, western Hungary, Croatia, etc.) caused by these inclement weather conditions have affected some types of fruit trees that were already in bloom or about to bloom. Extreme cold damages the early harvest of stone fruit in these regions.

In (Martínez-Núñez et al., 2015) a study of frost and cold hours in Spain is presented during the period 2002-2012 from November to April. The authors use different threshold temperatures for their study. Among them, temperatures below 7^{o}C are considered as a fixed value in the determine the hours of cold.

The map of frosts in Spain presented in that study is very interesting, analysing both the agricultural and economic repercussions of cold hours. They present maps describing the number of frost days/year, frost probabilities/year, dates of both the first and last frosts, etc. Among the areas described, the Region of Murcia stands out for its agricultural features. The Region of Murcia (Southeast Spain) suffers during the winter season and early spring of various stages of frost which causes damage to the flower and/or fruit. These frosts produce considerable losses to the sector, and there is a need for reliable warning systems to prevent such damage in some cases (the loss caused by frost in March 2019 is estimated at 14.7 million euros in several areas) (Fresh Plaza, 2019).

In (Snyder et al., 2010) there is also an analysis for frost in crops. The analysis of frost prediction and monitoring, and of the various passive and active methods for frost protection is very interesting. In this paper, the authors highlight the value of effective prediction involving complex analysis of decision making, but importantly, accurate prediction would allow farmers to prepare against them and potentially reduce the damage they can cause.

In conclusion, frosts produce significant losses to the agricultural sector which need to develop and have effective strategies and reliable warning systems to prevent or reduce damage to crops, loss of fruit quality and/or production losses. Thus, this manuscript focuses on the development of a decision system based on fuzzy models for the frost prediction with the aim of informing and alerting. With this information farmers can activate anti-frost techniques and thus avoid losing their crops. To this purpose, the objectives in this manuscript are as follows:

- First of all we focus on the information used for the modeling and prediction of frosts. The information provided by institutional systems related to weather can be used in a most appropriate way. For example, systems provide periodic information such as "Minimum, Mean and Maximum relative humidity", "Mean and Maximum wind speed ", etc. In general, models use this information as variables or attributes independent of the instances necessary for the frost prediction process. For example, the minimum, mean and maximum of relative humidity refer to the same attribute and not to three independent ones. For this reason, we propose the use of fuzzy information to represent the values of the same measure, interpreting those values more appropriately. In addition, with this representation, the instances are defined by fewer attributes which facilitates the manipulation and subsequent interpretation. Once the information has been preprocessed and the datasets have been obtained with the fuzzy information, the design and use of appropriate techniques for its manipulation are proposed.

- Next, we focus on the study and characterization of the relationships between the weather attributes to predict frosts, and on the use and/or design of adequate techniques to manipulate the datasets and obtain a good performance in the prediction. Therefore, the objective is to find relationships between the weather information obtained several hours before with the prediction that a frost will occur some time later. This helps us to find traces that can characterize these relationships and to build predictive models with good behavior. The final purpose is to build a decision system that helps predict frost. In addition, designed models must be simple using fewer attributes without loss of accuracy.

- Finally, a new approach based on k-nearest neighbors is proposed. This approach is able to manage imprecise information implicitly in data to tackle the regression task. This technique is applied to the problem presented in this manuscript, that is, to predict the minimum temperature considering several weather attributes.

In summary, the proposed decision support system provides the farmer with both qualitative information (whether there will be frost or not) using a classification technique and quantitative information (minimum temperature in the next hour) using a proposed new regression technique. In addition, the characterisation of the most important weather attributes indicates the measuring instruments needed to be able to obtain this information locally in each plot, saving costs since the number of these instruments is smaller.

Thus, this manuscript is organized as follows. In Section 2 a background on the automatic systems applied in agriculture is provided, paying special attention to frost prediction systems. In Section 3, a novel approach based on k-nearest neighbours for the regression task that is capable of supporting fuzzy information is presented. In Section 4 the datasets, techniques and methods used for the study of early prediction of frost in crops in south-east Spain are described. Specifically, in Subsection 4.1, the study areas where collecting weather information are presented. In addition, from the available information, a preprocessing is carried out with special attention to the transformation into fuzzy values. In Subsection 4.2, the techniques used in the experiments, and their configurations are indicated. In addition, in Subsection 4.3, the different measures and statistical tests used to evaluate the results are commented. In Section 5, all experiments aimed at answering the various questions raised are developed. For the result evaluation and conclusions obtained, statistical tests are applied. Finally, in Section 6, a decision system with its component elements is described and the conclusions are presented in Section 7.

2. Background

Automatic systems built for agriculture are often used for decision making at different levels (Yelapure and Kulkarni, 2012): a) in the operational level the system is often used to provide advice to producers; and b) in the planning level the system is used to predict the plantation needs.

Since the 1970s, decision support systems begin to be built and applied in the agricultural area. Since, these systems have been applied, among others, to the protection breeding, poultry raising, installation horticulture management, aquaculture activity, plant crops management as well as economical decision making. In (Liakos et al., 2018; Yelapure and Kulkarni, 2012), a very good compilation of these systems can be find. Specifically, a review of different expert system and fuzzy expert systems are detailed.

In this manuscript we focus on precision agriculture based on computational learning. More specifically, our proposal focus on the crop management activity of predicting weather conditions, trying to predict frost as one of the problems that affects the quality of producers and the economy of the farmer. Therefore, to focus on the prediction of weather conditions, next this agricultural activity is analyzed.

For the prediction and construction of alert systems for frosts in agriculture, different techniques have been proposed such as neural networks, self-organizing maps, decision trees, support vector machines (SVM), rule based systems, etc, using data provided by automatic systems witch obtain weather conditions given by time series. Then, some studies are analyzed emphasizing the data and attributes used and the approaches presented.

In (Lee et al., 2016) two models for frost prediction or warnings in the spring of Korea are developed using a decision tree and logistic regression. These models were compared using data obtained from 1973 to 2004 from six weather stations and seven attributes. The attributes used were the minimum temperature, grass minimum temperature, mean relative humidity, dew point, minimum relative humidity, wind and cloud. The conclusion reached indicates that the decision tree may be more useful for the frost alert system.

In (Fuentes et al., 2018) a neural network model is presented to predict the minimum air temperature of the next day. For the model construction the meteorological data are used such as wind direction and speed, relative humidity, air temperature, precipitation and radiation. The model was validated with 10 weather stations in central Chile for 8 years (from 2010 to 2017). The mean square error in the prediction of the minimum temperature was 2.99^{o}C; and a total average accuracy in the frost detection of 98% (86% sensitivity). The authors highlight that differences and errors in the frost detection can be attributed to factors mainly associated with the accuracy of the weather stations, local climatic and geographical conditions, and the parameter number in the construction of the models.

In (Yu et al., 2016) a model based on least squares SVM is proposed. The model parameters are optimized by particle swarm for the anticipated temperature prediction in the Chinese solar greenhouse. The model uses data on indoor and outdoor temperature, indoor air humidity, outdoor solar radiation, wind speed, and soil temperature and humidity of two greenhouses. The obtained conclusions indicate that the proposed model is accurate and therefore useful and effective in predicting the temperatures of the Chinese solar greenhouse.

In (Smith et al., 2009) a neural network for predicting air temperature based on near real-time data is applied. The learning and subsequent validation of the models was carried out based on meteorological data from the southeastern United States (Georgia). The models used current values and previous observations of relative humidity, wind speed, temperature and solar radiation. An improvement in the prediction accuracy could be observed when the rain attribute was used. The time values were coded as four values using triangular membership functions in the range [0,1] (midnight, morning, noon and evening). Similarly, day of the year values were coded using four triangular membership functions to represent seasonality. In total, data with 258 attributes were used. The neural network provided predictions for a test dataset of 0.516^oC of mean absolute error in the one hour horizon and of 1.873^oC of mean absolute error in the twelve hour horizon.

In (Efendi et al., 2017) a procedure based on a fuzzy random auto-regression time series is proposed to model the variability and temperature trend. A relevant topic of this work is the transformation of minimum-maximum data into triangular fuzzy numbers. The support of these fuzzy numbers is defined by the minimum and maximum values with membership degree 0 and the midpoint with membership degree 1.

A study to predict low temperatures is presented in (Guillén-Navarro et al., 2018). This initial study uses C4.5 decision tree and M5P rule techniques (implementations provided by Weka package) to classify possible frosts. For this, the authors use three datasets obtaining a classification error of 12%. Later the authors extend this study in (Guillén-Navarro et al., 2019) where they use ten datasets to predict temperatures from different weather attributes using those techniques. From the experiments, they obtain a root mean square error less than 0.6^oC.

3. kNN-RegID: A technique for regression from imprecise data based on *k*-nearest neighbors

The *k*-nearest neighbors technique is widely used in data mining. The technique can be applied to high-dimensionality problems where the attributes describing the instances can be both nominal and numerical. In addition, *k*-nearest neighbors has been successfully applied in solving both the regression and classification tasks in a variety of fields. In literature there are proposals for studies based on *k*-nearest neighbors where some of them are framed in the Fuzzy Set Theory to incorporate imprecision therein. On the one hand, some proposals incorporate imprecision in the class attribute (Keller et al., 1985). From that imprecision, several works were developed that focus on obtaining the final membership degrees of the different classes (Han and Kim, 1999). On the other hand, there are other works focus on the calculation of distances (Mitchell and Schaefer, 2001).

In this manuscript, a regression technique based on *k*-nearest neighbors that supports datasets with imprecise values is proposed. Specifically, the *k*-nearest neighbors technique is used in order to deal with data whose attribute values are defined from membership functions. This technique is denoted by kNN-RegID and is described below.

3.1. Description of the kNN-RegID technique

The kNN-RegID technique allows the imputation of missing values for numerical domain attributes from imprecise data. These imprecise input data can be both nominal and numerical. In order to homogenize the structure of the input data, the technique works with attribute values described by tuples. Each tuple is formed by elements of the form

$\{\mu_i/v_i\}$ where v_i is a domain value of the attribute i and $\mu_i \in [0,1]$ the membership degrees of those values. This representation by means of tuples allows us to formulate:

- Trapezoidal and triangular fuzzy values, missing values, interval values and crisp values for numerical attributes. The values of these attributes are represented by tuples in the form $[\mu_1/v_1, \mu_2/v_2, \mu_3/v_3, \mu_4/v_4]$.
- Fuzzy subsets, missing values and crisp values for nominal attributes. In these cases, the values of these attributes are represented by $\{\mu_1/v_1, \mu_2/v_2, \mu_3/v_3, \ldots\}$, with as many μ/v pairs as necessary to indicate the attribute value.

Therefore, the input datasets to the kNN-RegID technique are composed of instances with attributes defined as indicated above. In general, each instance $\mathbf{q} = (q_1, q_2, \ldots, q_{n-1}, q_n)$ is made up of n attributes where the attributes q_i, $i = 1, \ldots, n-1$ are described by tuples and the attribute to estimate q_n is always given by a crisp value described by a tuple $\{1/v_i\}$ or only as v_i. The operation of the kNN-RegID technique is defined in the Algorithm 1.

Algorithm 1: kNN-RegID - k-Nearest Neighbors for regression from imprecise data

Input Dataset D, instance to infer $\mathbf{q}$, value k ($1 \leq k \leq |D|$), value U_E ($U_E \in [0,1]$), similarity (dissimilarity) function $S(\cdot,\cdot)$, entropy function $f_e(\cdot)$

Obtain the set of k instances of D (denoted by K_S) more/less similar/dissimilar to $\mathbf{q}$ according to the $\mathrm{S}(\cdot,\cdot)$ function.

Calculate $f_e(\mathbf{q}')$, $\forall \mathbf{q}' \in K_S$ and $E_{K_S} = \frac{\sum_{\mathbf{q}' \in K_S} f_e(\mathbf{q}')}{|K_S|}$

if ($E_{K_S} \leq U_E$ **then**

  if similarity **then** $q_n = \sum_{\mathbf{q}' \in K_S} (1 - f_e(\mathbf{q}')) \cdot S(\mathbf{q}, \mathbf{q}') \cdot q'_n$

  if dissimilarity **then** $q_n = \sum_{\mathbf{q}' \in K_S} (1 - f_e(\mathbf{q}')) \cdot (1 - S(\mathbf{q}, \mathbf{q}')) \cdot q'_n$

  Output q_n

else

  Output Estimation is not performed

end if

In general, in the technique k-nearest neighbors plays a very important role the function used to obtain the set of k-nearest neighbors to a given instance. Since kNN-RegID is going to handle imprecise attribute values, the function is defined for this type of values. When the set K_S is formed by the k examples more similar to a given one it is considered fuzzy similarity functions and when the set K_S is composed by the k examples with less dissimilarity value it is considered a fuzzy dissimilarity function. In general, the function used to obtain K_S is defined as $S(\mathbf{q}, \mathbf{q}') = \sum_{i=1}^{n-1} \frac{S_t(q_i, q'_i)}{n-1}$ where $t = 1$ indicates a similarity/dissimilarity function defined over imprecise numerical attributes and $t = 2$ indicates a similarity/dissimilarity function defined for imprecise nominal attributes.

When working with imprecise data, the kNN-RegID technique uses a measure of the imprecision of the different attribute values. With this measurement, the technique can take into account that those less imprecise instances have a greater relevance in the estimates made. Fuzziness measures or fuzzy entropy functions ($f_e(\cdot)$) are used because they allow to measure the indefiniteness described by the memberships function of fuzzy sets. In general, the imprecision of an instance is measured as $f_e(\mathbf{q}') = \frac{\sum_{i=1}^{n-1} f_e(q'_i)}{n-1}$. In addition, the technique incorporates the parameter $U_E \in [0,1]$ that establishes a limit of imprecision for the set of neighbors from which an estimation is performed. If the imprecision of the set K_S, denoted by E_{K_S}, exceeds this threshold, the estimation is not carried out.

The q_n value is obtained by averaging the q'_n values of the neighbors in K_S. In this average, the value of each neighbor is applied weighted by two values: the value of similarity/dissimilarity of each $\mathbf{q}'$ with $\mathbf{q}$ and its imprecision $f_e(\mathbf{q}')$, in order that the more-similar/less-dissimilar to $\mathbf{q}$ and less is its imprecision, the greater is its contribution in the final result. This is reflected in the Algorithm 1 with the factors $1 - f_e(\mathbf{q}')$, and $S(\mathbf{q}, \mathbf{q}')$ or $1 - S(\mathbf{q}, \mathbf{q}')$ depending on whether the $S(\cdot, \cdot)$ function is similarity/dissimilarity, respectively.

4. Datasets, techniques, evaluation and validation of experiments.

4.1. Information collection and data preparation

4.1.1. Study areas in south-east Spain

The Murcian Institute of Agricultural and Food Research and Development, among other things, collects information on the weatherology of different areas through the Agricultural Information Service of Murcia Region (SIAM, http://siam.imida.es). The covered areas, and municipalities that integrates, are the following:

- Altiplano – Yecla, Jumilla, Abanilla and Fortuna.
- Noroeste – Moratalla, Caravaca de la Cruz, Cehegín and Bullas.
- Río Mula – Mula, Pliego, Albudeite and Campos del Río.
- Vega del Segura – Murcia, Beniel, Santomera, Alcantarilla, Molina de Segura, Torres de Cotillas, Alguazas, Ceutí, Lorquí, Archena, Ulea, Villanueva del Segura, Ojós, Ricote, Blanca, Abarán, Cieza and Calasparra.
- Valle del Guadalentín – Lorca, Puerto Lumbreras, Águilas, Mazarrón, Totana, Aledo, Alhama de Murcia and Librilla.
- Campo de Cartagena – Fuente Álamo, Cartagena, Unión (La), Torre Pacheco, San Javier, San Pedro del Pinatar and Alcázares (Los).

In (Martínez-Núñez et al., 2015), it is described that in the areas of the Altiplano, Noroeste, Río Mula and north of the Vega del Segura, the average percentage of frost occurring is between 80% and 100%. And the average date of the first frost starts from 1 to 15 December, and the last frost is until 31 March. For our study we have selected, from these regions, 5 weather stations that are shown in the Table 1. These stations are surrounded by stone fruit crops and therefore the results obtained are interesting to prevent frost on these crops.

4.1.2. Initial collection of information. Initial data

Each station is equipped with the following sensors and ephemeris: weather vane, radiometer, rain gauge, datalogger and thermo-hygrometer. The information collected corresponds to 7 years (2012-2018). The initial time series data obtained from SIAM sensors correspond to values obtained every 5 minutes. These are grouped 12 by 12 to show only values for each hour. For this reason, some of the measurements show the minimum, average and maximum values for each hour. The type of information obtained is shown in the Table 2.

Table 1: Description of the weather stations in study and places by hours of cold ($\leq 7^o$) per year of the Region of Murcia, Spain.

Caravaca – Noroeste Station CR12 – Altitude 869m Coordinate: 38^o 2’ 38.24” – 1^o 58’ 48,67”
Cehegín – Noroeste Station CR32 – Altitude 433m Coordinate: 38^o 6’ 39” – 1^o 19’ 27.58”
Calasparra – Vega del Segura Station CI52 – Altitude 275m Coordinate: 38^o 15’ 12.59” – 1^o 41’ 41,89”
Jumilla – Altiplano Station JU71 – Altitude 401m Coordinate: 38^o 23’ 40,01” – 1^o 14’ 21,58”
Jumilla – Altiplano Station JU81 – Altitude 341m Coordinate: 38^o 19’ 11,3” – 1^o 19’ 27,58”

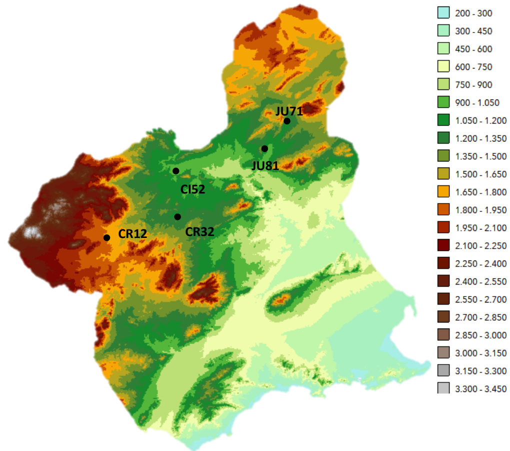

Table 2: Information collected every hour for each station.

Weather station code	Date of data reading
Hour of data reading	Min. relative humidity (%)
Mean relative humidity (%)	Max. relative humidity (%)
Mean radiation (W/m^2)	Max. radiation (W/m^2)
Accumulated radiation (W/m^2)	Mean wind speed (m/s)
Max. wind speed (m/s)	Mean wind direction (^oC)
Rainfall (mm)	Dew point (^oC)
Vapor pressure deficit (kPa)	Min. temperature (^oC)
Mean temperature (^oC)	Max. temperature (^oC)

4.1.3. Preprocessing of information

After the collection of information from the five weather stations, a data preprocessing process is carried out to generate the time series datasets. These datasets are used to obtain the classification and regression models of the decision support system.

Building the time series dataset to regression. From the initial data, the following preprocessing of the information is performed to construct the time series dataset to regression.

1. Attribute construction:

   In the Table 3 the constructed attributes are shown.

Table 3: Extended Attribute Description.

Attrib.	Description	Attrib.	Description
Stat	Wheater station code	Date	Date of data reading
H	Hour of data reading	RH_f	Relative humidity
R_f	Radiation	AR	Accumulated radiation
WS_f	Wind speed	WD	Mean wind direction
RF	Rainfall	DE	Dew point
VPD	Vapor pressure deficit	T_f	Temperature

   These attributes include RH_f, R_f, WS_f and T_f which take fuzzy values. These fuzzy attributes are constructed as follows:

- From the weather attributes that have minimun, mean and maximun values (specifically, the "relative humidity" and "temperature" measures) fuzzy values are constructed. These values are trapezoidal fuzzy numbers (Figure 1a and 1b) given by $[\mu_1/v_1, \mu_2/v_2, \mu_3/v_3, \mu_4/v_4]$ where $\frac{v_2+v_3}{2} = mean_{value}$ with $\mu_2 = \mu_3 = 1$; if $(max_{value} - mean_{value} \leq mean_{value} - min_{value})$ then $v_4 = max_{value}$ with $\mu_4 = 0$ and $v_1 = min_{value}$ with $\mu_1 = \frac{x-v_1}{v_2-v_1}$, else $v_1 = min_{value}$ with $\mu_1 = 0$ and $v_4 = max_{value}$ with $\mu_4 = \frac{v_4-x}{v_4-v_3}$. If the min_{value} or max_{value} are lower or higher, respectively, than the minimum or maximum global values, they will be defined by the global minimum and maximum.

- From the weather attributes that have mean and maximum values (specifically, the "radiation" and "wind speed" measures) fuzzy values are constructed. These values are trapezoidal fuzzy numbers (Figure 1c) given by $[\mu_1/v_1, \mu_2/v_2, \mu_3/v_3, \mu_4/v_4]$ where $\frac{v_2+v_3}{2} = mean_{value}$, $v_4 = max_{value}$ with $\mu_4 = 0$ and $v_1 = 2 \times mean_{value} - max_{value}$ with $\mu_1 = \frac{x-v_1}{v_2-v_1}$. If the min_{value} is less than the global minimum value, the value will be the global minimum value.

- The values of the measures corresponding to the missing values have been maintained. In these cases, the constructed attributes will contain trapezoidal fuzzy numbers $[\mu_1/v_1, \mu_2/v_2, \mu_3/v_3, \mu_4/v_4]$ where $v_1 = v_2 = min_{global}$ with $\mu_1 = \mu_2 = 1$ and $v_3 = v_4 = max_{global}$ with $\mu_3 = \mu_4 = 1$ (Figure 1d).

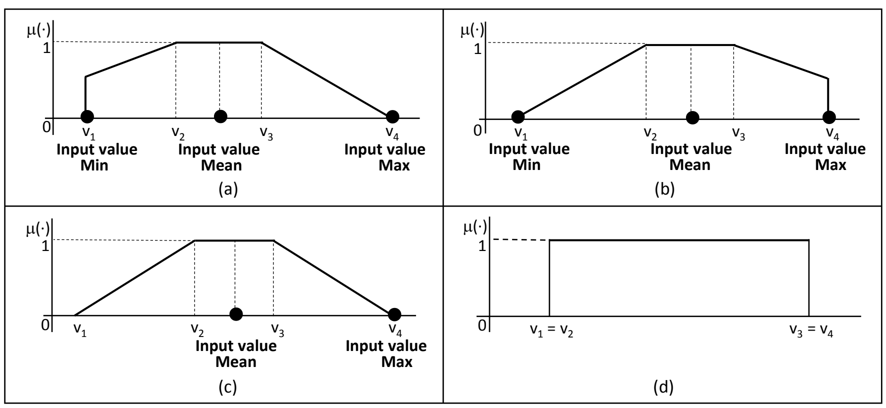

Figure 1: Representations of fuzzy values constructed from the known values minimun, mean, maximun and missing of a measure.

2. Instance construction. From the attributes in Table 3, the instances are formed as follows: The first nine attributes correspond to the values of RH_f, R_f, AR, WS_f, WD, RF, DE, VPD and T_f taken in the hour $H-1$; the following nine attributes correspond to the values of the same previous attributes but taken at H and the last attribute (attribute to be inferred) corresponds to the value of the minimum temperature during the hour $H+1$. Therefore, each instance will be composed of 19 attributes as shown in the Table 4.

Table 4: Constructed Instance.

H-1									H									H+1
RH_f	R_f	AR	WS_f	WD	RF	DE	VPD	T_f	RH_f	R_f	AR	WS_f	WD	RF	DE	VPD	T_f	T_{MIN}

3. The instances are filtered selecting only those that have the value $T_{MIN} \leq 7^oC$.

Building the time series dataset for classification. This dataset is constructed from the dataset already built for regression. Now, a new attribute is built, which it is denoted by CLASS_{FnF}, with two possible values: Frost or NoFrost. For each instance, and using the value $v_{T_{MIN}}$ of the T_{MIN} attribute, the value for CLASS_{FnF} is defined as:

- If $v_{T_{MIN}} > 0$ then CLASS_{FnF}=NoFrost, else CLASS_{FnF}=Frost

From dataset for regression, removing its T_{MIN} attribute and adding the new one CLASS_{FnF}, the dataset for classification is obtained.

4.1.4. Features of constructed datasets

Through the preprocessing and construction process, 10 datasets are obtained (5 for classification and 5 for regression) corresponding to the 5 stations (two datasets per station). The features of the 10 datasets are shown in Table 5. For each station, it is shown the number of instances, the total number attributes, the number of numerical and nominal attributes without taking into account the attribute to be inferred (that in classification will be nominal and in regression will be numerical), the number of classes for the attribute to be inferred in classification, the number of missing values, the number of fuzzy values and the number of instances with imprecises values (missing and/or fuzzy values). The last three values are expressed in percentage.

Table 5: Features of the datasets for the different stations, where $|D|$ - number of instances, Attr - number of attributes, Nu and No - number of numerical and nominal attributes, respectively, I - number of classes, NMV - number of missing values, NFV - number of fuzzy values and NI-MFV - number of instances with imprecises values.

Acron	$\|D\|$	Attr	Nu	No	I	%NMV	%NFV	%NI-MFV
CR12	15185	22	18	3	2	0.020	44.4	100
CR32	11855	22	18	3	2	0.010	44.4	100
CI52	9094	22	18	3	2	0.023	44.4	100
JU71	8470	22	18	3	2	0.001	44.4	100
JU81	8296	22	18	3	2	0.010	44.4	100

4.2. Techniques used and their configuration

The technique proposed in Section 3 is used for the regression process and to infer the temperature value. In addition, two classification techniques proposed in the literature are also applied. Specifically, the decision tree "FDT_{ii}" (Cadenas et al., 2012) and *k*-nearest neighbors "kNN_{imp}" (Cadenas et al., 2018) techniques, which support imprecise data, are applied.

First, the techniques used for classification FDT_{ii} and kNN_{imp} are briefly described. Then, the configurations used in the three techniques in the different experiments are detailed.

4.2.1. Brief description of techniques FDT_{ii} and kNN_{imp} used for classification

FDT_{ii}: a fuzzy decision tree for classification. The FDT_{ii} technique (Cadenas et al., 2012) is a fuzzy decision tree that can classify instances that contain imprecise values in their input attributes. This technique needs a fuzzy/crisp discretization of the numerical attributes that are part of the problem. In order to handle imprecise values and to be able to obtain in each node the best input attribute to split it, the technique uses a similarity function to measure how similar a fuzzy value is with the labels of a partition.

Using this similarity function each instance descends for each branch of the tree with a weight. The class value provided for the tree to classify a new instance **q** is determined by the majority class of the tree leaf reached for **q** with greater weight. For more details on this technique, refer to the paper (Cadenas et al., 2012).

kNN_{imp}: k-nearest neighbors for classification. The $\mathrm{kNN_{imp}}$ technique (Cadenas et al., 2018) is based on k-nearest neighbors and allows to classify instances from imprecise data. In addition, it provides class values that may be also expressed with imprecise values when there is no class value clearly highlighted from the others in the resulting classification. The vague concept "clearly highlighted" can be defined by the external parameter U_D of the technique.

$\mathrm{kNN_{imp}}$ technique has several aggregation methods of the information provided by the k-nearest neighbors to decide the classification (Cadenas et al., 2018). These methods provide high flexibility to the technique, allowing choose them according to the classification problem. Two of them are used in this work: WM_{sv} and WM_{cv}.

The $\mathrm{kNN_{imp}}$ technique with the WM_{sv} method assigns to q a crisp subset $\{1/\omega_h\}$ composed of ω_h defined as $h = \arg \max\limits_{c=1,\ldots,I} \sum_{j=1}^{k} \mu_{q_n^j}(\omega_c) \cdot S(\mathbf{q^j},\mathbf{q}) \cdot (1 - f_e(\mathbf{q^j}))$ where $(\mu_{q_n^j}(\omega_c) \cdot S(\mathbf{q^j},\mathbf{q}) \cdot (1 - f_e(\mathbf{q^j}))$ returns the score assigned by a neighbor q^j to each class value $c = 1,\ldots,I$, determined by the weight of that value in q_n^j $(\mu_{q_n^j}(\omega_c))$, by the weight of $\mathbf{q^j}$ according to its nearness $S(\mathbf{q^j},\mathbf{q})$ when $S(\cdot,\cdot)$ is a similarity function (or $1 - S(\mathbf{q^j},\mathbf{q})$ when $S(\cdot,\cdot)$ is a dissimilarity function) and by the weight of $\mathbf{q^j}$ according to its imprecision $((1 - f_e(\mathbf{q^j})))$.

While the $\mathrm{kNN_{imp}}$ technique with the WM_{cv} method returns as output a fuzzy subset $\{\mu(\omega_t)/\omega_t\}$ composed of the class values ω_t with $\mu(\omega_t) > 0$ defined as $\mu(\omega_t) = \frac{\sum_{j=1}^{k} \mu_{q_n^j}(\omega_t) \cdot S(\mathbf{q^j},\mathbf{q}) \cdot (1 - f_e(\mathbf{q^j}))}{\sum_{j=1}^{k} \sum_{c=1}^{I} \mu_{q_n^j}(\omega_c) \cdot S(\mathbf{q^j},\mathbf{q}) \cdot (1 - f_e(\mathbf{q^j}))}$

For further details on this technique, refer to the paper (Cadenas et al., 2018).

4.2.2. *Configuration of techniques for experiments*

In this subsection, the different values assigned to the parameters of the used techniques are detailed.

To kNN-RegID technique: A value of U_E=1 has been used. With respect to the value of k, in each test the value of k that obtains the best results has been selected and indicated in the test. In addition, the following functions have been used:

- For numerical attributes, S_1 function is the similarity function defined in (Dengfeng and Chuntian, 2002):

$$S_1(q_i, q_i') = 1 - \frac{\frac{|v_1 - v_1'| + |v_2 - v_2'| + |v_3 - v_3'| + |v_4 - v_4'|}{4}}{max_i - min_i} \tag{1}$$

  where q_i and q_i' are values for the numerical attribute i expressed by means of tuples.

- For nominal attributes, S_2 function is the similarity function based on the Minkowski r-metric (Beckenbach and Bellman, 1961):

$$S_2(q_i, q_i') = 1 - \frac{\sum_{j=1}^{|\Omega_i|} |\mu_{q_i}(v_j) - \mu_{q_i'}(v_j)|}{|\Omega_i|} \tag{2}$$

  where $|\Omega_i|$ is the cardinal of the domain of the nominal attribute i and q_i and q_i' are values for the nominal attribute i expressed by means of the tuples commented previously.

- The entropy function $f_e(\cdot)$ used is the one defined by Luca and Termini in (de Luca and Termini, 1972):

$$f_e(q_i) = \begin{cases} \frac{1}{|\Omega_{q_i}|} \sum_{v \in q_i} \mu_{q_i}(v) & \text{if } q_i \text{ has an imprecise value} \\ 0 & \text{other case} \end{cases} \quad (3)$$

To FDT$_{\text{ii}}$ technique: The stop condition of the fuzzy decision tree is double: if a node contains at least 95% of instances of the same class, the node is labeled as a leaf; if the number of instances of the node is smaller than 2.5% of the dataset size (M), the node is labeled as a leaf node.

To kNN$_{\text{imp}}$ technique with the WM$_{sv}$ and WM$_{cv}$ aggregation methods: The $S(\cdot,\cdot)$ and $f_e(\cdot)$ functions are those described in (1), (2) and (3). U$_D$=0.2 has been used. Regarding to the k value, in each test the value that obtains the best results has been selected, indicating this value in the several tables.

4.3. Measures used to assess results. Validating experiments

For all experiments, a 5-fold cross-validation is performed. In addition, an evaluation with complete training dataset is carried out.

For the classification task:

- Given a confusion matrix obtained by a classification technique, the different used measures are as follows: ACC (accuracy) $= \frac{TP+FN}{TP+TN+FP+FN}$; TPR (sensitivity) $= \frac{TP}{TP+FN}$; TNR (specificity) $= \frac{TN}{TN+FP}$; and F-score (harmonic mean of sensitivity and precision) $= \frac{2TP}{2TP+FP+FN}$, where TP: forecasted frost, FP: forecasted false frost, FN: frost not predicted, TN: no frost. The values of these measurements range from 0 to 1, and when the values are close to 1, they indicate good model results.

- To determine the economic impact of the decisions taken, the cost matrix is used. This matrix indicates the estimated cost to protect crops and the estimated cost of losing the crop to frost. The value of the classification cost is estimated by the expression E=$\alpha \cdot FP + \beta \cdot FN$ where α is the cost of protecting the crop and β is the estimated cost for the loss of the crop. As a particular case, for the experiments $\alpha = 1$ and $\beta = 10$ are used.

- Due to the possibility that the kNN$_{imp}$ technique obtains an imprecise classification, the confusion matrix must be extended (Table 6).

Table 6: Extended confusion matrix.

	Frost	NoFrost	{Frost,NoFrost}	← classified as
Frost	TP	FN	TP_FN	
NoFrost	FP	TN	FP_TN	

As shown the extended confusion matrix, a new value appears in the prediction {Frost,NoFrost} for CLASS$_{FnF}$. This new value appears when the technique has obtained for each class value its membership degree $\{\mu_1/\text{Frost}\}$ and $\{\mu_2/\text{NoFrost}\}$ verifying that $\frac{|\mu_1-\mu_2|}{max(\mu_1,\mu_2)} \leq U_D$. In this situation, the technique does not know which class value to choose and classify the instance with {Frost,NoFrost}. With these new values, and in a real situation, the user is the one who must make a decision about it. For the experiments, we make a conservative decision (the lowest cost class), deciding that:

  - If the inferred class is imprecise {Frost,NoFrost}, we will decide that the instance is of the Frost class. This decision reflects that we believe that a frost may occur and we apply the actions to try to reduce its impact. Therefore, the classifications {Frost,NoFrost} will be taken as Frost and in the extended confusion matrix we will have that TP_FN will be successes and FP_TN failures.

The different versions of $\mathrm{kNN_{imp}}$ technique will be denoted as $\mathrm{kNN^{C}_{imp}}$ and $\mathrm{kNN^{F}_{imp}}$, depending on whether it is used the aggregation method WM_{sv} that produces crisp outputs or the WM_{cv} method that produces fuzzy outputs, respectively. To denote the conservative decision from the technique $\mathrm{kNN^{F}_{imp}}$, $\mathrm{kNN^{Fcons}_{imp}}$ will be used.

For the regression task:

- MSE(X,Y) (mean square error)= $\frac{1}{n}\sum_{i=1}^{n}(X_i-Y_i)^2$ and MAE(X,Y) (mean absolute error) = $\frac{1}{n}\sum_{i=1}^{n}|X_i-Y_i|$.
- CC(X,Y) (Pearson correlation coefficient)= $\frac{Cov(X,Y)}{\sigma_X\sigma_Y}$ where Cov is the covariance and σ is the standard deviation.
- R^2(X,Y) (determination coefficient)= $1-\frac{\sum_{i=1}^{n}(X_i-Y_i)^2}{\sum_{i=1}^{n}(X_i-\overline{X})^2}$ where $\overline{X}$ is the mean of the real data.

Finally, and to check which technique has the best overall behavior, a statistical analysis is performed. To perform this analysis, non-parametric statistical tests are applied, (García et al., 2010). The Wilcoxon signed-rank test is used to perform pairwise comparison between two techniques. The Friedman test is used to analyze if there are significant differences in the behavior of the different techniques. If the null hypothesis is rejected, this indicates that there are differences between the techniques, although it does not indicate which technique or techniques are better. In order to identify which of the techniques has the best behaviour, a post hoc test is carried out. As a post hoc test, Holm's procedure is used. For this statistical analysis, the R package (Ihaka and Gentleman, 1996) is used.

5. Results and discussions

In this section, and in order to answer the different questions posed, a series of experiments with the different techniques and datasets are going to be conducted.

First, the relationship between temperature and the other weather attributes (of the current time and the previous one) is studied in order to predict a frost, and to infer the temperature value for the next hour. To achieve this target, the $\mathrm{FDT_{ii}}$ technique is applied to the dataset for classification using a 5-fold cross-validation. Once the model parameters have been validated using cross-validation, the model is obtained using the complete dataset and the information provided is analyzed in order to select the most relevant attributes for the decision system.

Secondly, an adjustment, selection and validation of the models that will be incorporated into the decision system is performed. For this purpose, the $\mathrm{kNN_{imp}}$ technique is applied to the dataset for classification, with and without selected attributes, using a 5-fold cross-validation. After performing a statistical analysis of the results, the classification model is obtained, which is incorporated into the decision system.

In addition, the kNN-RegID technique is applied to the dataset for regression, with and without selected attributes, using a 5-fold cross-validation and after performing a statistical analysis of the results obtained, the regression model is obtained and incorporated into the decision system.

5.1. Applying FDT$_{ii}$ technique: Characterization of the problem from the datasets for classification

Using the datasets for classification and the different models provided by FDT$_{ii}$ technique, certain characteristic elements for the detection of frost from weather attributes previously measured in certain geographical areas are obtained. For this purpose, FDT$_{ii}$ technique is applied to the dataset for the different stations using a cross-validation and training. In the Table 7, the results obtained for the different measures are shown.

Table 7: Results in % of a 5-fold cross-validation and training obtained with the FDT$_{ii}$ technique and datasets for classification (in subscript the standard deviation is shown). The *M* values obtaining these results for the different stations are as follows: 380, 296, 227, 212 and 207, respectively.

		CR12	CR32	CI52	JU71	JU81	mean
5cv	ACC	$95.19_{0.3}$	$95.31_{0.4}$	$96.82_{0.2}$	$96.10_{0.2}$	$97.60_{0.3}$	$96.20_{0.3}$
	TPR	$83.28_{1.0}$	$90.51_{1.2}$	$88.47_{1.7}$	$85.66_{2.1}$	$88.37_{2.4}$	$87.26_{1.7}$
	TNR	$97.61_{0.2}$	$97.08_{0.3}$	$98.30_{0.3}$	$97.86_{0.4}$	$98.74_{0.1}$	$97.92_{0.3}$
	F-score	$85.40_{0.6}$	$91.25_{1.0}$	$89.23_{0.7}$	$86.36_{0.5}$	$88.96_{1.4}$	$88.24_{0.8}$
	E	$918.2_{77.3}$	$658.4_{57.5}$	$340.4_{51.1}$	$381.0_{51.6}$	$230.6_{46.6}$	$505.7_{56.8}$
training	ACC	95.19	96.81	96.83	96.81	97.65	96.66
	TPR	82.68	89.02	87.40	89.02	86.70	86.96
	TNR	97.73	98.18	98.49	98.18	99.00	98.32
	F-score	85.29	89.28	89.17	89.28	89.00	88.40
	E	4727	1631	1827	1631	1284	2220

From the models obtained by train, the prediction of the classes Frost and NoFrost can be characterized because these models use a subset of the attributes. The attributes used by the different classification models for the different stations are as follows:

- CR12 station – The model uses 7 attributes: DE^{H-1}, R_f^H, AR^H, WS_f^H, DE^H, VPD^H, T_f^H.
- CR32 station – The model uses 7 attributes: VPD^{H-1}, R_f^H, AR^H, WD^H, DE^H, VPD^H,T_f^H.
- CI52 station – The model uses 5 attributes: VPD^{H-1}, R_f^H, AR^H, WD^H, T_f^H.
- JU71 station – The model uses 5 attributes: DE^{H-1}, VPD^{H-1}, R_f^H, AR^H, T_f^H.
- JU81 station – The model uses 6 attributes: RH_f^{H-1}, VPD^{H-1}, T_f^{H-1}, R_f^H, AR^H, T_f^H.

5.1.1. Analyzing the information and characterizing the problem

According to the different models learned through the FDT$_{ii}$ technique for the different stations, some conclusions referring to the weather conditions and the terrain orography that condition the occurrence of a frost are obtained.

First of all, it is important to emphasize that of the attributes selected and indicated above, and according to the models obtained, there is a subset of them that mainly characterize the Frost class. In general, the most relevant attributes are radiation, accumulated radiation and temperature. These three attributes appear in the first and second level of all the models being therefore the ones that better discriminate the instances labeled with Frost from those of

NoFrost. In all stations, and when the current temperature values are negative but close to 0 (values around [-2.0^{o}C,-0.5^{o}C]), the prediction of a frost depends on the values of radiation (less than 70w/m^2) and accumulated radiation (less than 0.25w/m^2).

When the temperature values are at [-1.0^{o}C,2.0^{o}C], it is when the other attributes are involved for the prediction of the label Frost. Specifically, and knowing that the stations JU71 and JU81 belong to the region of the "Altiplano" and present an orography of flatlands and mountains; that the stations CR12 and CR32 belong to the region of the "Noroeste", emphasizing the station CR12 for being located in the most mountainous zone and of greater altitude; and finally that, the station CI52 belongs to the "Vega del Segura" and is the one of lower altitude, the conclusions obtained are the following ones:

- The models indicate that in the region of the "Altiplano" the wind measure does not influence in the prediction of a frost. In the area of the JU71 station, when the temperature is at [-1.0^{o}C,1.0^{o}C], a frost is predicted for the next hour when the radiation is very low (less than 1.0w/m^2) and the dew point of the previous hour is at [-24.0^{o}C,-2.96^{o}C]. In the area of the JU81 station, and when the temperature is at [-1.0^{o}C,1.0^{o}C], a frost is predicted for the next hour when the radiation is very low (less than 0.1w/m^2).

- The "Noroeste" region is the highest. In the less mountainous area (mountains and valleys around the CR32 station), there is an influence of the wind direction for the prediction of a frost when the radiation is low (less than 0.13w/m^2). On the one hand, if the temperature is at [0.5^{o}C,2.0^{o}C] and the radiation is low (less than 0.13w/m^2) a frost is predicted when dew point is less than -3.5^{o}C; and when dew point is at [-3.5^{o}C,1.0^{o}C], a frost is predicted when the wind direction is south/southwest. On the other hand, if the temperature is at [-1.0^{o}C,1.0^{o}C] and the radiation is low (less than 0.13w/m^2), a frost occurs when the direction of the wind is not south/southwest; or if the direction is south/southwest but the dew point is at [-22.0^{o}C,4.0^{o}C].

  However, in the more mountainous areas (around CR12), there is an influence of wind speed (light wind, less than 8m/s), low dew point (at [-22.0^{o}C,4.0^{o}C]) and low radiation (less than 3.7w/m^2) in the prediction of a frost.

- In the region of the "Vega del Segura" also has an influence the wind direction in the prediction of a frost. When the temperature is at [-1.0^{o}C,1.0^{o}C] and the radiation is low (less than 0.1w/m^2), a frost is predicted when the wind direction is from the north/northeast, or if the wind direction is not from the north/northeast, a frost is predicted if the vapor pressure deficit is low (this at [0.0kPa,1.7kPa]).

5.2. Applying the kNN_{imp} technique: Selecting a classifier

5.2.1. Using datasets for classification

In this section the executions of the kNN_{imp} technique with the datasets for classification for the different stations are displayed. Table 8 shows the results obtained for the different assessment measures by the $\text{kNN}^{\text{C}}_{\text{imp}}$, $\text{kNN}^{\text{F}}_{\text{imp}}$ and $\text{kNN}^{\text{Fcons}}_{\text{imp}}$ techniques when a 5-fold cross-validation is used.

As it can be seen in Table 8, the cost of wrong classification obtained by the $\text{kNN}^{\text{Fcons}}_{\text{imp}}$ technique has improved significantly with respect to those obtained by the $\text{kNN}^{\text{C}}_{\text{imp}}$ technique.

Table 8: Results in % of a 5-fold cross-validation (subscripts show standard deviation). For each station and measure, the results obtained by $\mathrm{kNN}^{\mathrm{C}}_{\mathrm{imp}}$, $\mathrm{kNN}^{\mathrm{F}}_{\mathrm{imp}}$ and $\mathrm{kNN}^{\mathrm{Fcons}}_{\mathrm{imp}}$ techniques are shown, respectively. Values of k used are 7, 9, 5, 17 and 7 (in $\mathrm{kNN}^{\mathrm{C}}_{\mathrm{imp}}$) and 11, 9, 7, 13 and 11 (in $\mathrm{kNN}^{\mathrm{F}}_{\mathrm{imp}}$ and $\mathrm{kNN}^{\mathrm{Fcons}}_{\mathrm{imp}}$) for each station respectively.

	CR12	CR32	CI52	JU71	JU81	mean
ACC	$95.20_{0.3}$ $[93.7_{0.6},96.3_{0.5}]$ $95.13_{0.3}$	$95.30_{0.3}$ $[94.4_{0.3},95.9_{0.3}]$ $95.20_{0.4}$	$96.62_{0.2}$ $[95.7_{0.3},97.1_{0.1}]$ $96.39_{0.2}$	$95.80_{0.6}$ $[94.4_{0.4},96.7_{0.4}]$ $95.51_{0.4}$	$97.16_{0.4}$ $[95.9_{0.4},98.1_{0.2}]$ $96.94_{0.3}$	$96.02_{0.4}$ $[94.9_{0.4},96.8_{0.3}]$ $95.83_{0.3}$
TPR	$84.23_{1.3}$ $[78.9_{1.8},87.2_{1.7}]$ $87.15_{1.7}$	$91.54_{1.1}$ $[89.5_{0.7},92.4_{0.9}]$ $92.35_{0.9}$	$88.98_{1.1}$ $[85.9_{1.1},90.2_{1.4}]$ $90.24_{1.4}$	$83.98_{1.8}$ $[79.2_{2.1},86.8_{1.2}]$ $86.77_{1.2}$	$89.03_{1.3}$ $[82.7_{1.5},91.9_{1.4}]$ $91.83_{2.2}$	$87.55_{1.3}$ $[83.2_{1.4},89.7_{1.3}]$ $89.67_{1.5}$
TNR	$97.43_{0.3}$ $[96.8_{0.2},98.1_{0.3}]$ $97.70_{0.8}$	$96.69_{0.3}$ $[96.3_{0.5},97.2_{0.3}]$ $96.25_{0.5}$	$97.96_{0.3}$ $[97.5_{0.3},98.3_{0.2}]$ $97.47_{0.3}$	$97.79_{0.5}$ $[97.0_{0.4},98.4_{0.4}]$ $96.99_{0.4}$	$98.16_{0.3}$ $[97.6_{0.3},98.8_{0.3}]$ $97.58_{0.3}$	$97.61_{0.3}$ $[97.0_{0.3},98.2_{0.3}]$ $97.20_{0.5}$
F-score	$85.55_{0.7}$ $[81.0_{1.0},88.8_{1.0}]$ $85.79_{0.5}$	$91.33_{0.9}$ $[89.7_{0.8},92.3_{0.9}]$ $91.23_{1.0}$	$88.67_{1.0}$ $[85.7_{1.2},90.3_{1.1}]$ $88.12_{1.2}$	$85.24_{1.5}$ $[80.4_{0.8},88.4_{1.0}]$ $84.81_{1.0}$	$87.30_{1.5}$ $[81.7_{1.7},91.3_{0.8}]$ $86.82_{1.0}$	$87.62_{1.1}$ $[83.7_{1.1},90.2_{1.0}]$ $87.35_{0.9}$
E	$874.8_{87.7}$ $[709_{111.6},1168_{133.6}]$ $778.4_{60.7}$	$599.2_{49.8}$ $[539_{43.1},741_{32}]$ $554.8_{41.5}$	$329.6_{24.2}$ $[290_{34.8},421_{38.6}]$ $303.2_{34.1}$	$424_{58.7}$ $[552_{64.1},347_{43.3}]$ $367.6_{43.8}$	$227.2_{30.8}$ $[165_{31},352_{39.9}]$ $185.8_{45.6}$	$491_{50.2}$ $[451_{56.9},606_{57.5}]$ $438_{45.1}$

The results obtained by the $\mathrm{FDT_{ii}}$ technique (Table 7) and those obtained by the $\mathrm{kNN}^{\mathrm{C}}_{\mathrm{imp}}$ and $\mathrm{kNN}^{\mathrm{Fcons}}_{\mathrm{imp}}$ techniques (Table 8) show that techniques based on $\mathrm{kNN_{imp}}$ have globally good behavior. Now, a statistical analysis (Subsection 4.3) of the results is performed to decide which technique has the best behavior.

First, in order to select the technique that obtains the best accuracy with the least error for the Frost class, the ACC and TPR results are analyzed together for these three techniques. Friedman's test is applied, obtaining a rejection of the null hypothesis (p-value=1.095e-04) with a $\alpha = 0.01$. In other words, it is rejected that there are no significant differences. In this situation, Holm's test is performed on the comparison hypotheses between the $\mathrm{kNN}^{\mathrm{C}}_{\mathrm{imp}}$ and $\mathrm{FDT_{ii}}$ techniques and between the $\mathrm{kNN}^{\mathrm{Fcons}}_{\mathrm{imp}}$, $\mathrm{FDT_{ii}}$ and $\mathrm{kNN}^{\mathrm{C}}_{\mathrm{imp}}$ techniques. The p-values obtained from this test are 0.2961, 0.022665 and 0.004632 . Holm's procedure rejects the hypothesis for $\mathrm{FDT_{ii}}$ and $\mathrm{kNN}^{\mathrm{C}}_{\mathrm{imp}}$, indicating that the $\mathrm{kNN}^{\mathrm{Fcons}}_{\mathrm{imp}}$ technique is statistically better than the other techniques ($\alpha = 0.05$). With respect to the kNN^{C}_{imp} and $\mathrm{FDT_{ii}}$ techniques, there are no significant differences.

Secondly, the techniques are analyzed considering their behavior with respect to the E measure (the one that obtains less value of E is better). Friedman's test is applied, obtaining a rejection of the null hypothesis (p-value=4.287e-06) with a $\alpha = 0.01$. In other words, it is rejected that there are no significant differences. In this situation Holm's test is performed on the same comparison hypotheses previously carried out. The p-values obtained are 4.760e-02, 4.041e-05 and 4.041e-05. Holm's procedure rejects the null hypothesis for FDT_{ii} and $\mathrm{kNN}^{\mathrm{C}}_{\mathrm{imp}}$, indicating that the $\mathrm{kNN}^{\mathrm{Fcons}}_{\mathrm{imp}}$ technique is statistically better than the other techniques ($\alpha = 0.05$). In addition, the $\mathrm{kNN}^{\mathrm{C}}_{\mathrm{imp}}$ technique is better than the $\mathrm{FDT_{ii}}$ technique ($\alpha = 0.05$).

Therefore, the behavior of the three techniques is very satisfactory, highlighting the techniques based on k-nearest neighbors, specifically, the $\mathrm{kNN}^{\mathrm{Fcons}}_{\mathrm{imp}}$ technique. With this analysis, this technique is a candidate for the construction of the decision system.

5.2.2. *Using datasets for classification with the attributes selected by FDT$_{ii}$ technique*

Now, the same previous process is performed using only the attributes selected by the FDT$_{\mathrm{ii}}$ technique.

The kNN$^{\mathrm{F}}_{\mathrm{imp}}$ technique is applied to the datasets for classification with selected attributes (the used attributes are those indicated in the Subsection 5.1). Table 9 shows the results obtained for the different assessment measures by the kNN$^{\mathrm{F}}_{\mathrm{imp}}$ and kNN$^{\mathrm{Fcons}}_{\mathrm{imp}}$ techniques with selected attributes.

Table 9: Results in % of a 5-fold cross-validation with the datasets with selected attributes (subscripts show standard deviation). For each station and measure, the results obtained by kNN$^{\mathrm{F}}_{\mathrm{imp}}$ and kNN$^{\mathrm{Fcons}}_{\mathrm{imp}}$ are shown, respectively. Values of k used are 9, 9, 13, 11 and 9, for each station respectively.

	CR12	CR32	CI52	JU71	JU81	mean
ACC	$[95.2_{0.5},96.5_{0.4}]$ $95.84_{0.4}$	$[94.8_{0.2},96.1_{0.3}]$ $95.40_{0.3}$	$[96.5_{0.2},97.7_{0.2}]$ $97.02_{0.3}$	$[96.1_{0.4},97.3_{0.4}]$ $96.73_{0.4}$	$[97.0_{0.3},97.8_{0.3}]$ $97.36_{0.25}$	$[95.9_{0.3},97.1_{0.3}]$ $96.47_{0.3}$
TPR	$[84.3_{1.1},88.1_{1.0}]$ $88.10_{1.0}$	$[89.5_{0.8},91.8_{0.6}]$ $91.80_{0.6}$	$[87.5_{1.9},91.0_{1.2}]$ $90.98_{1.2}$	$[85.0_{1.8},89.7_{1.4}]$ $89.70_{1.4}$	$[85.9_{1.8},89.6_{1.3}]$ $89.59_{1.34}$	$[86.5_{1.5},90.0_{1.1}]$ $90.03_{1.1}$
TNR	$[97.4_{0.3},98.2_{0.3}]$ $97.59_{0.6}$	$[96.7_{0.3},97.7_{0.3}]$ $96.74_{0.3}$	$[98.1_{0.4},98.8_{0.2}]$ $98.09_{0.4}$	$[97.9_{0.5},98.6_{0.4}]$ $97.92_{0.5}$	$[98.3_{0.1},98.9_{0.1}]$ $98.32_{0.14}$	$[97.7_{0.3},98.4_{0.3}]$ $97.73_{0.4}$
F-score	$[85.6_{0.9},89.5_{0.7}]$ $87.78_{0.6}$	$[90.3_{0.8},92.8_{0.7}]$ $91.53_{0.8}$	$[88.2_{0.3},92.0_{0.5}]$ $90.12_{0.7}$	$[86.2_{1.0},90.7_{0.9}]$ $88.80_{0.9}$	$[86.1_{1.5},90.1_{1.0}]$ $88.15_{1.09}$	$[87.3_{0.9},91.0_{0.8}]$ $89.28_{0.8}$
E	$[657_{81.5},875_{103.8}]$ $694_{63.9}$	$[565_{25.2},728_{35.6}]$ $582.4_{24.9}$	$[264_{42.5},370_{54.4}]$ $275.6_{41.6}$	$[272_{40.3},396_{50}]$ $282.2_{38.6}$	$[207_{31.6},281_{37.1}]$ $214.8_{30.9}$	$[393_{44.2},530_{56.2}]$ 409.8_{40}

The behavior of kNN$^{\mathrm{Fcons}}_{\mathrm{imp}}$ technique with all attributes and with selected attributes is satisfactory as is shown in the results in Table 8 and 9, respectively. Now, a statistical analysis of the results is performed to decide which technique has a better behavior.

In order to decide which technique obtains the best accuracy with the least error number for the Frost class, the ACC and TPR results are analyzed together for these two techniques. Wilcoxon test is applied, obtaining a rejection of the null hypothesis (p-value=0.000481) with a $\alpha = 0.01$ (at 99.95%). In other words, the test indicates that the kNN$^{\mathrm{Fcons}}_{\mathrm{imp}}$ technique with selected attributes is statistically better than the other technique. If the behavior of the techniques with respect to the E measure is compared, the conclusions are as follows: Wilcoxon test rejects the null hypothesis (p-value= 0.01732) with a $\alpha = 0.02$ (at 99.27%). Test indicates that the kNN$^{\mathrm{Fcons}}_{\mathrm{imp}}$ technique with selected attributes is statistically better than the other technique. And, if the behavior of the techniques is compared with respect to the measure $F-score$, the conclusions are as follows: Wilcoxon test rejects the null hypothesis (p-value= 9.835e-07) with a $\alpha = 0.01$ (at 99.99%). Test indicates that the kNN$^{\mathrm{Fcons}}_{\mathrm{imp}}$ technique with selected attributes is statistically better than the the kNN$^{\mathrm{Fcons}}_{\mathrm{imp}}$ technique with all attributes.

In Figure 2, radial plots to compare the results obtained by the techniques FDT$_{\mathrm{ii}}$, kNN$^{\mathrm{Fcons}}_{\mathrm{imp}}$ with all attributes and kNN$^{\mathrm{Fcons}}_{\mathrm{imp}}$ with selected attributes regarding the three areas of study are shown. “Noroeste” area includes CR12 and CR32 stations, “Vega del Segura” area includes CI52 station, and Altiplano area includes JU71 and JU81 stations.

In Figure 3, bar graphs to compare kNN$^{\mathrm{C}}_{\mathrm{imp}}$, kNN$^{\mathrm{F}}_{\mathrm{imp}}$ and kNN$^{\mathrm{Fcons}}_{\mathrm{imp}}$ techniques, with and without attributes selected, regarding to the mean of E measure (classification cost mean) are shown.

Figures 2 and 3 show that kNN$^{\mathrm{Fcons}}_{\mathrm{imp}}$ technique with selected attributes obtains the highest values for ACC, TPR, TNR and F-Score measures, and the lowest value for E measure, for all areas.

In summary, globally and statistically, the technique with the best behaviour is kNN$^{\mathrm{Fcons}}_{\mathrm{imp}}$ with selected attributes.

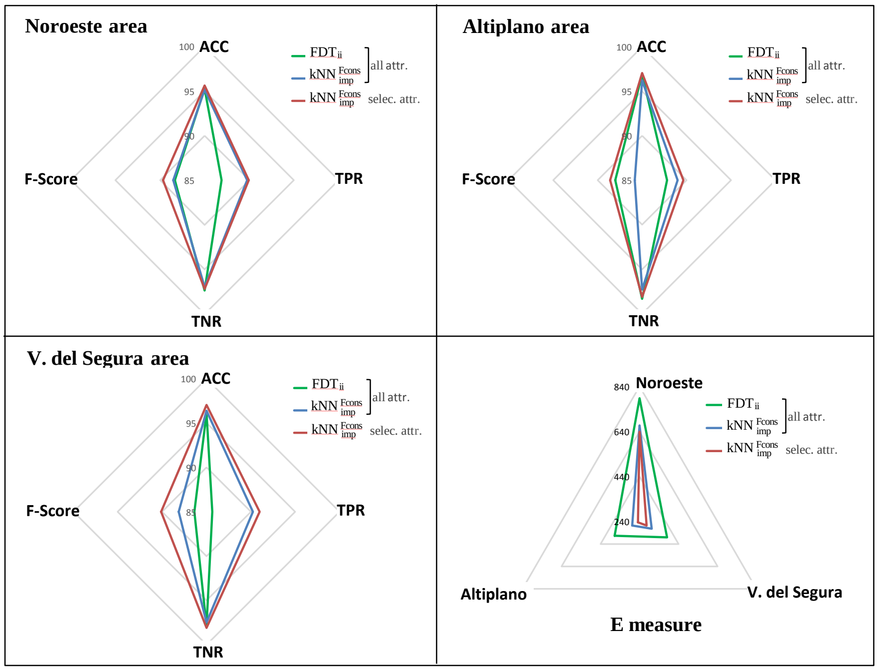

Figure 2: Comparison of the FDT$_{ii}$, kNN$^{Fcons}_{imp}$ using all attributes and kNN$^{Fcons}_{imp}$ with selected attributes regarding the different measures grouped by the different covered areas.

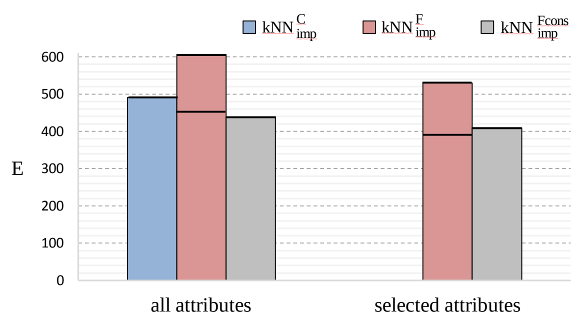

Figure 3: Comparison of the kNN$^{C}_{imp}$, kNN$^{F}_{imp}$ and kNN$^{Fcons}_{imp}$ techniques with respect to measure E using the mean for all stations.

Table 10 shows the results obtained by this technique using the training (complete datasets) and the same parameters used for 5-fold cross-validation.

Table 10: Results in % of train obtained with the kNN$^{Fcons}_{imp}$ technique and datasets for classification with selected attributes.

	CR12	CR32	CI52	JU71	JU81	mean
ACC	96.23	96.11	97.53	97.47	97.70	97.01
TPR	87.01	92.08	90.42	89.75	89.89	89.83
TNR	98.10	97.61	98.77	98.77	98.66	98.38
F-score	88.62	92.78	91.60	91.10	89.55	90.73
E	3570	2756	1395	1339	1019	2015.8

5.3. *Applying kNN-RegID technique: Predicting the temperature value*

The proposed kNN-RegID technique is applied to the datasets for regression to infer the value T_{MIN}^{H+1} from the attributes taken at the current hour and one hour before. Also the different datasets of the stations using only the attributes selected by FDT_{ii} technique are used (Subsection 5.1). The results obtained are shown in the Table 11.

Table 11: Measurements obtained from 5-fold cross-validaton for regression when estimating the T_{MIN}^{H+1} attribute with the kNN-RegID technique and datasets without and with selected attributes. The used k values are 9, 7, 7, 9, 5 and 13, 11, 7, 9, 7, respectively.

	CR12	CR32	CI52	JU71	JU81	mean
MSE	$0.8323_{0.027}$	$0.8634_{0.031}$	$0.6951_{0.041}$	$0.9534_{0.057}$	$0.5942_{0.012}$	$0.7877_{0.034}$
	$0.6654_{0.016}$	$0.7612_{0.037}$	$0.5343_{0.039}$	$0.6676_{0.041}$	$0.5125_{0.020}$	$0.6282_{0.031}$
MAE	$0.6909_{0.011}$	$0.6812_{0.014}$	$0.6166_{0.010}$	$0.7310_{0.022}$	$0.5796_{0.008}$	$0.6599_{0.013}$
	$0.6063_{0.006}$	$0.6348_{0.011}$	$0.5397_{0.014}$	$0.6055_{0.016}$	$0.5389_{0.009}$	$0.5850_{0.011}$
CC	$0.9486_{0.001}$	$0.9634_{0.001}$	$0.9536_{0.004}$	$0.9370_{0.004}$	$0.9556_{0.002}$	$0.9516_{0.002}$
	$0.9590_{0.001}$	$0.9676_{0.002}$	$0.9642_{0.003}$	$0.9559_{0.003}$	$0.9616_{0.002}$	$0.9617_{0.002}$
$R^2(\%)$	$89.945_{0.273}$	$92.751_{0.298}$	$90.836_{0.726}$	$87.661_{0.683}$	$91.252_{0.314}$	$90.489_{0.459}$
	$91.959_{0.229}$	$93.606_{0.376}$	$92.956_{0.638}$	$91.354_{0.601}$	$92.455_{0.349}$	$92.466_{0.439}$

Applying the statistical tests to the results obtained for MSE and MAE in the Tables 11, it is obtained that there are significant differences between them with a $\alpha = 0.01$ (p-value=2.98e-08), being the kNN-RegID technique with selected attributes the best. In addition, as shown these tables, the correlation and determination coefficients have increased for this technique. Table 12 shows the results for train using this technique with the same values of k as those used in 5-fold cross-validation. As an example, Figure 4 shows the behavior, during the indicated time periods, of the inference made by the technique for the attribute T_{MIN}^{H+1} compared with the real values.

Table 12: Measurements obtained from training for regression when estimating the T_{MIN}^{H+1} attribute with the kNN-RegID technique and datasets for regression with selected attributes.

	CR12	CR32	CI52	JU71	JU81	mean
MSE	0.5518	0.6154	0.3884	0.5149	0.3647	0.4870
MAE	0.5504	0.5683	0.4588	0.5290	0.4545	0.5122
CC	0.9661	0.9739	0.9742	0.9661	0.9728	0.9706
$R^2(\%)$	93.33	94.84	94.89	93.34	94.64	94.207

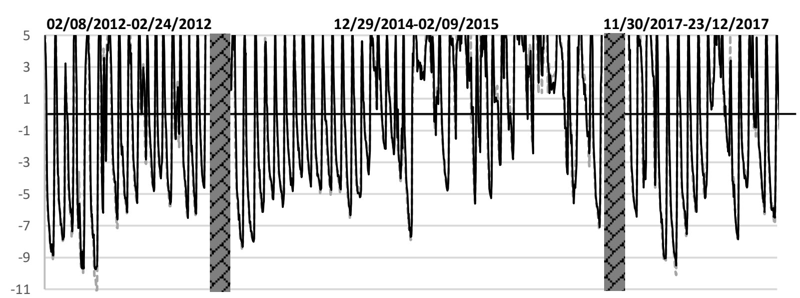

Figure 4: Regression on time periods where different frost events have occurred at CR32 station. The discontinuous gray line represents the real value of the instances and the black line represents the values inferred by kNN-RegID technique and datasets for regression with selected attributes.

6. AFROC: A decision system to Avoid FROst on Crops

The AFROC decision system helps the user to make a decision regarding frost alerts, and therefore, take appropriate actions to reduce its impact. It is therefore a system designed at the operational level. In Figure 5, an illustrative scheme of the system is shown.

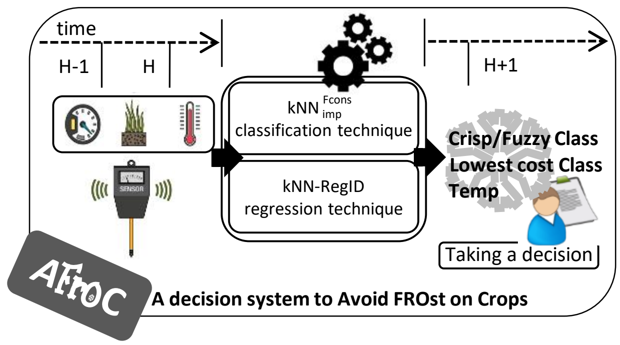

Figure 5: A decision system to prediction of frost.

The system uses datasets with all instances and with selected attributes. Therefore, the technique for classification that is incorporated into the decision system is $\mathrm{kNN}^{\mathrm{Fcons}}_{\mathrm{imp}}$, with an average accuracy of 97.01% (89.83% sensitivity). Using this technique, the system provides the fuzzy prediction $\{\mu_F/\mathrm{Frost}, \mu_{nF}/\mathrm{NoFrost}\}$ when $\frac{|\mu_F - \mu_{nF}|}{max\{\mu_F, \mu_{nF}\}} \leq U_D$, or the crisp decision Frost or NoFrost in another case. In addition, the kNN-RegID technique to regression that is incorporated into the decision system has MSE value of 0.4870 and MAE value of 0.5122.

During operation, the system takes as input data the required weather attributes of the current and previous hour as indicated in the classification and regression models. From these attributes, it is carry out both the frost prediction (CLASS_{FnF}) and the temperature value prediction (T_{MIN}) for the next hour. More specifically, the classification model provides the crisp prediction Frost or NoFrost, if possible, or the imprecise prediction $\{\mu_F/\mathrm{Frost}, \mu_{nF}/\mathrm{NoFrost}\}$. In the latter situation, the system reports that the Frost decision is the one with the lowest economic risk (conservative decision or lowest cost class), but the user, as an expert, can make the most suitable decision based on the degrees μ_F and μ_{nF} and the temperature value estimated for T_{MIN}.

In Table 13 the estimated reply time of the system to the input data is shown, both for obtaining class value and temperature. The computer used to make these estimates is an Intel Core i7-6700HQ 2.6 GHz with 16GB RAM.

Table 13: Time (in ms) estimated to obtain the inference (classification and regression) in the decision system.

	CR12	CR32	CI52	JU71	JU81	mean
$\mathrm{kNN}^{\mathrm{F}}_{\mathrm{imp}}$	25.17	17.04	11.69	10.25	12.32	15.29
kNN-RegID	20.40	16.30	10.78	9.47	11.06	13.60

Finally, it is important to highlight that the system is scalable and extensible to any agricultural crop plot. It simply requires of the technology necessary to collect information of the indicated weather attributes. With this information, the classification and regression models associated with this plot are created and the AFROC system can now begin to help farmers. In addition, the U_D parameter must be taken into account when implanting the system. This parameter defines the vagueness of the class value provided by the technique according to the needs of the user. When the system provides a fuzzy class value (fuzzy subset), the uncertainty is maximum since the values of fuzzy subset will have very close membership degrees. Adjusting this parameter causes the user to make their own decisions on a smaller situation number.

7. Conclusions

This manuscript has been focused on the problem of frost prediction in crops. The problem has been addressed through a process of intelligent data analysis. From the initial dataset composed of a time series data with weather attributes provided by Agricultural Information Service of Murcia Region, the preprocessing phase has allowed to obtain a dataset with fewer instances that maintain the time dependencies and with fewer attributes. The reduction of attributes has been carried out using two strategies: 1) grouping attributes that refer to the same measure in a single attribute expressed by a fuzzy value and 2) performing a selection of the most relevant attributes to carry out the frost prediction from the fuzzy model generated by the FDT_{ii} technique. The information provided by this model has allowed to analyze the relationship between the weather attributes and the orography to predict a frost.

As a result of the preprocessing phase, the datasets obtained contain attributes expressed by imprecise values. Because of this, classification and regression techniques capable of dealing with this type of values are needed. Specifically, a technique based on the *k*-nearest neighbors has been proposed for regression and the FDT_{ii} and kNN_{imp} classification techniques have been used. Several experiments have been performed on the available datasets in order to find the classification and regression models that provide the best results. The results obtained, supported by statistical tests, indicate that the models obtained from the datasets containing the selected attributes obtain a better result. More specifically, using $\text{kNN}^{\text{F}}_{\text{imp}}$ technique, which can provide fuzzy class values as an output, and taking a conservative decision (a decision that implies a lower economic cost) the best results are obtained with an average accuracy of 97.01% (89.83% sensitivity). On the other hand, the regression model provided by the kNN-RegID proposed technique obtains an MSE precision of 0.4870 and MAE of 0.5122 in the prediction of the temperature value from previous weather values which is quite acceptable.

Finally, as a result of the above process, the AFROC decision system has been designed to inform and alert about possible frosts with the aim that the farmer can take appropriate actions to reduce its impact. With the data obtained from the sensors at the current hour and at a previous hour, the system is able to provide the Frost or NoFrost prediction together with the temperature value T_{MIN} for the next hour. The system also provides the degrees of Frost and NoFrost (μ_F and μ_{nF}) allowing the farmer to make the final decision based on his/her experience as an expert. In addition, because the system provides its output from a reduced set of attributes, the number of technological sensors needed decreases, making it easier to implement and maintain the system.

Acknowledgment

Supported by the projects TIN2017-86885-R (AEI/ FEDER, UE), RTC-2017-6389-5 and TIN2016-78799-P (AEI/FEDER, UE), granted by the Ministry of Economy and Competitiveness of Spain (including ERDF support).

Conflict of interest

The authors declare that they have no conflict of interest.

References

Beckenbach, E., Bellman, R., 1961. An Introduction to Inequalities. The L.W. Singer Company.

Cadenas, J.M., Garrido, M.C., Martínez, R., Bonissone, P.P., 2018. A fuzzy k-nearest neighbor classifier to deal with imperfect data. Soft Computing 22, 3313–3330. doi:10.1007/s00500-017-2567-x.

Cadenas, J.M., Garrido, M.C., Martínez, R., Muñoz, E., Bonissone, P.P., 2012. Extending information processing in a fuzzy random forest ensemble. Soft Computing 16, 845–861. doi:10.1007/s00500-011-0777-1.

Dengfeng, L., Chuntian, C., 2002. New similarity measures of intuitionistic fuzzy sets and application to pattern recognitions. Pattern Recognition Letters 23, 221–225. doi:10.1016/S0167-8655(01)00110-6.

Efendi, R., Samsudin, N.A., Arbaiy, N., Deris, M.M., 2017. Maximum-minimum temperature preduction using fuzzy random auto-regression time series model, in: 5th International Symposium on Computational and Bussiness Intelligence, Dubai, United Arab Emirates. pp. 57–60.

Fresh Plaza, 2018. Daily newsletter of the Europe edition. Https://www.freshplaza.com/article/2190407/southern-europe-concerns-about-frost-damage-on-early-stone-fruit/.

Fresh Plaza, 2019. Daily newsletter of the Europe edition. Https://www.freshplaza.com/article/9089079/spain-murcia-estimates-losses-caused-by-frost-at-14-7-million-euro/.

Fuentes, M., Campos, C., García-Loyola, S., 2018. Application of artificial neural networks to frost detection in central Chile using the next day minimum air temperature forecast. Chilean Journal of Agricultural Research 78, 912–915. doi:10.4067/S0718-58392018000300327.

García, S., Fernández, A., Luengo, J., Herrera, F., 2010. Advanced nonparametric tests for multiple comparisons in the design of experiments in computational intelligence and data mining: Experimental analysis of power. Information Sciences 180, 2044–2064. doi:10.1016/j.ins.2009.12.010.

García-Pedraza, L., García-Vega, J., 1991. Las heladas de irradiación en España. Hojas divulgadoras - Ministerio de Agricultura, Pesca y Alimentación de España 1/91, 1–21. Https://www.mapa.gob.es/ministerio/pags/biblioteca/hojas/hd_1991_01.pdf.

Guillén-Navarro, M.A., Cadenas, J.M., Garrido, M.C., Ayuso, B., Martínez-España, R., 2019. Minimum temperature prediction models in plots to forecast frost in crops, in: Muñoz, A., Park, J. (Eds.), Agriculture and Environment Perspectives in Intelligent Systems. IOS Press. volume 26, pp. 91–106.

Guillén-Navarro, M.A., Cadenas, J.M., Garrido, M.C., Ayuso, B., Martínez-España, R., 2018. A preliminary study to solve crop frost prediction using an intelligent data analysis process, in: Chatzigiannakis, I., Tobe, Y., Novais, P., Amft, O. (Eds.), Intelligent Environments 2018 - Workshop Proceedings of the 14th International Conference on Intelligent Environments. IOS Press. volume 23, pp. 97–106.

Han, J.H., Kim, Y.K., 1999. A fuzzy k-nn algorithm using weights from the variance of membership values, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition (IEEE CVPR'99), Ft. Collins, Colorado, USA.. pp. 394–399.

Ihaka, R., Gentleman, R., 1996. R: a language for data analysis and graphics. Journal of computational and graphical statistics 5, 299–314. doi:10.1080/10618600.1996.10474713.

Keller, J.M., Gray, M.R., Givens, J.A., 1985. A fuzzy k-nearest neighbor algorithm. IEEE transactions on systems, man, and cybernetics SMC-15, 580–585. doi:10.1109/TSMC.1985.6313426.

Lee, H., Chun, J.A., Han, H.H., Kim, S., 2016. Prediction of frost occurrences using statistical modeling approaches. Advances in Meteorology 2016, 1–9. doi:10.1155/2016/2075186.

Liakos, K.G., Busato, P., Moshou, D., Pearson, S., Bochtis, D., 2018. Machine learning in agriculture: A review. Sensors 18, 2674. doi:10.3390/s18082674.

de Luca, A., Termini, S., 1972. A definition of a nonprobabilistic entropy in the setting of fuzzy sets theory. Information and Control 20, 301–312. doi:10.1016/S0019-9958(72)90199-4.

Martínez-Núñez, L., Moreno, J.V., Chazarra, A., Gallego-Abaroa, T., Avello, E., Botey, M.R., 2015. Mapas de riesgo: Heladas y horas de frío en la España peninsular (2002-2012). "Agencia Estatal de Meteorología" of Spain. Https://repositorio.aemet.es/bitstream/20.500.11765/8799/1/Mapas_de_riesgo_2002-2012.pdf.

Mitchell, H., Schaefer, P., 2001. A "soft" k-nearest neighbor voting scheme. International journal of intelligent systems 16, 459–468. doi:10.1002/int.1018.

Smith, B.A., Hoogenboom, G., McClendon, R.W., 2009. Artificial neural networks for automated year-round temperature prediction. Computers and Electronics in Agriculture 68, 52–61. doi:10.1016/j.compag.2009.04.003.

Snyder, R.L., Melo-Abreu, J.P., Villar-Mir, J.M., 2010. Protección contra las heladas: fundamentos, práctica y economía. volume 1. Organización Naciones Unidas para la Agricultura y la Alimentación - FAO.

Yelapure, S., Kulkarni, R., 2012. Literature review on expert system in agriculture. International Journal of Computer Science and Information Technologies 3, 5086–5089. Http://www.ijcsit.com/docs/Volume 203/vol3issue5/ijcsit2012030530.pdf.

Yu, H., Chen, Y., Hassan, S.G., Li, D., 2016. Prediction of the temperature in a Chinese solar greenhouse based on LSSVM optimized by improved PSO. Computers and Electronics in Agriculture 122, 94–102. doi:10.1016/j.compag.2016.01.019.