ADVANCES IN FOREST FIRE RESEARCH 2022

Edited by
DOMINGOS XAVIER VIEGAS
LUÍS MÁRIO RIBEIRO

Extreme Fire Severity Classification using Clustering and Decision Tree

Henrique Coelho[1]; Susana Nascimento*[1]; Carlos Viegas Damásio[1]; Lourdes Bugalho[2]; Gonçalo Severino[1]

[1]*NOVA LINCS and Departamento de Informática da NOVA School of Science and Technology. Quinta da Torre, 2829-516 Caparica, Portugal,*
{ha.coelho@campus.fct.unl.pt, snt@fct.unl.pt, cd@fct.unl.pt, g.severino@campus.fct.unl.pt}
[2] *Instituto Português do Mar e da Atmosfera. IPMA,I.P, Rua C do Aeroporto 1749-077 Lisboa, Portugal,*
{lourdes.bugalho@ipma.pt}

**Corresponding author*

Keywords

Extreme fires, fire risk scales, Canadian Fire Weather Index, Fuzzy c-means, decision tree

Abstract

With climate change, large, unpredictable, and difficult to suppress forest fires are increasingly frequent. To increase the ability to anticipate and respond to these extreme events it is necessary to characterize the meteorological conditions associated with the risk levels of these events. The main objective of this work is to automatically identify those severity conditions and extract classification rules to characterize extreme forest fires with at least 100ha of burned area (90% percentile) in mainland Portugal for the period 2001-2020.

The conditions characterizing the extreme fires are elicited by applying fuzzy clustering and predictive methods to forest fire data and corresponding fire risk indices, namely the Canadian Forest Fire Risk Index (FWI), and subindices, as well as the Continuous Haines Index (CHI), provided by the Portuguese Institute of Sea and Atmosphere (IPMA). The dates and localization of fires are obtained from the shapefiles provided by the Portuguese Institute for Nature Conservation and Forests (ICNF), and complemented with data from the MODIS Global Burned Area Product MCD64A1 downloaded from the University of Maryland repository.

The popular fuzzy c-means (FCM) algorithm is applied to group fires into five and seven clusters, with no pre-specified ground-truth severity. Then each cluster is labelled with the fire risk scale class assigned to the cluster's prototype considering the EEFIS scale (European-Forest-Fire Information System) for five clusters and IPMA fire risk scale for seven clusters, respectively. Fuzzy Sammon mapping has been used to visualize and validate the fuzzy partitions.

Using the data from 2001-2018, decision trees (DT) were induced in order to obtain the conditions and thresholds that characterize the obtained clusters, and tested with the data from 2019 and 2020. To ensure the quality of the classification results robust validation techniques such as cross-validation and bootstrapping as well as evaluation metrics are applied.

The DT rules described by conjunctions of the fire risk indices and thresholds, were not always in agreement with the reference forest fire risk prediction scales, revealing the importance of adapting the indices values according to the region in question and taking into account several factors (forest fire risk indices) in the analysis of the conditions associated with the level of risk of an extreme forest fire. The proposed approach shown to be a proof of concept to derive an empirical fire severity risk scale for the collection of used indices and to compare the results with the two fire risk scales used by IPMA and EEFIS.

1. Introduction

With climate change, those large, unpredictable, and difficult to suppress forest fires will become increasingly frequent (Petroliagkis et al., 2015). In Portugal, wildfires continue to be one of the most serious natural catastrophes, due to their high frequency and intensity, and with climate change they will become more frequent. To increase our ability to anticipate and respond to these phenomena, it is necessary to study and characterize the meteorological conditions that favor them.

There is no generally accepted definition of what an extreme fire is. This difficulty in defining what an extreme fire is clearly explained by (Viegas, 2012) and (Tedim et al., 2018). However, there is consensus (Tedim et al., 2018; Fernandes, 2005) that three descriptive parameters, burned area (BA), rate of spread (ROS) and fire line intensity (FLI), are necessary indicators to assess what an extreme fire is, as they characterize three essential aspects of fires: damage caused, unpredictability and suppression capacity, respectively. Table 1 presents the limits used by the authors Tedim et al. (2018) and Fernandes (2005) to define an extreme fire.

Table 1. Adapted from (Tedim et al., 2018; Fernandes, 2005)

Aspects	Parameter	Reference Values
Damage	BA	>= 100 ha
Unpredictability	ROS	>= 50m/min (in forests)
Suppression Capacity	FLI	10000-30000kW/m

Although these descriptive parameters are important to define an extreme fire, in practice, only the burned area (BA) will be used, due to the lack of other data and compatibility problems between them. More specifically, we will use the 90% percentile of the size of fires in Mainland Portugal from 2001 to 2020, which corresponds to about 100 ha, which is in accordance with the criteria in Table 1.

To characterize the meteorological conditions for the occurrence of forest fires, it is used the values of the Continuous Haines Index (CHI) (Mills and McCaw, 2010) and the Fire Weather Index (FWI) (Turner and Lawson, 1978; Van Wagner and Pickett, 1985), and their subindices (FFMC, DMC, DC, BUI and ISI). To characterize the level of risk, the ordinal risk scales based on the Fire Weather Index (FWI) defined by IPMA (IPMA-FWI, 2022) and by the European-Forest-Fire Information System (EEFIS) (Joint Research Centre, 2020) will be adopted in our study.

The objective of this work is to explore fuzzy clustering (Ruspini et al., 2019) to unsupervisedly group fire risk data into distinct fire risk classes, such that a fire may have a positive degree of belongingness to more than one risk class. After the data is clustered, decision tree based methods are used (Quinlan, 1986) to derive if-then classification rules characterizing the meteorological conditions associated with the different classes of risk of fire occurrences.

2. Materials and Methods

The shapefiles of forest fires in Portugal in the period 2001-2020 were obtained from the available geocatalog provided by ICNF (https://geocatalogo.icnf.pt/metadados/area_ardida.html). The data was filtered by burned area in QGIS, and the fires with area >= 100ha were retained (of which 113 are in years 2019 and 2020). When not available, the date of the fire ignition data was obtained from the burned area MODIS Global Burned Area Product MCD64A1 (https://modis.gsfc.nasa.gov/data/dataprod/mod45.php).

The data regarding the FWI, FFMC, DMC, DC, BUI, ISI and CHI indexes were provided by IPMA. The data for the FWI are daily and for the CHI are data every 3 hours, for the period 2001 to 2020, calculated from the operational model analysis data of the European Center for Medium-Term Weather Forecasts (ECMWF), with a regular geographic grid of 0.125º latitude and 0.125º longitude, covering the territory of Mainland Portugal. For each fire, the fire risk indexes corresponding to the ignition date of the network with the largest area of intersection were considered.

There were built two data samples considering fires with burned area greater than to 100ha: one taking the five indices (FWI, CHI, ISI, DC, FFMC), and the other one with seven indices (FWI, CHI, ISI, DC, FFMC, BUI and DMC).

The conducted experimental study comprises three stages: (i) pre-processing and exploratory pre-analysis of fire data and risk indicators; (ii) application of the fuzzy c-means algorithm (FCM) (Bezdek, 1981) to generate fuzzy partitions of fires; and (iii) induction of decision trees (Quinlan, 1986) from the fuzzy clusters.

For each data sample the FCM was run looking for five and seven clusters. The clusters prototypes (centroids) of the obtained fuzzy partitions with five clusters (FCM-5) were labelled according to the EEFIS severity risk scale while the ones of the fuzzy partitioning with seven clusters (FCM-7) use the IPMA scale for the classification. Then, the fire fuzzy partitions, originally in a 5/7-dimensional data space, were projected in 2D with fuzzy Sammon mappings (FUZZYSAM) (Feil et al., 2007) for visual inspection and validation. Finally, the data fires were assigned to the cluster with highest belongingness and labeled with the classification of the corresponding cluster prototype.

With all observations thus labeled with a certain fire risk class, there were generated decision trees to extract if-then rules that associate meteorological and terrain conditions with a certain level of extreme fire risk.

To avoid overfitting, a decision tree was induced from the data of period 2001-2018 with depth between 1 and 9, with the hyperparameters tuned using stratified shuffle splitting (using a 50%/50% for training and validation). Finally, the data of 2019 and 2020 fires was classified with respect to the FCM prototypes and used to evaluate the performance of the induced trees, namely with the standard precision, recall and F1 metrics.

3. Results

In this Section we discuss the main results of our approach and summarize the assessment of DT classification with the evaluation metrics.

3.1. Phase 1 – Fuzzy Clustering

Figures 2 and 3 show the FUZZYSAM mappings for the four fuzzy partitions resulting from the four combinations of parameters: **some -** five indices (FWI, CHI, ISI, DC, FFMC) or **all** – seven indices (FWI, CHI, ISI, DC, FFMC, BUI and DMC), and **C5 -** five clusters or **C7-** seven clusters, hereafter designated as **someC5/C7**, **allC5/C7**, respectively. The projected clustered fires are represented as blue dot points, the colored star points represent the clusters prototypes labelled according to the EEFIS/IPMA severity risk scales, while the iso-lines represent degrees of belongingness (from closest prototypes values of 0.9, 0.8, 0.7 etc) derived from the FCM membership function. This projection guarantees the interpoint fuzzy weighted distances of the projected data approximating the corresponding weighted distance in the original (5D/7D) space. It is useful for interpretation of the clustering results since it is based on the Euclidean distance between the clusters prototypes (representative fires with average indices values) and the data fires.

It is interesting to observe that for the two partitions with five clusters, only the fire risk class labels “Very Low / Low”, “High” and “Very High” are used to classify the clusters. In case of fuzzy partitions with seven clusters, the situation depends on the used data sample with all/some indices. Anyway, the fuzzy partition derived from the configuration someC7 has a continuum of the risk class labels “Very Low”, “Low”, “Moderate”, “High”, and “Extreme”, being the best partition.

3.2. Phase 2 – Induced trees

The decision trees induced from the FCM clustered data (configuration someC7), at depths 3 and 4, are presented in Figures 4 and 5. The obtained trees are human interpretable and show that most fire occurrences are concentrated in three (four) leaves, one for each class, indicating that the result can be quite robust. It is interesting to observe that in both situations are used indices for fuel modelling and fire behaviour, and indices modelling the instability and dryness of the atmosphere. It is clear also that some of the leaves could be pruned, getting the same classification. For the case of seven clusters (someC7) all depth 3 or greater present good quality trees for use (Table 2).

3.3. Evaluation Metrics

The evaluation of the classification results from DT with depths 1 to 9, derived from the fuzzy partition someC7, are presented in Table 2. For DTs with depth above 2 the results appear to be very robust, with score values of recall, precision and F1 above 0.8. However, the metric values for depths above 6 are very close indicating that the deeper trees are overfitted. Furthermore, the shallower depth 3 tree obtains a good performance being very effective at separating the “Very Low’’ and “Extreme’’ fire classes from the others. The confusion occurs essentially in the middle classes and all the risk indices are used except the FWI. So, the decision tree with depth 3 is a good compromise between correct classification and rules simplicity presenting a classification F1 score value above 0.80, which is considered very reliable for real data.

Table 2. Evaluation metrics of the induced trees for configuration some C7

Metric (weighted)	DT depth 1	DT depth 2	DT depth 3	DT depth 4	DT depth 5	DT depth 6	DT depth 7	DT depth 8	DT depth 9
Recall	0.458	0.703	0.838	0.867	0.875	0.919	0.931	0.928	0.928
Precision	0.609	0.736	0.836	0.864	0.873	0.918	0.927	0.927	0.927
F1	0.520	0.712	0.832	0.860	0.873	0.917	0.926	0.926	0.926

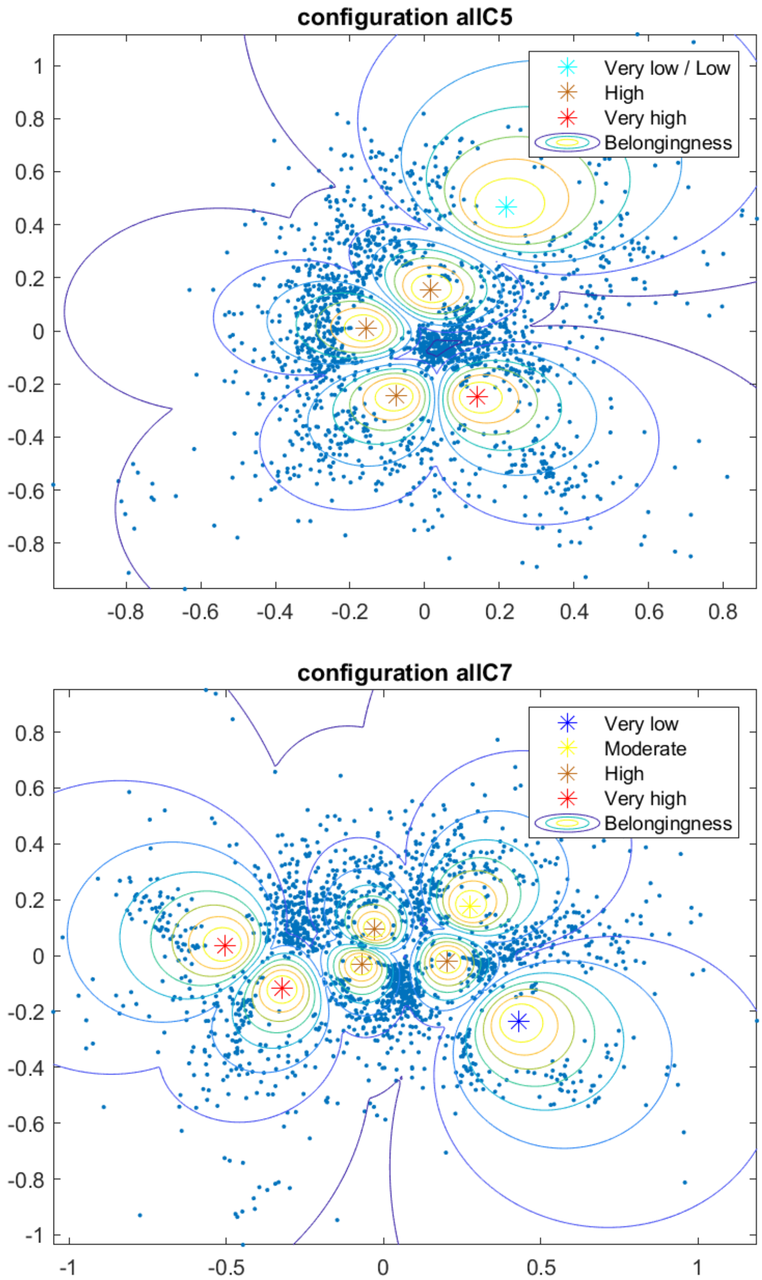


Figure 2. Fuzzy Sammon mapping projection of fuzzy partitions with all the indices: five clusters (top figure) and seven clusters (bottom figure)

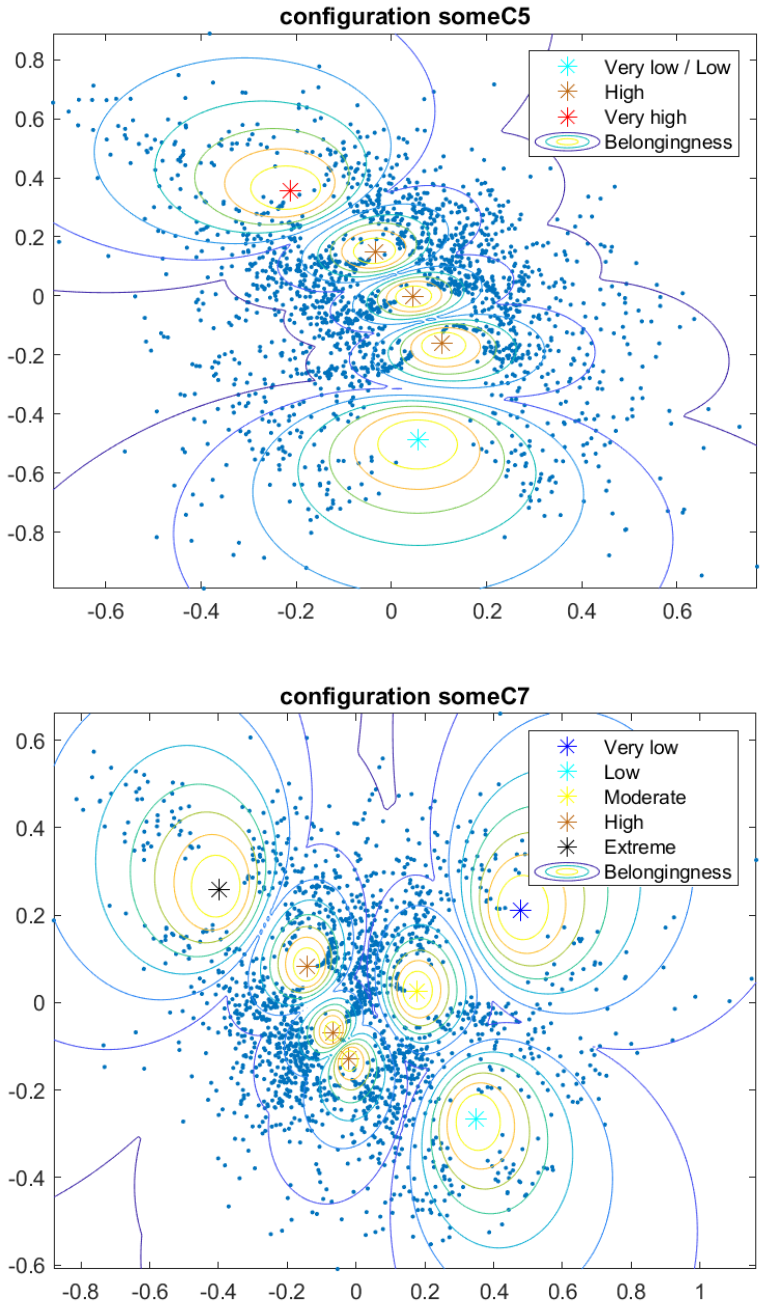
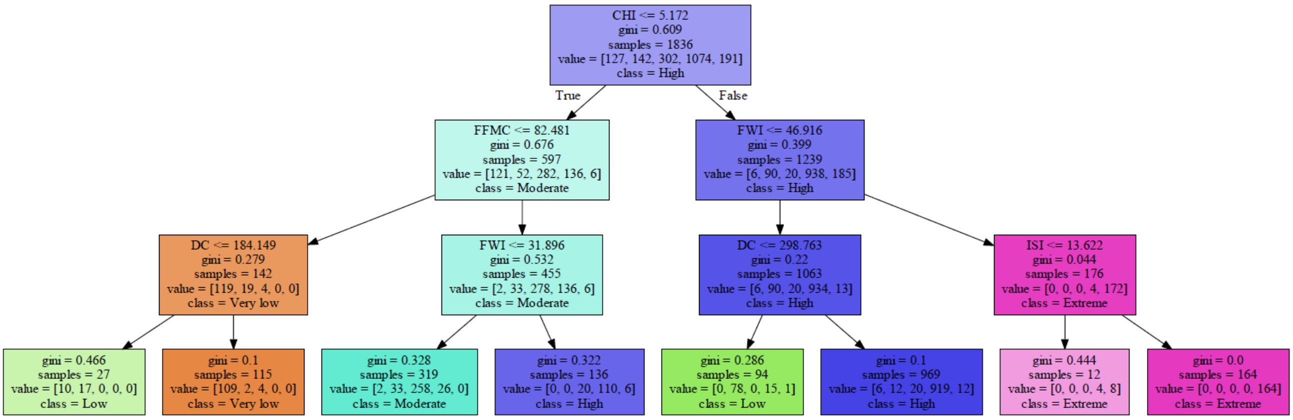


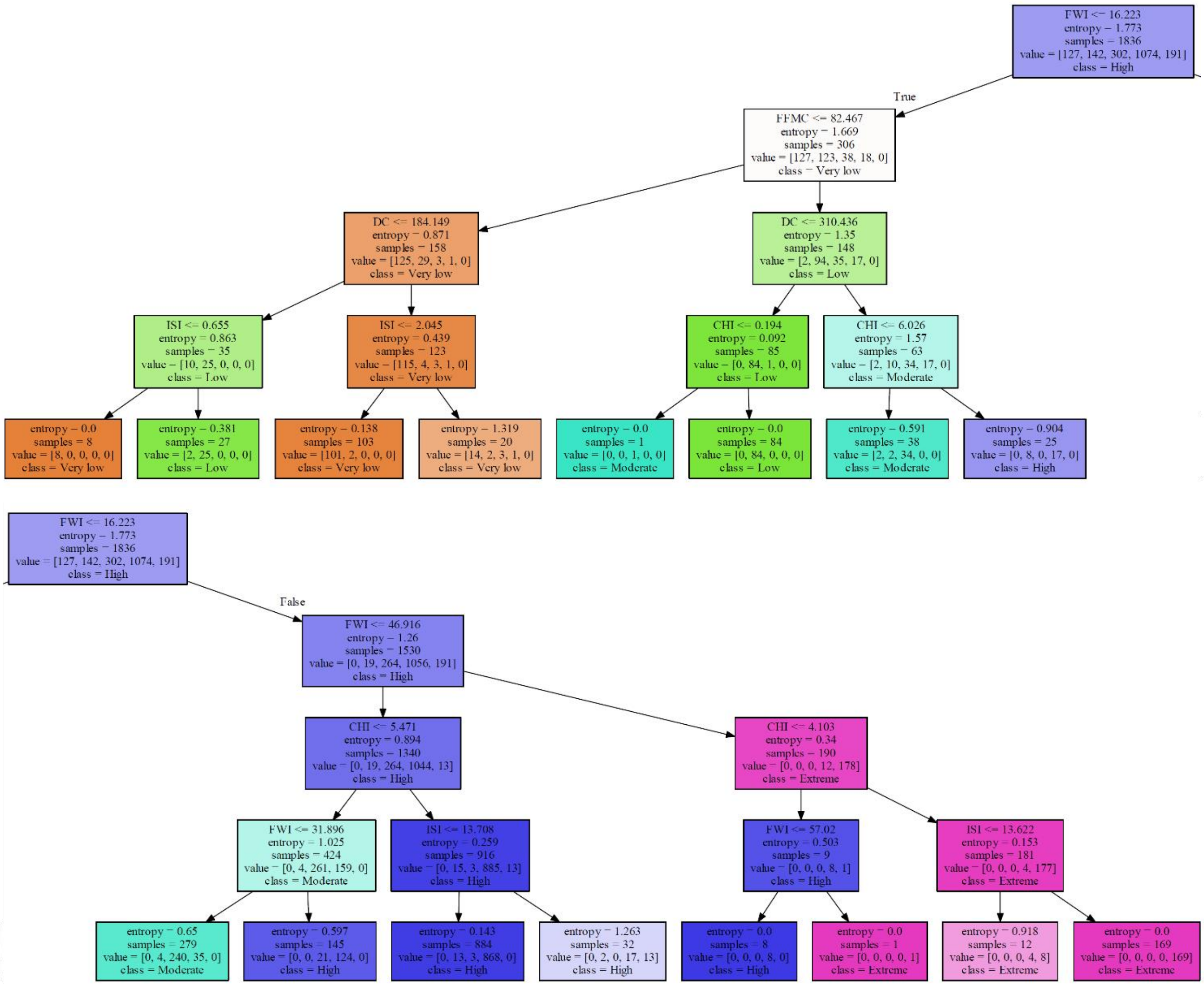
Figure 3. Fuzzy Sammon mapping projection of fuzzy partitions with some of the indices: five clusters (top figure) and seven clusters (bottom figure)



Simplified rules

if (CHI <= 5.172) and (FFMC <= 82.481) and (DC > 184.149) then class: Very low
if (CHI > 5.172) and (FWI <= 46.916) and (DC <= 298.763) then class: Low
if (CHI <= 5.172) and (FFMC <= 82.481) and (DC <= 184.149) then class: Low
if (CHI <= 5.172) and (FFMC > 82.481) and (FWI <= 31.896) then class: Moderate
if (CHI > 5.172) and (FWI <= 46.916) and (DC > 298.763) then class: High
if (CHI <= 5.172) and (FFMC > 82.481) and (FWI > 31.896) then class: High
if (CHI > 5.172) and (FWI > 46.916) then class: Extreme

Figure 4. Induced tree with depth 3 for configuration someC7



Simplified rules

if (FWI <= 16.223) and (FFMC <= 82.467) and (DC > 184.149) then class: Very low
if (FWI <= 16.223) and (FFMC <= 82.467) and (DC <= 184.149) and (ISI <= 0.655) then class: Very low
if (FWI <= 16.223) and (FFMC > 82.467) and (DC <= 310.436) and (CHI > 0.194) then class: Low
if (FWI <= 16.223) and (FFMC <= 82.467) and (DC <= 184.149) and (ISI > 0.655) then class: Low
if (FWI > 16.223) and (FWI <= 46.916) and (CHI <= 5.471) and (FWI <= 31.896) then class: Moderate
if (FWI <= 16.223) and (FFMC > 82.467) and (DC > 310.436) and (CHI <= 6.026) then class: Moderate
if (FWI <= 16.223) and (FFMC > 82.467) and (DC <= 310.436) and (CHI <= 0.194) then class: Moderate
if (FWI > 16.223) and (FWI <= 46.916) and (CHI > 5.471) then class: High
if (FWI > 16.223) and (FWI <= 46.916) and (CHI <= 5.471) and (FWI > 31.896) then class: High
if (FWI <= 16.223) and (FFMC > 82.467) and (DC > 310.436) and (CHI > 6.026) then class: High
if (FWI > 16.223) and (FWI > 46.916) and (CHI <= 4.103) and (FWI <= 57.02) then class: High
if (FWI > 16.223) and (FWI > 46.916) and (CHI > 4.103) then class: Extreme
if (FWI > 16.223) and (FWI > 46.916) and (CHI <= 4.103) and (FWI > 57.02) then class: Extreme

Figure 5. Induced tree with depth 4 for configuration someC7

4. Discussion and Conclusion

This work consisted of a comparative experimental study between the rules that associate meteorological conditions with extreme forest fire risk levels, generated from data mining and machine learning techniques, with the two ordinal scales of fire risk prediction of reference, the EEFIS scale and the IPMA scale. Specifically, the fuzzy clustering algorithm FCM was explored for the grouping of extreme forest fire data, occurred between 2001 and 2018, into severity fire risk classes, and subsequent visualization in 2D FUZZYSAM maps. Then a classification model was induced by decision tree algorithm for the characterization and extraction of rules, that is, to highlight the meteorological conditions associated with different levels of extreme fire risk. The applicability of these algorithms was tested for this purpose.

The FCM algorithm shown to be effective in the segmentation and classification of extreme fires, for the number of clusters five and seven, providing flexibility to the clustering effort with degrees of membership that may distinguish and rank the fires occurrences with respect to their severity. The partition with seven clusters was chosen with some of the indices (FWI, CHI, ISI, DC, FFMC) because the clusters recognize the "Low''/"Very Low'', "High''/"Very High'' and "Extreme" classes of the classification used in the IPMA. The decision tree classification model, once its hyperparameters have been trained and tuned, through hyper-parameterization and validation techniques, stratified shuffle split, also proved to be effective in extracting the meteorological conditions associated with different levels of extreme fire risk. The decision trees induced from the fire data with 5 indices labelled with the FCM-7 cluster prototypes, presented promising results, with classification validity scores higher than 0.8. The rules generated by these trees were used to determine the meteorological conditions associated with different levels of extreme fire risk, having been analyzed and compared with the EEFIS and IPMA scales, showing that the thresholds do not correspond exactly. This is expected since the EEFIS and IPMA scales were obtained from a single index, while our approach combines conditions of several indices: FWI, subindices and Continuous Haines Index. The generated rules clearly separate "Low'', "High'' and "Extreme''. However, while the rules for "Low'' and "Extreme'' are well defined, for high risk there are several sets of rules that define it. It is also interesting that the danger index CHI, linked to the instability and humidity of the lower atmosphere, is considered as important in the classification of danger (especially extreme) for both rules, with depth 3 (Figure4) and rules with depth 4 (Figure5). In the first case, depth 3, the extreme classification is given for the case of CHI>5.172 and FWI>46.9. Other indices are used to define the low, moderate or high classification rules, always having CHI as the first classification division. In the second case, depth 4, the extreme classification is achieved by two rules: 1) FWI>46.9 and CHI>4.103 or 2) If CHI≤4.103 and FWI>57.02 (extremely high value of FWI), Figure 5. Rules for depth 4 are more complex than those for depth 3, especially for the definition of "Low'', "Moderate'' and "High'' classes.

In the future, the same approach will be applied to particular edaphoclimatic regions of Portugal mainland in order to understand if the results are consistently obtained, as well the expected differences among regions. Furthermore, vegetation and topographic indices will be added in order to be able to induce risk level conditions for wildfires.

5. Acknowledgments

This work was supported by Fundação para a Ciência e a Tecnologia (FCT.IP) through project Floresta Limpa (PCIF/MOG/0161/2019) and NOVA LINCS (UIDB/04516/2020). The authors are indebted to the reviewers for their helpful comments that allowed to improve the paper.

6. References

Bezdek, James C., 1981. Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum press.

Feil, B., Balasko, B. Abonyi, J., 2007. Visualization of fuzzy clusters by fuzzy sammon mapping projection: application to the analysis of phase space trajectories. Soft Computing, 11 (5) 479–488.

Fernandes, P. M., 2005. Estudo de adaptação para Portugal do Sistema Canadiano de Indexação do Perigo de Incêndio. Relatório para a Agência de Prevenção de Incêndios Florestais. Departamento Florestal, Universidade de Trás-os-Montes e Alto Douro.

IPMA-FWI (2022). https://www.ipma.pt/pt/riscoincendio/fwi/ (Accessed 20 June 2022).

Joint Research Centre (2020). EFFIS - fire danger forecast. https://effis.jrc.ec.europa.eu/about-effis/technical-background/fire-danger-forecast (Accessed 20 June 2022).

Mills, G. A. e McCaw, L., 2010. Atmospheric Stability Environments and Fire Weather in Australia - extending the Haines Index. CAWCR Technical Report No. 020, The Centre for Australian Weather and Climate Research.

Petroliagkis, T., Liberta, G, Artes, T., Rodriguez-Aseretto, D., Di Leo, M. e San-Miguel, Ayanz. J., 2015. Stability of atmospheric flow and low-level jets influencing forest fire behavior - An EFFIS report. EUR–Scientific and Technical Research series, EUR 27362, 30, 3-28.

Quinlan, J.R., 1986. Induction of decision trees. Machine Learning 1, 81–106. DOI: https://doi.org/10.1007/BF00116251.

Ruspini, E. H., Bezdek, J. C., Keller, J. M., 2019. Fuzzy clustering: A historical perspective. IEEE Computational Intelligence Magazine 14.1, 45-55.

Tedim, F., Leone, V., Amraoui, M., Bouillon, C., Coughlan, M.R., Delogu, G.M., Fernandes, P.M., Ferreira, C., McCaffrey, S., McGee, T.K. et al., 2018. Defining Extreme Wildfire Events: Difficulties, Challenges, and Impacts. Fire 1, 1, 9.

Turner, J. A. e Lawson, B. D., 1978. Weather in the Canadian Forest Fire Danger Rating System: a user guide to national standards and practices. Environment Canada, Canadian Forest Service, Information Report BC-X-177. Pacific Forest Research Centre: Victoria, BC.

Van Wagner C. E. e Pickett, T.L., 1985. Equations and FORTRAN program for the Canadian Forest FireWeather Index System. Government of Canada, Forestry Technical Report 33, Canadian Forest Service: Ottawa, ON.

Viegas, D.X., Bovio, G., Ferreira, A., Nosenzo, A. e Sol, B., 1999. Comparative study of various methods of fire danger evaluation in southern Europe. International Journal of wildland fire, 9, 4, 235-246.

Viegas, D.X., Reis, R.M., Cruz, M.G. e Viegas, M.T., 2004. Calibração do sistema canadiano de perigo de incêndio para aplicação em Portugal. Silva Lusitana 12, 1, 77-93.

Viegas, D. X., 2012. Extreme fire behaviour. In Forest Management: Technology, Practices and Impact. Bonilla Cruz, AC, Guzman Correa, RE, Eds, 1-56.