A Multi-label Approach for Diagnosis Problems in Energy Systems using LAMDA algorithm

Carlos Quintero Gull
Dpto de Ciencias Aplicadas y Humanísticas. Doctorado en Ciencias Aplicadas Facultad de Ingeniería Universidad de Los Andes
Mérida 5101, Venezuela
carlgull@gmail.com

Jose Aguilar
Universidad de Alcalá, Escuela Politécnica Superior, ISG,
Alcalá de Henares, 28805, Spain;
CEMISID, Universidad de Los Andes,
Mérida, 5101, Venezuela;
GIDITIC, Universidad EAFIT,
Medellín, 50022, Colombia
jose.aguilar@uah.es

Maria D. R-Moreno
Universidad de Alcalá, Escuela Politécnica Superior, ISG,
Alcalá de Henares, 28805, Spain;
TNO, Intelligent Autonomous Systems Group (IAS),
The Hague, The Netherlands
malola.rmoreno@uah.es

***Abstract*—In this paper, we propose a supervised multilabel algorithm called Learning Algorithm for Multivariate Data Analysis for Multilabel Classification (LAMDA-ML). This algorithm is based on the algorithms of the LAMDA family, in particular, on the LAMDA-HAD (Higher Adequacy Grade) algorithm. Unlike previous algorithms in a multi-label context, LAMDA-ML is based on the Global Adequacy Degree (GAD) of an individual in multiple classes. In our proposal, we define a membership threshold (Gt), such that for all GAD values above this threshold, it implies that an individual will be assigned to the respective classes. For the evaluation of the performance of this proposal, a solar power generation dataset is used, with very encouraging results according to several metrics in the context of multiple labels.**

Keywords—Multilabel classification, fuzzy systems, LAMDA, diagnosis problems

I. Introduction

In Machine Learning (ML), the supervised learning for the classification problem is referred to the assignment of each individual to a known class, which has a label (single label) [1, 2]. In this context, we can find several cases for the classification problem, for example, the binary classification (2 classes) or multiclass classification (more than 2 classes) [3, 4].

In multilabel classification, each individual can simultaneously belong to a set of classes [3, 4]. This kind of classification is motivated by problems like text categorization, where a text could belong to different documents. Nowadays, multilabel classification is increasingly used in other contexts, like medical diagnosis, and semantic annotation, among other domains [5, 6, 7]. In general, the multilabel classification algorithms, same as the traditional classification algorithms, are of different types: instance-based methods, decision tree-based methods, and kernel-based methods, to name a few.

In the literature, we can find several articles about multilabel classification [8, 9, 10, 11]. This kind of algorithm is split into transformation methods and adaptation methods. The transformation methods are those that transform any multilabel classification problem into several independent binary classifications, for example, binary relevancy and pairwise ranking [11]. In the case of the adaptation methods, the traditional algorithms of classification are modified, adapting them to the multilabel classification, for example, we can mention versions of k-Nearest Neighbor [10], neural networks [12], and decision trees [13], among others. A current problem with these algorithms is that they make a misclassification when the classes have big variations [12].

On other hand, the Learning Algorithm for Multivariate Data Analysis (LAMDA) algorithm is a method based on fuzzy logic, which focuses on calculating the Global Adequacy Degree, from an individual to a class (classification) or group (clustering) [14. 15, 16, 17]. Particularly, LAMDA has the capability of creating new classes/clusters after the training stage. If an individual does not have enough similarity to the preexisting classes/clusters, it is evaluated with respect to a threshold called the Non-Informative Class (NIC), which allows an adaptive online learning. In spite of the LAMDA algorithm has been modified by different works for different purposes, we highlight the contribution [16], where has been developed LAMDA-HAD (Higher Adequacy Degree), which makes an improvement significate in the performance of the traditional LAMDA algorithm in the classification context. In LAMDA-HAD, the efficiency of LAMDA for classification problems is improved based on two strategies [16]. In the first one, an adaptable Non-Informative Class (NIC) for each class is calculated in order to avoid that, correctly classified individuals create new classes. In the second one, the HAD metric is calculated to grant more robustness to the classification process.

In another way, it has not been developed [18, 19, 20, 21] a multilabel classification-based diagnosis approach for energy systems. Traditionally, the Artificial Intelligence (AI)-based diagnosis methods for energy systems are classified into two categories, data driven-based and knowledge driven-based [21, 22]. The data driven-based methods can be classification-based, regression-based, and unsupervised learning-based, among others, and learn patterns from training data but they need a lot of training data. The knowledge driven-based methods simulate the diagnostic thinking of experts but they depend on expert knowledge. In none of the previous works has the diagnostic problem in energy systems been treated using a multi-label classification approach, for simultaneously diagnosing multiple problems. Multilabel classification is gained popularity in diagnostic applications as an efficient method for fault detection and monitoring [13, 18, 20, 21, 23].

In this work, we propose a multilabel algorithm, called LAMDA-ML, which is capable to assign individuals to several classes according to the adequacy degree of one individual to each class. The contribution of this work is the definition of the multilabel algorithm based on the degree of membership of an individual to each class. This algorithm is based on LAMDA-HAD, which is extended with this functionality, and it is capable to assign individuals to several classes according to their adequacy degree in each class.

This work is organized as follows: section 2 presents the related works. Section 3 introduces the fundamentals of LAMDA-HAD. Section 4 describes our multilabel proposal. Section 5 shows the experiments and presents the analysis of the results, and finally, Section 6 shows the conclusions and future works.

II. Related works

In this section, we present some recent works in multilabel classification. In the diagnosis medical field, Xia et al. [24] developed a multilabel approach, which through a stacked ensemble exploits the label correlations and the learning processes of ensemble members. For that, they made a weighted stacked ensemble with sparsity regularization to facilitate the classifier selection and the ensemble member's construction for multilabel classification. Then, the label correlations are considered for creating the weights of the ensemble members, and finally, an algorithm of optimization is made to achieve an optimal ensemble.

Li et al. [26] proposed an approach to multilabel classification based on the correlation of labels and instances. This is possible because they introduce regularization norms to simultaneously learn and label specifics. Then, they use a constrain label correlation on label outputs instead of the coefficient matrix, and finally, the correlations are used in a k-nearest neighbor mechanism to define the labels.

To improve the optimization of positive and negative labels in the multilabel settings of a picture, Ridnik et al. [6] developed an asymmetric loss ("ASL") metric, which operates differently on positive and negative labels. The loss metric enables down-weights and hard-thresholds of negative samples, and also, discards possibly mislabeled samples. This method has applicability in complex tasks like object detection in others.

Wehrmann et al. [7] developed a neural network architecture for Hierarchical Multilabel Classification, for optimizing local and global loss functions to discover local hierarchical classes. This proposal can be used in classification problems where the classes are hierarchically structured and objects can be assigned to multiple paths of the class hierarchy at the same time. For example, for the text classification, the image annotation.

In the work of Dinevaet al. [13], the authors proposed an approach for the detection of concurrent failures in the presence of disturbing noises or when the multiple faults cause overlapping features. They propose a Multi-label classification for simultaneously diagnosing multiple faults and evaluating the fault severity under noisy conditions of electrical machines using an Electrical Signature Analysis as well as traditional vibration data for modeling. Furthermore, the performance of various multi-label classification models is compared.

III. Fundamentals of LAMDA HAD algorithm for task classification

In this section, we will present briefly the basis of the LAMDA-HAD algorithm, which is based on our proposal (for major details, see Morales et al.'s work [16]). LAMDA is an algorithm based on fuzzy logic that assigns individuals to a class according to the Global Adequacy Degree. In LAMDA, the object/individual X is represented by a vector of descriptors:

$$X = [x_1;\ x_2;\ \ldots;\ x_j\ ;\ \ldots;\ x_n]$$

Where x_j: descriptor j of the object X.

Before using a LAMDA family's algorithm, it is necessary to standardize the individuals, according to the minimum and maximum values, as shown in Eq. (1):

$$\bar{x}_j = \frac{x_j - x_{jmin}}{x_{jmax} - x_{jmin}} \quad (1)$$

Where: $\bar{x}_j$: Standardized descriptor/feature; $\bar{x}_{jmin}$: Minimum Value of descriptor j; $\bar{x}_{jmax}$: Maximum value of descriptor j

Below, we present the main definitions of LAMDA.

Definition 1. Marginal Adequacy Degree (MAD): it determines how similar a descriptor of an individual is with respect to the same descriptor in a given class. For MAD calculation, density functions are used, and one of the most common is the Fuzzy Binomial function, shown in Eq. (2).

$$MAD(\bar{x}_j/\rho_{k,j}) = \rho_{k,j}{}^{\bar{x}_j}\ (1-\rho_{kj})^{(1-\bar{x}_j)} \quad (2)$$

Where: $\bar{x}_j$: standardized descriptor (see Eq. (1)); $\rho_{k,j}$: is the average value of the descriptor j that belongs to the class k, calculated using Eq. (3)

$$\rho_{kj} = \frac{1}{n_{kj}} \sum_{t=1}^{n_{kj}} \bar{x}_j(t) \quad (3)$$

Where: n_{kj} is the number of observations of class k and descriptor j.

Definition 1. Global Adequacy Degree (GAD): it determines the adequacy of an individual to each class. This value is calculated using MAD and can be determined according to Eq. (4):

$$\begin{aligned} GAD_{k,\bar{X}} = {} & (MAD_{k,1}, MAD_{k,2}, \ldots, MAD_{k,n}) \\ & = \alpha T(MAD_{k,1}, \ldots, \quad MAD_{k,n,}) \\ & + (1 \\ & - \alpha) S(MAD_{k,1,} \ldots MAD_{k,n}) \end{aligned} \quad (4)$$

Where $\alpha \in [0, 1]$ is the exigency parameter; T and S are the linear operators for a classes /clustering type [15, 16]. For example, example of T is $T(a,b) = \min(a,b)$.

Definition 3. Let $p = \{1, \ldots, m\}$ the number of existing classes in a dataset. The object $\bar{X}$ is assigned to the class with the maximum GAD, where the index corresponds to the number of the class.

$$index = \max(GAD_{1,\bar{X}}, GAD_{k,\bar{X}}, \ldots, GAD_{m,\bar{X}}, GAD_{NIC,\bar{X}})$$

NIC is used to create new classes after the training, when an object is unrecognized (it is sent to the NIC), making the algorithm more adaptive (online learning). It is considered $\rho_{NIC} = 0.5$, because, with this value in the probabilistic function (Eqs. (2)), the $MAD_{NIC} = 0.5$ for any value of the descriptor $\bar{x}_j$.

Below, we present the main definitions of LAMDA-HAD:

Definition 4: Let $MGAD_{k,p}$ the average of $GAD's$ of the individual in the class p in the class k:

$$MGAD_{k,p} = \frac{1}{n_k}\sum_{t=1}^{t=n_k} GAD_{p,t} \quad (5)$$

Where $MGAD_{k,p}$: Average of the Global Adequacy Degree of class k in class p; n_k is the number of objects belonging to class k, and $GAD_{p,t}$ is the GAD of the individual t for the class p, in the class k.

Definition 5: Let GAD_{NIC_p} the GAD of the NIC for the class p computed as:

$$GAD_{NIC_p} : \frac{1}{m}\sum_{p=1}^{p=m} MGAD_{k,p} \quad (6)$$

Definition 6: Let $AD_{GAD_{k,p,\bar{X}}}$ the new Global Adequacy Degree (GAD), which is a parameter that allows comparing the similarity between the GAD of an individual and each $MGAD_{k,p}$:

$$AD_{GAD_{k,p,\bar{X}}} = MGAD_{k,p}{}^{GAD_{p,\bar{X}}} (1 - MGAD_{k,p})^{\left(1-GAD_{p,\bar{X}}\right)} \quad (7)$$

Definition 7: The Highest Degree of Adequacy of an individual to a class $\left(HAD_{k,\bar{X}}\right)$ is determined by adding all the $AD_{GAD_{k,p,\bar{X}}}$ in class p:

$$HAD_{k,\bar{X}} = \sum_{p=1}^{m} AD_{GAD_{k,p,\bar{X}}} \quad (8)$$

Let E_I the class that the individual has the highest probability of belonging:

$$E_{I,\bar{X}} = max\left(HAD_{1,\bar{X}}, HAD_{2,\bar{X}}, \ldots HAD_{k,\bar{X}}, \ldots . HAD_{m,\bar{X}}\right) \quad (9)$$

Definition 8: Let index the value that identifies the class that an individual will be assigned, which is obtained by comparing the maximum value between E_I and the $GAD_{NIC_{E_I}}$:

$$index = max\left(HAD_{E_I,\bar{X},}, GAD_{NIC_{E_I}}\right) \quad (10)$$

Thus, it is verified if the maximum value of $HAD_{E_I,\bar{X}}$ is greater than the $GAD_{NIC_{E_I}}$ (the GAD_{NIC} adapted to each class).

Once the LAMDA-HAD algorithm is finished, the result will be the number of the class to which the individual is assigned; or otherwise, the individual is sent to the non-informative class (NIC).

IV. LAMDA-ML

In this section, we will describe the basis of our proposal. The fundamental idea is the assignment of individuals to several classes, for that we will define a threshold of membership, which will establish the classes that the individual will be assigned. In this section, we will describe the definitions of our proposal, through the next definitions.

Definition 9: Let G_t be a membership threshold. If an individual obtains a GAD greater than or equal to this value in a given class, then it will be labeled to this class. That implies that an individual can belong to different classes. This value G_t is obtained by calibration o just defined by the user

Definition 10: Let $B_{i,\bar{X},}$ a class to which an individual can belong, defined by Eq. 11:

$$B_{i,\bar{X}_r} = \quad HAD_{i,\bar{X}} > G_t \quad (11)$$

Definition 11: Let $L_{\bar{X}}$ the set of classes to which the individual will belong, defined by Eq. 12:

$$L_{\bar{X}} = \quad \{i, \forall i = 1, m \quad if\ HAD_{i,\bar{X}} > G_t\ and\ HAD_{i,\bar{X}} \quad > GAD_{NIC_i}\} \quad (12)$$

Definition 11: Let indexes the set that identifies the classes,

$$indexes = \begin{cases} L_{\bar{X}} & if\ L_{\bar{X}}\ is\ not\ empty \\ new\ class & else \end{cases} \quad (13)$$

Macro algorithm 1 shows the LAMDA-ML algorithm.

MACRO ALGORITHM 1. PSEUDO CODE OF LAMDA-ML

Procedure
1. *Input(x_i)*
2. *Run LAMDA-HAD until step 5 according to Macro algorithm 2*
    a. *Find the indexes of the classes of the individual using Eq. (13).*

End
Output: Class Indexes

And Macro algorithm 2 shows the LAMDA-HAD algorithm.

MACRO ALGORITHM 2. PSEUDO CODE OF LAMDA-HAD

Input $X = \left((x_1), \dots (x_j) \dots, (x_n)\right)$

Procedure

1. *Normalize the descriptors of the individual using Eq. (1).*
2. *Calculate the MAD for each descriptor and each class based on the selected probability density function described by Eq (2).*
3. *Calculate the GAD of each class using Eq. (4).*
4. *Calculate the MGAD in each class through Eq. (5) of Definition 3.*
5. *Calculate de* $GADNIC$ *for each class k. Eq(6)*
6. *Calculate the adequacy degree* $AD_{GAD_{k,p,\bar{X}}}$ *with Eq. (7), and the HAD as the sum of them with Eq. (8).*
7. *Find the estimation index* $E_{I,\bar{X}}$ *to the class to which the individual could belong with Eq. (9).*
8. *In the estimated class determine if the maximum GAD is greater than the corresponding* $GADNIC$

End

Output: Class Index

V. EXPERIMENTS

A. Experimental Protocol

Energy Context

Table 1 shows the dataset for evaluating the performance of our proposal. This dataset contains 2920 records of different features of solar energy generation, such as distance to solar noon, temperature, wind direction, wind speed, average pressure, sky cover, visibility, humidity, average wind speed, average pressure, and power generated. This last one is conformed into 3 classes, and is the classification target of this work.

TABLE 1. DATASET FOR EVALUATING THE MULTILABEL TASK

Dataset	Size	Features	Characteristics
Power Generated	2920	10	3 classes unbalanced

The dataset is formed by 10 variables, and has 3 classes unbalanced. 80% of the data is chosen for the training, and 20% is later for validation.

Particularly, the goal is the classification of the power generated by a solar plant. Nowadays, it is necessary to determine the ideal conditions for the installation of solar cells to obtain the maximum amount of energy out of them. With this dataset, we can build a classification model to classify the power output for a particular array of solar power generators, knowing the environmental conditions of the context. This dataset can be downloaded from [28].

Metrics for Evaluating the Multilabel Task

For the evaluation of the experiments, the next metrics have been used, which are normally used in the context of multilabel classification problems [5, 13, 25, 26, 27].

Accuracy (Acc): Proportion of individuals correctly classified, determined by Eq. (14):

$$ACC = \frac{1}{N}\sum_{i=1}^{N} I(Z_i = Y_i) \tag{14}$$

Where: Z_i: the set of labels in the dataset in the test stage; Y_i: the set of labels predicted by the model; N: number of observations in the test stage; I: function indicator ($I(True) = 1 \; and \; I(False) = 0$).

Precision (P): it is the proportion of correct predictions among all predictions of a certain class, determined by Eq. (15):

$$P = \frac{1}{N}\sum_{i=1}^{N} \frac{|Y_i \cap Z_i|}{|Y_i|} \tag{15}$$

Exact Match Radio (EMR): is an extension of accuracy for single-labels to a multi-label classification problem. It is calculated by Eq. (16):

$$EMR = \frac{1}{N}\sum_{i=1}^{N} \frac{|Y_i \cap Z_i|}{|Y_i|} \tag{16}$$

F1-macro L: This metric evaluates the performance by classes. It is calculated by Eq. (17):

$$F_1 \; macro \; L = \frac{1}{L}\sum_{i=1}^{L} F_1(Y_i, Z_i) \tag{17}$$

Where F_1: F Measure metric, L: numbers of elements of L class. F1 is:

$$F_1 = \frac{1}{N}\sum_{i=1}^{N} \frac{2|Y_i \cap Z_i|}{|Y_i| + |Z_i|}$$

And in this case, to calculate F1 then N=1

B. Result Analysis

Figure 1 shows a perceptual map of the dataset studied in this work. In this case, we observe 3 classes marked with different colors. Also, it is highlighted that the class marked with red color is a class that is overlapped with the blue and purple classes. So, some individuals of the red class also belong to the other classes. This is a typical case of a multilabel problem, where the classes aren't mutually exclusive.

This dataset originally is not ready for ML applications. So, before using the LAMDA-ML algorithm, we have developed a data preprocessing, which consists of the next steps: i) Detection and deletion of missing values; ii) Detection and deletion of outliers; iii) Execution of a feature engineering process for selection of variables and determination of classes in the dataset [29]. This process includes a Principal Component Analysis (PCA); iv) Labeled of individuals according to the PCA.

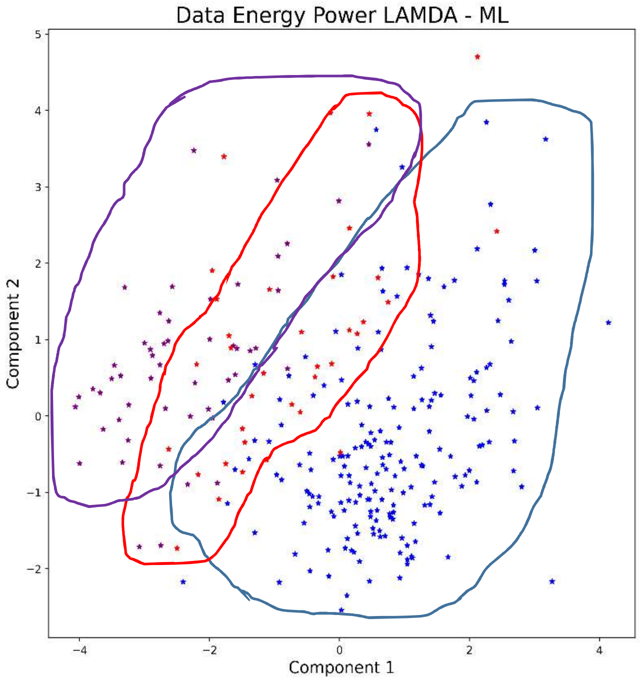

FIGURE 1. PERCEPTUAL MAP OF DATASETS STUDIED POWER GENERATED

Table 2 shows the results of the metrics. We want to highlight that in a multilabel context, we can compute the performance by classes. In this sense, we can observe that in the accuracy metric, the best performance is for the blue class, which is the densest class. With respect to the red class, which is the overlapped class with the other classes, the performance is the lowest for all the metrics. So, in this case, the behavior may be happening by the Gt values, which in this experiment is 0.15. In the purple class, we can observe that the performance is acceptable. Finally, if we take into account the average values of all the metrics in Table 2, we can observe that the best performance is with P and ACC, which means that the proportion of labels identified is acceptable according to the individuals in the training stage. In general, LAMDA-ML has a deficient performance for the overlapped class.

TABLE 2. METRICS FOR THE MULTILABEL PROBLEM OF THE SOLAR POWER GENERATED FOR GT=0.15

Metrics / Classes	**ACC**	**P**	**EMR**	**F1-Macro L**
Class Blue	0.89	0.85	0.78	0.80
Class Red	0.72	0.82	0.68	0.73
Class Purple	0.82	0.8	0.77	0.74
Average	0.81	0.82	0.74	0.76

In Table 3 are shown the results for different values of Gt. We can see that LAMDA-ML is very sensitive to the Gt values. For high values. the classification is very demanding. so the individuals will only be classified into a few classes and a few new classes are created. LAMDA-ML tries to classify all in the existing classes, tending to a unilabeled behavior (multi-classification). In the case of low values of Gt two things happen: new classes are eventually generated due to the NIC value. or existing classes are updated with less demand. which means that an individual can easily have several labels. Thus. for low Gt's. the multilabel quality metrics are better. In addition. in general. they are quite similar for values of Gt less than 0.5.

TABLE 3. METRICS FOR THE MULTILABEL PROBLEM OF THE GENERATED SOLAR POWER FOR DIFFERENT GT'S

Metrics /G_t	**ACC**	**P**	**Exact – Match**	**F1-Macro L**
0.15	0.81	0.82	0.74	0.77
0.25	0.82	0.84	0.75	0.8
0.5	0.8	0.8	0.73	0.76
0.75	0.71	0.78	0.71	0.7
0.9	0.68	0.72	0.7	0.68

VI. CONCLUSIONS

In this work. we have proposed a new algorithm for the LAMDA family's algorithms. called LAMDA-ML. for the multilabel context. This algorithm has been tested in a dataset to study solar energy generation. Particularly. this implementation is focused on a multilabel situation in the energy field. with interesting initial results. Our approach has enabled the discovery of hidden behavior in the solar energy generation context (multiple states of overlapping power generation). useful for future diagnostic models.

The definition of this multilabel approach from the LAMDA-HAD algorithm has been very natural. The process to extend the LAMDA-HAD algorithm for the multilabel context is not expensive and is easy. We have exploited the characteristics of the LAMDA algorithms that allow determining the degree of membership of an individual in each class. and depending on whether that value is high or not (determined by the Gt parameter). each individual is labeled with that class. therefore. allows it to belong to multiple classes.

For future works. it is necessary to design a methodology for calibrating the parameter Gt. which initially is introduced by the user. with the aim to improve the performance of the LAMDA-ML algorithm. Also. this algorithm must be tested in other types of real-world multilabel problems. like text categorization. medical diagnosis. among other domains.

ACKNOWLEDGMENT

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under the Marie Skłodowska Curie Grant Agreement No. 754382. M.D. R-Moreno is supported by the JCLM project SBPLY/19/180501/000024 and the Spanish Ministry of Science and Innovation project PID2019-109891RB-I00. both under the European Regional Development Fund (FEDER).

REFERENCES

[1] D. Abdulqader. A. Abdulazeez. D. Zeebaree. "Machine learning supervised algorithms of gene selection: A review". Machine Learning. vol. 62. 2020.

[2] S. Uddin. A. Khan. M. Hossain. M. Moni. “Comparing different supervised machine learning algorithms for disease prediction”. BMC medical informatics and decision making. vol. 19. 2019.

[3] S. Carta. A. Corriga. R. Mulas. D. Recupero. R. Saia. “A Supervised Multi-class Multi-label Word Embeddings Approach for Toxic Comment Classification”. In Proc. 11th International Joint Conference on Knowledge Discovery. Knowledge Engineering and Knowledge Management. (pp. 105-112). 2019).

[4] N. Hameed. A. Shabut. M. Ghosh. M. Hossain. “Multi-class multi-level classification algorithm for skin lesions classification using machine learning techniques”. Expert Systems with Applications. vol. 141. 2020.

[5] A. Lagopoulos. A. Fachantidis. G. Tsoumakas. “Multi-label modality classification for figures in biomedical literature”. In Proc. IEEE 30th International Symposium on Computer-Based Medical Systems (pp. 79-84). 2017.

[6] T. Ridnik. E. Ben-Baruch. N.. Zamir. A.. Noy. I. Friedman. M. Protter. L. Zelnik-Manor. “Asymmetric loss for multi-label classification”. In Proc IEEE/CVF International Conference on Computer Vision (pp. 82-91). 2021.

[7] J. Wehrmann. R. Cerri. R. Barros. “Hierarchical multi-label classification networks. In Proc. International Conference on Machine Learning (pp. 5075-5084). 2018.

[8] Q. Chen. A. Allot. R. Leaman. R. Doğan. Z. Lu. “Overview of the BioCreative VII LitCovid Track: multi-label topic classification for COVID-19 literature annotation”. In Proc seventh BioCreative challenge evaluation workshop. 2021.

[9] B. Ferreira. J.. Costeira. J. Gomes. “Explainable Noisy Label Flipping for Multi-Label Fashion Image Classification”. In Pro.e IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 3916-3920). 2021.

[10] A. Ruiz. P. Thiam. F. Schwenker. G. Palm. “A k-Nearest Neighbor Based Algorithm for Multi-Instance Multi-Label Active Learning”. In Proc. IAPR Workshop on Artificial Neural Networks in Pattern Recognition (pp. 139-151). 2018.

[11] M. Wever. A. Tornede. F. Mohr. E. Hullermeier. “AutoML for Multi-Label Classification: Overview and Empirical Evaluation”. IEEE Transactions on Pattern Analysis & Machine Intelligence. vol.1. 2021.

[12] M. Bello. G. Nápoles. R. Sánchez. R. Bello. K. Vanhoof. “Deep neural network to extract high-level features and labels in multi-label classification problems”. Neurocomputing. vol. 413. pp. 259-270. 2020.

[13] A. Dineva. A. Mosavi. M. Gyimesi I.Vajda N. Nabipour. T. Rabczuk “Fault Diagnosis of Rotating Electrical Machines Using Multi-Label Classification”. Applied Sciences. vol. 9. 2019.

[14] T. Kempowsky. A. Subias. J. Aguilar-Martin. “Process situation assessment: From a fuzzy partition to a finite state machine”. Engineering Applications of Artificial Intelligence.vol. 19. pp. 461-477. J. 2006.

[15] L. Morales. J. Aguilar. "An Automatic Merge Technique to Improve the Clustering Quality Performed by LAMDA." IEEE Access. vol. 8. pp. 162917-162944. 2020.

[16] L. Morales. J. Aguilar. D. Chavez. C. Isaza. “LAMDA-HAD. an extension to the Lamda classifier in the context of supervised learning”. International Journal of Information Technology & Decision Making. vol. 19. pp. 283-316. 2018.

[17] J. Waissman R. Sarrate T. Escobet J. Aguilar B. Dahhou “Wastewater treatment process supervision by means of a fuzzy automaton mode”l. In Proc. IEEE International Symposium on Intelligent Control (pp. 163-168). 2000.

[18] J. Aguilar A. Garces-Jimenez M. R-Moreno R. García “A systematic literature review on the use of artificial intelligence in energy self-management in smart buildings”. Renewable and Sustainable Energy Reviews. vol. 151. 2021.

[19] L. Hernández. A. Hernández-Callejo. Q.. Zorita-Lamadrid. F. Duque-Pérez. A. Santos García. “Review of strategies for building energy management system: Model predictive control. demand side management. optimization. and fault detect & diagnosis”. Journal of Building Engineering. vol. 33. 2021.

[20] Z. Yang. C. Zhang Y. Zhang Z. Wang J. Li. “A review of data mining technologies in building energy systems: Load prediction. pattern identification. fault detection and diagnosis”. Energy and Built Environment. vol. 1. pp. 149-164. 2020.

[21] Y. Zhao. T. Li. X. Zhang. C. Zhang. “Artificial intelligence-based fault detection and diagnosis methods for building energy systems: Advantages. challenges and the future". Renewable and Sustainable Energy Reviews. vol. 109. pp. 85-101. 2019.

[22] M. Araujo J. Aguilar H. Aponte “Fault detection system in gas lift well based on artificial immune system”. In Proc. International Joint Conference on Neural Networks (pp. 1673-1677). 2003.

[23] T. Wang J. Wang Z. Fan X. Wei M. Pérez Jiménez T. Zang T. Huang “Fault diagnosis for multi-energy flows of energy internet: Framework and prospects”. In Proc. IEEE Conference on Energy Internet and Energy System Integration. 2017.

[24] Y. Xia. K. Chen. Y. Yang. “Multi-label classification with weighted classifier selection and stacked ensemble”. Information Sciences. vol. 557. pp. 421-442. 2021.

[25] J. Aguilar, “A general ant colony model to solve combinatorial optimization problems”, Rev. colomb. comput., vol. 2, pp. 7-18, 2001.

[26] J. Li. P.. Li. X. Hu. K. Yu. “Learning common and label-specific features for multi-Label classification with correlation information”. Pattern Recognition. vol. 121. 2022.

[27] S. Vluymans. C. Cornelis. F. Herrera. Y. Saeys. “Multi-label classification using a fuzzy rough neighborhood consensus”. Information Sciences. vol. 433. pp. 96-114. 2018.

[28] neuraldesigner. (sf). “Explainable AI Platform”. Available from: https://www.neuraldesigner.com/

[29] F. Pacheco C. Rangel J. Aguilar M. Cerrada A. Altamiranda “Methodological framework for data processing based on the Data Science paradigm" In Proc. XL Latin American Computing Conference. 2014.