Journal of Universal Computer Science, vol. 30, no. 1 (2024), 130-150
submitted: 8/8/2022, accepted: 12/9/2023, appeared: 28/1/2024

Dimensionality Reduction for Hierarchical Multi-Label Classification: A Systematic Mapping Study

Raimundo Osvaldo Vieira
(Universidade Tecnológica Federal do Paraná, Ponta Grossa, Paraná
Instituto Federal de Educação, Ciência e Tecnologia do Maranhão, São Luís, Maranhão
https://orcid.org/ 0000-0002-7086-749X, raimundo.vieira@ifma.edu.br)

Helyane Bronoski Borges
(Universidade Tecnológica Federal do Paraná, Ponta Grossa, Paraná
https://orcid.org/ 0000-0002-9153-3819, helyane@utfpr.edu.br)

Abstract: Hierarchical multi-label classification problems typically deal with datasets with many attributes and labels, which can negatively impact the classifier performance. The application of dimensionality reduction methods can significantly improve the performance of classifiers. Dimensionality reduction can be performed by feature extraction or feature selection, according to the problem domain and datasets characteristics. This work carried out a systematic literature mapping to identify the approaches and techniques of dimensionality reduction that have been used in hierarchical multi-label classification tasks. Searches were performed on 7 important databases for the Computer Science field. From a list of 184 retrieved papers, 12 were selected for analysis, from which it was possible to determine a general overview of studies conducted from 2010 to 2022. It was identified that feature selection was the most frequent reduction method, with filter approach standing out. In addition, it was detected that most of the works used tree hierarchical structure. As its main outcome, this paper presents the state of the art of dimensionality reduction problem for hierarchical multi-label classification, indicating trends and research issues in the field.

Keywords: Hierarchical Multi-label Classification, Dimensionality Reduction, Feature Selection, Feature Extraction
Categories: I.2.0, I.2.5, I.2.6
DOI: 10.3897/jucs.91309

1 Introduction

The classification problem is one of the most important in the field of machine learning and it consists of assigning a label to an element of the target dataset of the classification task. The hierarchical multi-label classification (HMC) corresponds to a variation of the usual classification task in which an example can belong to more than one class simultaneously, and the classes maintain a hierarchical structure among themselves [Melo, 19].

This type of problem is common in document categorization, classification of images and musical genres, and predicting protein functions, among others [Silla Jr., 11]. In general, this type of application has a large volume of data and a high number of attributes and labels, which negatively interferes with the performance of the classifier. To deal with this limitation, it is necessary to use techniques to reduce the dimensionality of the data without changing its intrinsic meaning.

The application of dimensionality reduction techniques can improve the classification, visualization, and data compression processes [Van Der Maaten, 09], additionally, it incorporates benefits such as the reduction of noise and redundancy in attributes, promoting an increase in the learning potential. The most-known approaches to deal with the dimensionality reduction problem are the selection and extraction of attributes [Borges, 12]. The choice between the two approaches relies on the application domain and its dependence on the training data available. It justifies the relevance of this work, whose objective is to present an overview of the studies already carried out on the theme of dimensionality reduction for the classification multi-label hierarchy.

This paper describes the performance of a systematic literature mapping on the application of dimensionality reduction techniques in HMC problems. The works developed between 2010 to 2022 and that are available in the IEEE Xplore, Scopus, Science Direct, Springer, InderScience, ArXiv, and Emerald repositories were identified. Through mapping, we sought to identify the most used approaches and techniques for dimensionality reduction and analyze the results presented in each of the selected works.

The rest of this document is organized as follows: sections 2 and 3 are intended for the theoretical background of the work, covering, respectively, the basic concepts of hierarchical multi-label classification and dimensionality reduction. Section 4 describes the systematic mapping method used. Section 5 presents the results obtained and, in section 6, the conclusions of the work.

2 Hierarchical Multi-label Classification

According to [Vens, 08], [Carvalho, 11] and [Stojanova, 13], the HMC task can be formally defined as finding the function $f: X \longrightarrow \wp(C)$ which associates each instance to a set of classes $C_j \in \wp(C)$, where X is the space of instances, $\wp(C)$ is the power set of C and $C = \{c_1, c_2, \ldots, c_k\}$ is the set of all possible class labels, which, in turn, are hierarchically organized according to a partial order $\preccurlyeq_h$, which represents the superclass relationship, that is, $\forall c_1, c_2 \in C : c_1 \preccurlyeq_h c_2$ if c_1 and only if it is a superclass of c_2. So, given a set of examples T, where each example has the form $t_i = (x_i, C_i)$, the function must be such that $c \in f(x) \Rightarrow \forall c' \preccurlyeq_h c : c' \in f(x)$.

In the HMC, when an instance receives a label from a class, it is also classified as being of the type of all predecessor classes to the class from which it received the label. For instance, in Figure 1, an example that receives the label Class A1 also receives the label Class A, since it is a hierarchy of classes.

HMC problems can be characterized from three basic aspects performed [Silla Jr., 11]: 1) the type of hierarchy used to implement the relationship between classes; 2) whether the data can follow only one or more than one path in the hierarchy; 3) and the hierarchical level where the predictions are. Regarding the first aspect, the relationship between the classes can be represented as a tree or as a Directed Acyclic Graph (DAG). Figure 1 and Figure 2 show these two types of structures. The essential difference between the tree structure and the DAG structure lies in the fact that in a tree (Figure 1) each node, except the root, has one, and only one, parent node, while in the DAG structure (Figure 2), each node other than the root can have more than one parent node.

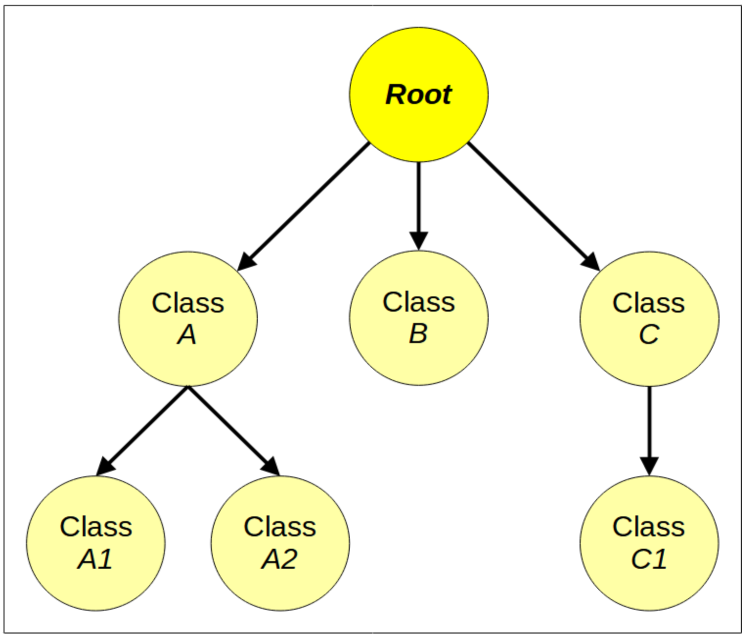

Figure 1: Class structure in tree hierarchy format

In the second aspect, the labeling path, there are cases where predictions can follow more than one path in the hierarchy. For example, an instance can belong simultaneously to classes *A1* and *C1* of the structure in Figure 2. In this case, there is specifically a hierarchical multi-label classification problem.

The third aspect, the hierarchical level where predictions are performed, corresponds to the depth of data labeling. In this case, the classification of new instances can be complete when the classification takes place in a leaf node or partial, when the classification occurs at higher levels [Faceli, 11].

There are works that propose methods for handling HMC tasks in different domains such as bioinformatics [Melo, 19], text classification [Gargiulo, 19], image classification [Dimitrovski, 10], among others. However, there is no consensus on which approach to use to deal with HMC problems. In this context, some classification algorithms were proposed, highlighting the Clus-HMC [Blockeel, 02], the HMC-GA (*Hierarchical Multi-Label Classification with Genetic Algorithm*) [Cerri, 18] and the MHC-CNN (*Multi-label Hierarchical Classification using the Competitive Neural Network*) [Borges, 12].

The performance of classification methods for global hierarchical classification problems depends on the number of examples, attributes and classes. Often, such problems present high dimensionality of attributes, scarcity of training examples, and variation in the number of classes to which each example belongs. In this case, dimensionality reduction methods are necessary to overcome such barriers.

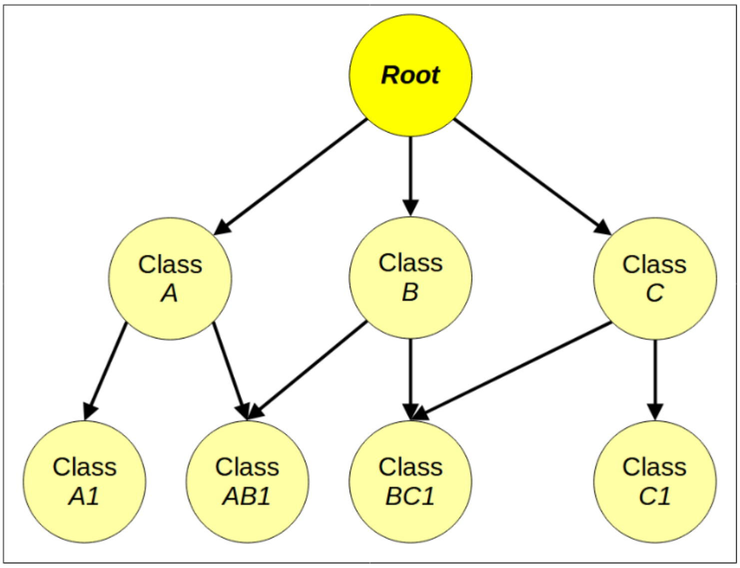

Figure 2: Class structure in DAG hierarchy format

3 Dimensionality Reduction

Dimensionality reduction consists of reducing the number of attributes, labels, or both to improve the performance of classifiers [Ghodsi, 06]. In other words, finding a meaningful representation in reduced dimensionality for high-dimensional data, keeping a minimum number of parameters that preserve the observed properties in the data. Dimensionality reduction can be performed by selecting attributes or extracting attributes.

The selection of attributes consists of identifying the relationships between them in a dataset and choosing the most significant to compose a simplified dataset capable of producing results equal to or very close to those obtained from the analysis of the complete set of data for a given task. The main approaches to attribute selection are filter, wrapper, and inline [Kumar, 14].

Attribute extraction is a process that creates new features from the original dataset by transforming or combining features from the original set [Ghodsi, 06]. These new features tend to be more expressive and better represent the variability of the data. For [Jain, 2000], given a feature space of dimension n, the methods of feature extraction determine an appropriate subspace of dimensionality m such that $m < n$. Several techniques can be applied to reduce dimensionality through attribute extraction: *Principal Component Analysis* (PCA) [Ghodsi, 06], Multidimensional Scaling – MDS [Kruskal, 64], *Self Organizing Map* – SOM [Kohonen, 90].

In HMC problems, the excessive number of attributes in many domains hinders the extraction of knowledge from the dataset. Therefore, the application of dimensionality reduction techniques becomes necessary in this context [Borges, 12], as it provides an increase in the generalization capacity of machine learning methods.

The choice between adopting the attribute selection or extraction approach depends on the problem domain and the characteristics of the training dataset. In this context, studies were conducted in different application domains. It is necessary to know the approaches and techniques applied, as well as the results obtained. A systematic literature mapping makes it possible to know such works and their results, which provides an overview of the studies developed in the research area in question. Section 4 describes the systematic mapping method used in this work.

4 Systematic Mapping

This section presents the bibliographic survey aimed at establishing the current state of the art regarding dimensionality reduction techniques in hierarchical multi-label classification databases.

4.1 Description of the Systematic Mapping Method

The proposed method for performing the systematic mapping was inspired by the protocols developed by [Rattan, 13] and [Pagani, 15], and it has the following steps: 1) planning; 2) execution; and 3) data extraction and results. Figure 3 presents an overview of the method and the recommended sequence of steps.

In the planning step, the research topic and mapping objectives are defined in addition, the research questions to be answered are elaborated. Research questions should align with the intent of the research and guide the extraction of information during mapping. In addition, in this step, the inclusion and exclusion criteria must be defined for the selection of works to be considered in the mapping. The time interval to be considered in the searches must also be defined, in addition to the area and language of publication. These last three criteria make it possible to carry out a previous selection of papers during the execution step, using resources made available by the search bases. As they are used as parameters during searches, these criteria will be called search criteria.

Once the mapping objectives, research questions, and inclusion and exclusion criteria have been defined, the proposed method leads to the execution step. This step encompasses the following activities: definition of search bases, the definition of keywords, definition of search strings, carrying out searches in the bases, and selection of works retrieved from the inclusion and exclusion criteria.

The definition of the search bases consists of selecting the data sources to conduct the searches. The bases chosen must be the most appropriate, considering that all the papers to be used in the following steps are extracted from these sources.

The definition of keywords corresponds to the choice of terms relevant to the research topic and the mapping objectives. The keywords form the basis of search strings. A search string is formed by one or more keywords linked through logical conjunction, disjunction, or negation operations. Each string is also a search criterion and must be used to perform searches together with the search criteria already defined.

Conducting searches refers to the retrieval of studies that meet the established search criteria. Such criteria, in turn, may have their syntax adapted to meet the format required by each of the chosen search bases.

Once the searches have been conducted, the selection of works proceeds by choosing the studies that are in alignment with the purposes of the work. This choice complies with the inclusion and exclusion criteria established in the initial planning step. In this search step, bibliography management software can be used to facilitate the control and organization of studies.

In the extraction of data and results step, the selected papers are read and analyzed, based on the research questions elaborated in the initial planning step. Then, each question is answered, and the information that has been inferred from reading the papers is presented. Finally, the mapping results are compiled and discussed, generating a systematic literature mapping document.

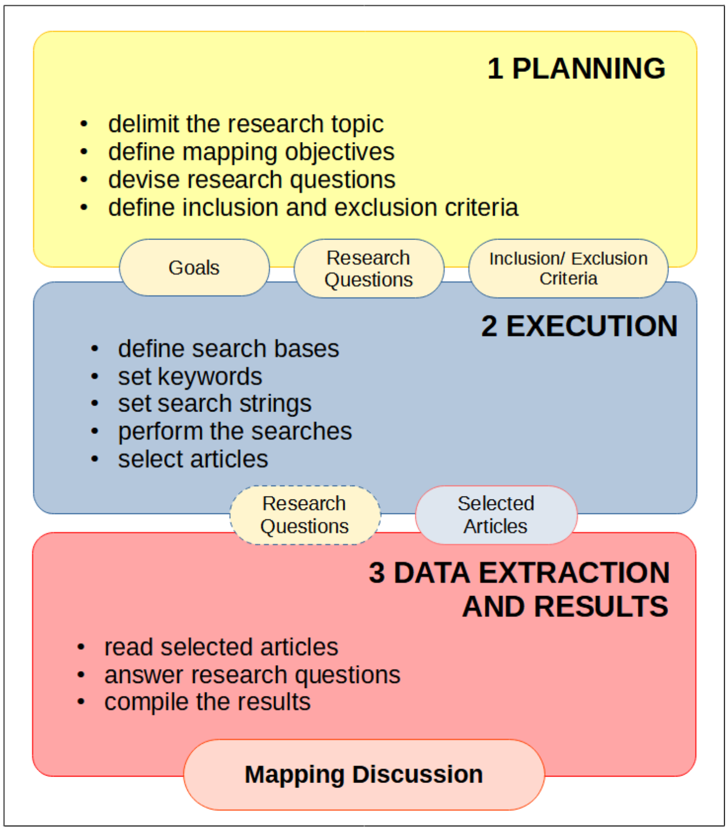

Figure 3: Overview of the systematic mapping method

In the following subsections, the application of the proposed method in the context of this work is described.

4.2 Planning

The research topic for this work is “Dimensionality Reduction in Hierarchical Multi-label Classification Databases”. In this sense, the main objective of this study is to identify the main applications of dimensionality reduction methods in hierarchical multi-label classification problems. Thus, to guide the data extraction, four research questions were elaborated, which are identified by Q_i, where $i = 1, 2, 3, 4$.

- Q_1: What are the methods (selection or extraction) and dimensionality reduction approaches/techniques used in hierarchical multi-label classification?
- Q_2: What were the areas where dimensionality reduction was applied? What is the format of the hierarchical structure of the classes (tree or DAG)?
- Q_3: What were the scientific contributions of the authors in the works to the problem of dimensionality reduction?
- Q_4: The result obtained by the reduced dataset was relevant when compared to the dataset formed by all attributes?

As an inclusion criterion, it was defined that only works published in conferences and journals, whose full version is available in the researched sources, should be considered. The exclusion criteria used are the exclusion of duplicate items, as they can eventually be retrieved from different databases, and the exclusion of studies whose title or abstract are not compatible with the topic in question. In addition, these search criteria were adopted: the publication period is from 2010 to 2022, publications are from the large area of Computer Science and in the English language.

4.3 Execution

For the search, the following databases were defined: IEEE Xplore, Scopus, Science Direct, Springer, Inderscience, arXiv and Emerald. These databases were chosen because it is possible to search for publications in Computer Science, therefore, to retrieve the highest possible number of documents in the research area of this work.

In order to obtain accurate search results within the researched topic, a set of keywords was defined, in English, related to the research questions. The keywords chosen are the following: Hierarchical Multi-label Classification, Dimensionality Reduction, Feature Extraction and Feature Selection.

Search strings match combinations of these keywords, using the logical operators AND and OR. Variations of these terms were used to improve the quality of the results obtained, namely: Hierarchical Multi-label Classification, Attribute Extraction, and Attribute Selection are, respectively, variations of the keywords listed above, except for the term Dimensionality Reduction. After some refinements during string calibration, aiming for the best trade-off between precision and accuracy, the strings identified by S_i $(i = 1, 2, 3)$ were obtained. Search strings are as follows:

- S_1: ("*Hierarchical Multi-label Classification*" OR "*Hierarchical Multi-label Classification*") AND "*Dimensionality Reduction*"
- S_2: ("*Hierarchical Multi-label Classification*" OR "*Hierarchical Multi-label Classification*") AND ("*Feature Extraction*" OR "*Attribute Extraction*")
- S_3: ("*Hierarchical Multi-Label Classification*" OR "*Hierarchical Multi-Label Classification*") AND ("*Feature Selection*" OR "*Attribute Selection*")

In this step, additionally to the search strings, the search criteria already defined were used to carry out searches in the selected databases. The result in the search bases, using the defined search strings and the adopted selection criteria, is presented in Table 1.

At the end of this stage, a total of 184 works were retrieved. They were imported into the *Zotero* bibliography manager (https://www.zotero.org/), so that a better management of the results obtained through the resources made available by the tool.

Search database	**S_1**	**S_2**	**S_3**	**Total**
IEEE Xplore	02	32	06	40
Scopus	00	08	05	13
Science Direct	03	14	17	34
Springer	28	30	38	96
InderScience	00	00	01	01
ArXiv	00	00	00	00
Emerald	00	00	00	00
Totals	**33**	**84**	**67**	**184**

Table 1: Total search results

The paper selection process was conducted based on the inclusion and exclusion criteria defined in the initial planning step. Initially, duplicates were eliminated by joining results. Then, since the exclusion criteria defined that only papers published in conferences or journals are considered, the publications of books and book chapters were excluded. This procedure was performed in an automated way using the Zotero tool, resulting in 117 papers.

Subsequently, papers whose title and abstract were not directly related to the topic of this mapping were discarded, producing a total of 12 papers selected for reading and data extraction. The application of the data extraction and results obtained step is described in section 5.

5 Results

This section shows the results obtained from the data extraction performed on the studies selected for analysis, consisting of the presentation of the answers obtained for the questions defined in subsection 4.2. The authors and titles of the selected papers are shown in Table 2. An identification number (Paper ID) was adopted to facilitate the citation of the papers.

Paper ID	**Authors**	**Title**	**Year**
1	Dimitrovski, I.; Kocev, D.; Loskovskaya, S.; Džroski, S.	Detection of visual concepts and annotation of images using ensembles of trees for hierarchical multi-label classification	2010
2	Dimitrovski, I.; Kocev, D.; Loskovskaya, S.; Džroski, S.	Hierarchical annotation of medical images	2011

3	Dimitrovski, I.; Kocev, D.; Loskovskaya, S.; Džroski, S.	Hierarchical classification of diatom images using ensembles of predictive clustering trees.	2012
4	Slavkov, I.; Karcheska, J.; Kocev, D.; Kalajdziski, S.; Džroski, S.	ReliefF for hierarchical multi-label classification	2014
5	Yan, S.; Wong, K. C.	Elucidating high-dimensional cancer hallmark annotation via enriched ontology	2017
6	Cerri, R.; Mantovani, R. G.; Basgalupp, M. P.; Carvalho, A. C. P. L. F.	Multi-label Feature Selection Techniques for Hierarchical Multi-label Protein Function Prediction	2018
7	Slavkov, I.; Karcheska, J.; Kocev, D.; Kalajdziski, S.; Džroski, S.	HMC-*ReliefF*: Feature Ranking for Hierarchical Multi-label Classification	2018
8	Melo, A.; Paulheim, H.	Local and global feature selection for multilabel classification with binary relevance	2019
9	Prabowo, F. A.; Ibrohim, M. O.; Budi, I.	Hierarchical Multi-label Classification to Identify Hate Speech and Abusive Language on Indonesian Twitter	2019
10	Huang, H.; Liu, H.	Feature selection for hierarchical classification via joint semantic and structural information of labels	2020
11	Aljedani, N.; Alotaibi, R.; Taileb, M.	HMATC: Hierarchical multi-label Arabic text classification model using machine learning	2021
12	Silva, L; Cerri, R.	Feature Selection for Hierarchical Multi-label Classification	2021

Table 2: Selected papers

Regarding the frequency of publication of the works selected for the study, it can be seen from the analysis of Table 1 that there were few publications per year within the time interval considered in this study. The number of publications per year is presented in a compiled form in Figure 4. It is noteworthy that only in the years 2018, 2019, and 2021 there was the publication of more than one paper.

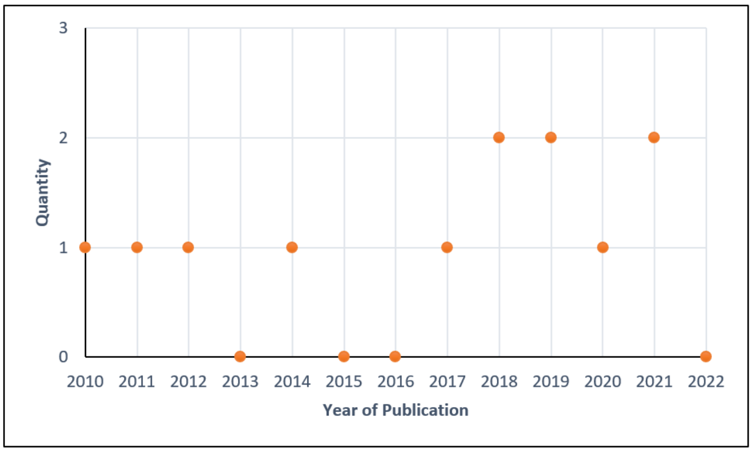

Figure 4: Number of publications per year

5.1 Q_1: What are the methods (selection or extraction) and dimensionality reduction approaches/techniques used in hierarchical multi-label classification?

Of the 12 works analysed in this mapping, 4 use the attribute extraction method and 8 use attribute selection. The techniques adopted in each work are listed in Table 2, making a total of 30 dimensionality reduction techniques, of which 10 are attribute extraction techniques and 20 are feature selection techniques. Regarding the selection of attributes, the approach used was also identified, with 1 work using the wrapper approach and 7 works using the filter approach. None of the selected works applied the embedded approach.

Figure 5 and Figure 6 illustrate two word clouds with terms referring to the approaches and techniques used in the selected works.

Regarding the extraction of attributes, as can be concluded from Figure 5, the Scale-Invariant Feature Transform (SIFT) technique was the most used (three works). Each of the other techniques was adopted in only one work.

The most prominent term in Figure 6 (Filter) refers to the attribute selection approach adopted in the largest number of works. As already mentioned, 7 studies used this approach, corresponding to 87.5% of the total of papers that use attribute selection as a dimensionality reduction technique. Regarding the techniques adopted, the *ReliefF*, *Binary Relevance* and *Label Powerset* transformation techniques stand out. *Binary Relevance* and *Label Powerset* were used in 2 works in combination with techniques for resources ranking, as shown in Table 3. For resource ranking the *ReliefF* technique stands out, having been adopted in 3 works, and in 2 works it was used in combination with other techniques. Considering that the *ReliefF* measure is a variation of the *Relief* technique, it can be said that 50% of the selected works adopted this technique for feature selection.

Paper ID	Method	Technique/ Approach
1	Extraction	Scale-Invariant Feature Transform (SIFT) Harris-Laplace detector
2	Extraction	Raw Pixel Representation (RPR) Local Binary Patterns (LBP) Edge Histogram Descriptors (EHD) Scale-Invariant Feature Transform (SIFT)
3	Extraction	Fourier Descriptors (FD) Scale-Invariant Feature Transform (SIFT)
4	Selection	HMC-ReliefF/ Filter
5	Selection	United Decision Tree (UDT)/ Filter United GSS Coefficient (UGSS)/ Filter United NGL Coefficient (UNGL)/ Filter
6	Selection	Clus-HMC/ Wrapper
7	Selection	HMC-ReliefF/ Filter
8	Selection	Information Gain (IG)/ Filter
9	Extraction	Frequency term word n-grams Character n-grams
10	Selection	Semantic and Structural Information (FSSS)/ Filter
11	Selection	Binary Relevance with Chi-Square (BR-χ^2)/ Filter Label Powerset with Chi-Square (LP-χ^2)/ Filter Binary Relevance with Gain Ratio (BR-GR)/ Filter Label Powerset with Gain Ratio (LP-GR)/ Filter Binary Relevance with RelieF (BR-RF)/ Filter Label Powerset with RelieF (LP-RF)/ Filter Binary Relevance with Information Gain (BR-IG)/ Filter Label Powerset with Information Gain (LP-IG)/ Filter
12	Selection	ReliefF based on the Binary Relevance transformation (RF-BR)/ Filter ReliefF based on the Label Powerset transformation (RF-LP)/ Filter Information Gain based on the Binary Relevance transformation (IG-BR)/ Filter Information Gain based on the Label Powerset transformation (IG-LP)/ Filter

Table 2: Dimensionality reduction methods and techniques/ approaches

The prevalence of the filter approach among the attribute selection techniques can be justified by the fact that it is a type of technique independent of the learning algorithm; that is, the selection takes place before the induction step of the classifier. In this case, the selection of the best attributes considers only the characteristics of the data itself, being, therefore, computationally efficient. In addition, there are relatively few multi-label hierarchical classifiers available in the literature to be able to use the wrapper and embedded approaches.

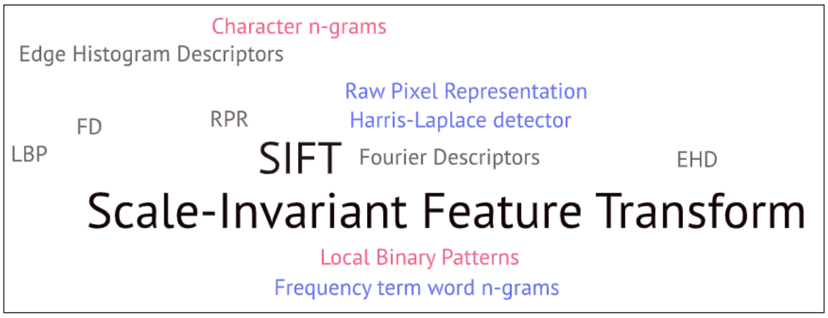

Figure 5: Word cloud of feature extraction techniques identified in the selected works

Figure 6: Word cloud of the attribute selection approaches and techniques identified in the selected works

5.2 Q_2: What were the areas where dimensionality reduction was applied? What is the format of the hierarchical structure of the classes (tree or DAG)?

The studies are distributed in three different areas: bioinformatics, text classification and image processing. The area of image processing was addressed in 7 studies, being the most frequent among the analyzed studies, followed by bioinformatics, approached in 6 studies, and the text classification in 3 works. Table 3 summarizes this information, where the hierarchical structure used to organize the data sets is also presented.

Paper ID	Area	Hierarchy
1	Image Processing	Tree
2	Image Processing	Tree
3	Image Processing	Tree

4	Bioinformatics Image Processing	DAG Tree
5	Image Processing	Tree
6	Bioinformatics	DAG
7	Bioinformatics Image Processing	DAG Tree
8	Bioinformatics Text Classification	(DAG) Tree Tree
9	Text Classification	Tree
10	Bioninformatics Image Processing	DAG Tree
11	Text Classification	Tree
12	Bioinformatics	Tree

Table 4: The application area and type of hierarchy used

The most discussed hierarchy is the tree type, used in 11 of the 12 works. The DAG-like structure is considered in only five studies, all in bioinformatics. Regarding study 8, although the data set is hierarchically organized in the form of a DAG, there was simplification to a tree-like structure, a requirement of the classification tool used in the study.

5.3 Q_3: What was the scientific contribution of the authors in the works to the problem of dimensionality reduction?

This question aims to identify whether the analyzed studies proposed the improvement of existing methods, the creation of a new method for the task of dimensionality reduction or the performance of experimental studies with existing methods. Table 5 summarizes the contributions of each of the eleven analyzed works.

Contribution type	**Papers ID**
New method for dimensionality reduction	4, 5, 6, 10
Adaptation of existing methods	11, 12
Application of existing methods	1, 2, 3, 9
Experimental studies with existing methods	7, 8

Table 5: Type of contribution in the works

In work 4 [Slavkov, 14], a new attribute selection method for hierarchical multi-label classification (HMC-*ReliefF*) was presented, consisting of an adaptation of the *ReliefF* algorithm for the multi-label hierarchical context. This new method can identify the most expressive features in the dataset, in addition to having the ability to deal with the class hierarchy without having to decompose the problem into several flat classification problems.

Paper 5 [Yan, 17] proposes a new approach to HMC in textual data. Such an approach is composed of 3 steps, highlighting the step of representing the resources. For this step, a new attribute selection technique is suggested, which seeks to select the most discriminating attributes in relation to each label. The proposed techniques are the improvement of three existing ones: information gain (IG), GSS Coefficient and NGL Coefficient.

Paper 6 [Cerri, 18] used the wrapper approach and proposed the Clus-HMC to select attributes in the HMC. In order to validate the experiments, two nonlinear classifiers based on neural networks (HMC-LMLP) and genetic algorithms (HMC-GA) were used.

In paper 10 [Huang, 20], the authors present a framework based on semantic and structural information labels: Feature Selection based on Semantic and Structural information of labels (FSSS). The procedure consists of calculating similarity between labels as semantic regularization and extracting parent-child and sibling relationships as structural regularization. This information is passed to a learning model created for feature selection.

Work 11 [Aljedani, 21] proposes a model for HMC of texts written in Arabic. The proposed model incorporates an attribute selection method to reduce the total amount of attributes resulting from the data preparation and pre-processing steps. The methods evaluated correspond to combinations of Binary Relevance and Label Powerset techniques with Chi-Square, Gain Ratio, RelieF and Information Gain techniques. In addition to the proposed classification model, the evaluation of the impact of the attribute selection methods and the dimensions of the attribute space on the proposed model corresponds to a contribution of the work.

Paper 12 [Silva, 21] presents four strategies to apply attribute selection methods in the HMC. The strategies combine Binary Relevance and Label Powerset techniques with ReliefF and Information Gain techniques; these assess the importance of attributes and correspond to multi-label transformation techniques. The four strategies are applied at each level of the class hierarchy and are considered a non-hierarchical multi-label problem. Thus, the attributes selected at each level are combined to compose a new hierarchical set of data. The main contribution of this work is evaluating the ability to select attributes of each proposed strategy.

Work 1 [Dimitrovski, 10] provides an approach that creates a classifier for detecting visual concepts and labeling images. It defines two steps: feature extraction and classification. The feature extraction step uses some existing extraction techniques to obtain global and local descriptions of the images and, thereby, simultaneously predict all labels in the data sample.

Work 2 [Dimitrovski, 11] presents an HMC system for annotating medical images. In the proposed system, the data pre-processing step uses several approaches and combinations of attribute extraction techniques in x-ray images. In addition to the comparative study of the predictive performances of Predictive Clustering Tree (PCT) ensembles with SVM, another contribution of the paperis is the analysis of which combination of attribute extraction techniques produces better predictive results.

In paper 3 [Dimitrovski, 12], the authors propose a new global multi-label hierarchical classifier. This classifier consists of a processing step, where the extraction of attributes is performed, and a classification step. Two approaches and combinations of feature extraction techniques are used to represent diatom images. In addition, a study is carried out to verify if the combination of these techniques increases the predictive performance.

Work 9 [Prabowo, 19] addresses the Indonesian language text HMC to identify targets, groups, and levels of hate speech on Twitter. The proposed approach makes use of classification algorithms such as Random Forest Decision Tree (RFDT), Naive-Bayes (NB), and Support Vector Machine (SVM). In addition, it applies attribute extraction techniques to improve the classification task. The extraction of attributes takes place after the acquisition and pre-processing steps of the textual data and is based on the frequency of the terms word n-gram and character n-gram. Among the contributions of the paper is the analysis of the impact of the extraction of attributes on the proposed approach.

In paper 7 [Slavkov, 18], an extension of the study conducted by [Slavkov, 14] is carried out. The results of experimental studies of the HMC-ReliefF algorithm are presented without any extension or modification in the method.

Work 8 [Melo, 19] performs a systematic comparison between the local and global approaches to attribute selection for flat and hierarchical multi-label classification based on the binary relevance approach.

Figure 7 summarizes the number of published works according to the type of contribution. From the studies evaluated, four present as their main contribution a new method for reducing dimensionality in the context of hierarchical multi-label classification, which corresponds to 31% of the selected works. All these jobs are related to attribute selection. Three works (23%) propose adaptations of existing methods, one for extraction and another for attribute selection. Two papers (15%), both on the selection of attributes, approach the accomplishment of experimental studies on existing methods without presenting any proposal of alteration or adaptation of these methods. In addition, four studies use existing attribute extraction techniques as part of their work proposals, representing 31% of the total selected works.

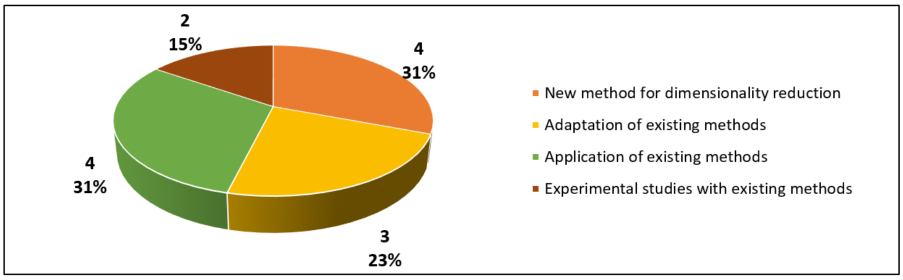

Figure 7: Quantity and Percentage of Works by Type of Contribution

5.4 Q_4: Was the result obtained by the reduced dataset relevant when compared to the dataset formed by all attributes?

This question aims to identify whether the analyzed works presented relevant results for the classification task using dimensionality reduction methods. Table 6 summarizes the answers to research question Q_4.

Was the result obtained by the reduced dataset relevant when compared to the dataset formed by all attributes?	Papers ID
Yes	2, 3, 4, 5, 6, 9, 10, 11
Partially	7, 8, 12
No	1

Table 6: Comparison of the results obtained in the works

In work 2 [Dimitrovski, 11], the experimental results, in terms of efficiency and error, are superior to the SVM approach. In addition, it is inferred, through the results obtained, that the approach provides good performance, is generalizable, and can be applied in different domains. In particular, the results certify that the inclusion of more than one type of feature in the classification process contributes to a better representation of the hierarchical nature of the images and favors improving the predictive performance.

The results presented in work 3 [Dimitrovski, 12] show that the developed method has the superior predictive performance to the approaches compared for the classification of images from the diatom image bank. Regarding the attribute extraction techniques evaluated, the results suggest that the combination of two contour-based and texture-based attributes proved more suitable for the automatic classification process.

The tests in work 4 [Slavkov, 14] were conducted on two datasets in two important domains for hierarchical multi-label classification: functional genomics and image annotation. The results presented indicate a better result in classifying images, although they demonstrate that the HCM-RelieF correctly identifies elements from both domains.

In paper 5 [Yan, 17], the experimental results showed that the proposed approach for selecting attributes successfully performed the proposed task, reducing the space of attributes, preserving the most informative ones, and filtering the noise, in addition to decreasing the dispersion in the data set. The results showed a good rate of reduction in the number of attributes, improving prediction effectiveness and performance.

The results in work 6 [Cerri, 18] indicate that the selection of attributes performed with the Clus-HMC algorithm is more suitable for the hierarchical multi-label classification than the application of multi-label methods already known in the literature, but which do not consider the hierarchical nature of the class structure. Tests performed with the HMC-LMLP approach, based on neural networks, showed better results when the algorithm worked with the original set of attributes. In HMC-GA, the genetic approach, the results are improved when the selection of attributes is performed with the Clus-HMC.

Work 9 [Prabowo, 19] conducted experiments in five different scenarios, varying the number of labels between scenarios. For each scenario, attribute extractions were performed using the word n-gram and character n-gram techniques, with the best result achieved with the uni-gram word extraction feature. The results showed that the proposed approach is able to improve classification performance.

In paper 10 [Huang, 20], the authors demonstrated, through experimental tests, that the techniques used in the study proved to be more effective than other attributes selection methods used in the scope of hierarchical classification.

In the experiments carried out in work 11 [Aljedani, 21], each method, that is, each combination of techniques, was used to select 2 thousand attributes of a space of 11 thousand, resulting from the previous steps of the model. The results acquired showed that the BR -χ^2 combination obtained better results. In addition, the influence of dimensionality size on classification was evaluated, noting that the selection of 4 thousand features offers better performance. The proposed model proved to be significantly better than the other models evaluated in a wide range of evaluation metrics.

In paper 7 [Slavkov, 18], the results indicate that the algorithm performs the resource classification process with good stability, and this result improves with the increase in the number of instances. However, the tests showed that the algorithm is not very sensitive to the size of the neighborhood, being stable with 25 neighbors. The assessment also considers that the quality of a feature ranking algorithm generates a list of features with the good features at the top. This corresponds to a test method called forward feature addition (FFA). The results of this test show that, for most data sets, the classifications of the HCM-ReliefF algorithm were better than the method with which it was compared (binary relevance).

The comparative results in paper 8 [Melo, 19] demonstrate that the local attribute selection approach is better than the global attribute selection in terms of the classification accuracy measure without compromising performance and execution time.

The results of the experiments carried out in paper 12 [da Silva, 21] showed that three of the four proposed strategies (IG-BR, RF-BR, and RF-LP) were able to select relevant subsets of attributes so that with the reduction of the attribute space, the capacity classifier prediction was maintained or improved.

In paper 1 [Dimitrovski, 10], the results show that the proposed approach presents good performance in relation to some methods with which it was compared; but in some cases, the results were not satisfactory. The authors suggest that the extraction step should be revised since the 10% reduction in the number of original attributes may have caused a significant loss of information.

Figure 8 shows the quantity and percentage of works for each of the answers to the question Q_4 (yes, partially, and no).

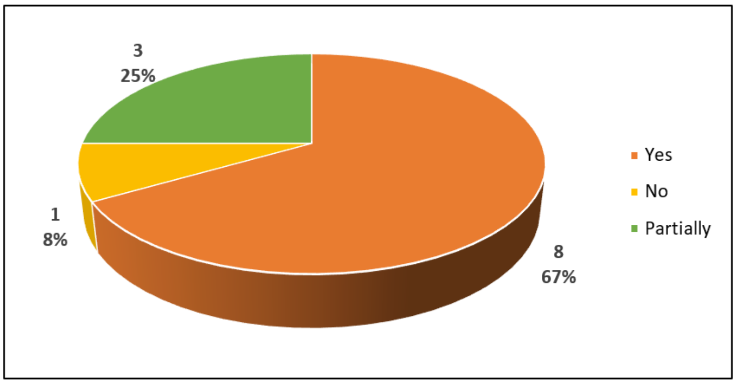

Figure 8: Number and percentage of publications per answer to the question Q_4

In 67% of the analyzed works, that is, eight works, the experimental results were relevant for the reduced dataset, while in 25%, that is, three works, the results were satisfactory only in part of the experiments. Only one study, which corresponds to 8% of the total number of analyzed papers, presented irrelevant results for the reduced dataset.

The papers analyzed in the context of this work showed in their experimental results that the feature selection techniques used ware able to perform the dimensionality reduction without compromising the performance of the classification task. Out of the eight works that approach the selection of attributes, five presented significant results for the reduced dataset in the total set of experiments; and three studies presented good results for the reduced dataset in part of the experiments.

As for the works that approach the extraction of attributes, one did not present satisfactory results with the reduced dataset; the other works presented good results. The selected studies that use the feature extraction method focus on techniques for image representation (papers 1, 2, and 3) and text representation. In general, these works do not have as objective to propose new techniques or adapt existing ones for dimensionality reduction. However, as the focus of these studies is the hierarchical multi-label classification, it was decided to analyze such works from the perspective of the application of the feature extraction techniques they use. Therefore, there are no relevant publications on the central theme of this research that adopt the extraction of attributes as a method.

5.5 Discussion

The Figure 9 summarizes the execution of this work. Starting from an information need and through the detailed systematic mapping carried out in section 4, this work presents the state of the art of dimensionality reduction problem in HMC and indicates trends and research issues.

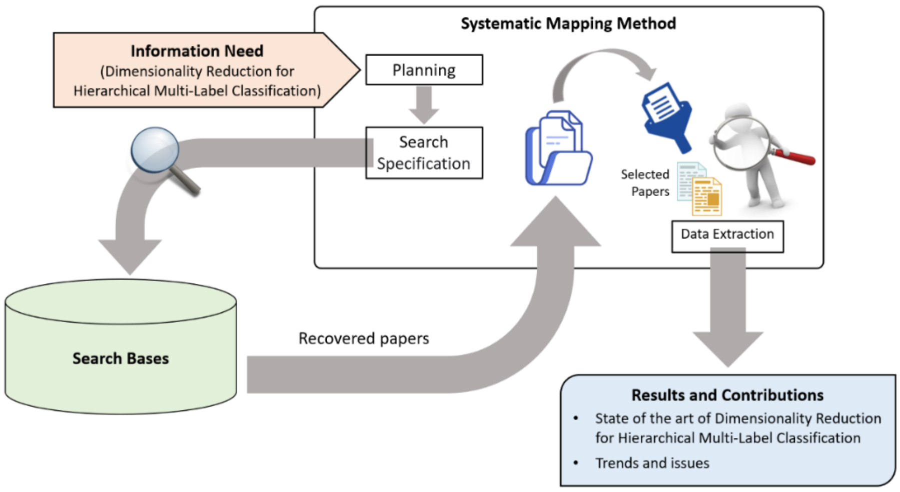

Figure 9: Execution of Systematic Mapping, Results and Contributions

It is known that in traditional classification, redundant and irrelevant attributes can hinder the classifier prediction process. In HMC, this problem can be aggravated due to the complexity of the classification process. Thus, dimensionality reduction should also be adopted as a pre-processing step in HMC.

Dimensionality reduction methods for HMC problems have been proposed. Feature selection has been more frequently used in the studies. Most of the identified methods consist of adaptations of traditional dimensionality reduction techniques, such as Relief, Information Gain, Binary Relevance. In addition, most of them were developed for a hierarchical tree structure. One of the reasons for this is the greater complexity of the DAG hierarchy, as a class can have more than one ancestor, and classes that are at deeper levels tend to have fewer samples.

Based on this mapping, it can be observed that dimensionality reduction for HMC is still a relatively unexplored area, presenting several research possibilities. Regarding the type of hierarchical structure, there is a lack of solutions mainly for DAG structure. In addition, it is relevant to investigate the possibility of developing and/or adapting other dimensionality reduction techniques for HMC problems.

6 Conclusions

This paper presented a systematic mapping to identify which methods and techniques have been used to reduce the dimensionality of databases in the HMC.

The analysis of the selected studies revealed that the works propose new methods or adaptations of existing methods for selecting features, with the adoption of the filter approach prevailing. Regarding the extraction of attributes, it was found that the works only apply such an approach as part of a larger process. There were no studies whose main objective was to develop, adapt or evaluate methods for extracting features regarding hierarchical multi-label classification.

In the time interval considered in this work, it is noted that the number of publications on the subject is still very few, and there is room to investigate the possibility of using other techniques of selection and extraction of attributes in the context of dimensionality reduction for HMC. It is possible to investigate the suitability of traditional techniques, such as Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA), or even evaluate the use of Deep Learning to extract features in HMC scenarios. Investigate the possibility of extending traditional metrics, such as the Fisher Score, or the wrapper or embedded approaches for feature selection in the context of the HMC.

Furthermore, as most of the analyzed works use the hierarchical tree structure, it is possible to develop studies to analyze how the proposed methods can be extended to be applied in HMC problems structured as DAG.

References

[Aljedani, 21] Aljedani N, Alotaibi R, Taileb M. Hmatc: Hierarchical multi-label arabic text classification model using machine learning. Egyptian Informatics Journal vol. 22 num. 3 pp. 225-37, 2021. https://doi.org/10.1016/j.eij.2020.08.004

[Blockeel, 02] Blockeel, H. et al. Hierarchical multi-classification. In: Proceedings of the ACM SIGKDD 2002 Workshop on Multi-relational Data Mining (MRDM 2002). Edmonton, 2002. https://lirias.kuleuven.be/1653518?limo=0

[Borges, 12] Borges, H. B.; Nievola, J. C. Multi-Label Hierarchical Classification using a Competitive Neural Network for Protein Function Prediction. In: 2012 International Joint Conference on Neural Networks (IJCNN 2012), 2012, Brisbane, Austrália. 2012 International Joint Conference on Neural Networks (IJCNN 2012). Piscataway, NJ: IEEE Press, v. 1. p. 1-8, 2012. https://ieeexplore.ieee.org/document/6252736

[Cerri, 12] Cerri, R.; Barros, R. C.; de Carvalho, A. C. P. L. F. A genetic algorithm for Hierarchical Multi-Label Classification. In Proceedings of the 27th Annual ACM Symposium on Applied Computing (SAC '12). Association for Computing Machinery, New York, NY, USA, 250–255, 2012. https://doi.org/10.1145/2245276.2245325

[Cerri, 18] Cerri, R., Mantovani, R. G., Basgalupp, M. P., Carvalho, A. C. P. L. F. Multi-label Feature Selection Techniques for Hierarchical Multi-label Protein Function Prediction. In: IEEE. 2018 International Joint Conference on Neural Networks (IJCNN). pp. 1-7, 2018. https://doi.org/10.1109/IJCNN.2018.8489247

[Silva, 21] Silva, L.; Cerri, R. Feature Selection for Hierarchical Multi-label Classification. In: Abreu, P.H., Rodrigues, P.P., Fernández, A., Gama, J. (eds) Advances in Intelligent Data Analysis XIX. IDA 2021. Lecture Notes in Computer Science, vol 12695, pp. 196-208. Springer, Cham, 2021. DOI: 10.1007/978-3-030-74251-5_16

[Dimitrovski, 10] Dimitrovski, I. et al. Detection of visual concepts and annotation of images using ensembles of trees for hierarchical multi-label classification. In: SPRINGER. International Conference on Pattern Recognition. pp. 152-161, 2010. DOI: 10.1007/978-3-642-17711-8_16

[Dimitrovski, 11] Dimitrovski, I. et al. Hierarchical annotation of medical images. Pattern Recognition, Elsevier, v. 44, n. 10-11, pp. 2436-2449, 2011. https://doi.org/10.1016/j.patcog.2011.03.026

[Dimitrovski, 12] Dimitrovski, I. et al. Hierarchical classification of diatom images using ensembles of predictive clustering trees. Ecological Informatics, Elsevier, v. 7, n. 1, pp. 19-29, 2012. https://doi.org/10.1016/j.ecoinf.2011.09.001

[Faceli, 11] Faceli, K. et al. Inteligência Artificial: uma abordagem de aprendizagem de máquina. Rio de Janeiro: LTC, 2011.

[Freitas, 07] Freitas, A.; Carvalho, A. A tutorial on hierarchical classification with applications in bioinformatics. In: Research and trends in data mining technologies and applications. IGI Global, pp. 175-208, 2007. DOI: 10.4018/978-1-59904-271-8.ch007

[Gargiulo, 19] Gargiulo, F. et al. Deep neural network for hierarchical extreme multi-label text classification. Applied Soft Computing, Elsevier, v. 79, pp. 125-138, 2019. https://doi.org/10.1016/j.asoc.2019.03.041

[Ghodsi, 06] Ghodsi, A. Dimensionality reduction a short tutorial. Department of Statistics and Actuarial Science, Univ. of Waterloo, Ontario, Canada, v. 37, pp. 38, 2006. https://www.math.uwaterloo.ca/~aghodsib/courses/f06stat890/readings/tutorial_stat890.pdf

[Huang, 20] Huang, H.; Liu, H. Feature selection for hierarchical classification via joint semantic and structural information of labels. Knowledge-Based Systems, v. 195, p. 105655, mai. 2020. https://doi.org/10.1016/j.knosys.2020.105655

[Kohonen, 90] Kohonen, T. The self-organizing map. Proceedings of the IEEE, IEEE, v. 78, n. 9, pp. 1464-1480, set. 1990.

[Kruskal, 64] Kruskal, J. B. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. Psychometrika, Springer, v. 29, n. 1, pp. 1-27, 1964

[Kumar, 14] Kumar, V.; Minz, S. Feature selection: a literature review. SmartCR, v. 4, n. 3, pp. 211-229, 2014. https://faculty.cc.gatech.edu/~hic/CS7616/Papers/Kumar-Minz-2014.pdf

[Melo, 19] Melo, A.; Paulheim, H. Local and global feature selection for multilabel classification with binary relevance. Artificial Intelligence Review, Springer, v. 51, n. 1, pp. 33-60, 2019. https://link.springer.com/article/10.1007/s10462-017-9556-4

[Pagani, 15] Pagani, R. N.; Kovaleski, J. L.; Resende, L. M. Methodi Ordinatio: a proposed methodology to select and rank relevant scientific papers encompassing the impact factor, number of citation, and year of publication. Scientometrics, Springer, v. 105, n. 3, pp. 2109-2135, 2015. https://link.springer.com/article/10.1007/s11192-015-1744-x

[Prabowo, 19] Prabowo, F. A.; Ibrohim, M. O.; Budi, I. Hierarchical Multi-label Classification to Identify Hate Speech and Abusive Language on Indonesian Twitter, in 6th International Conference on Information Technology, Computer and Electrical Engineering (ICITACEE), pp. 1-5, 2019. https://doi.org/10.1109/ICITACEE.2019.8904425

[Rattan, 13] Rattan, D; Bhatia, R; Singh, M. Software clone detection: a systematic review. Information and Software Technology, v. 55, n. 7, p. 1165-1199, July 2013. https://doi.org/10.1016/j.infsof.2013.01.008

[Slavkov, 14] Slavkov, I. et al. ReliefF for hierarchical multi-label classification. Lecture Notes in Computer Science, LNAI, v. 8399, pp. 148-161, 2014. DOI: 10.1007/978-3-319-08407-7_10

[Slavkov, 18] Slavkov, I. et al. HMC-ReliefF: Feature Ranking for Hierarchical Multi-label Classification. Computer Science and Information Systems, v. 15, n. 1, p. 187-209, 2018. DOI: 10.2298/CSIS170115043S

[Silla Jr., 11] Silla Jr., C. N.; Freitas,A. A. A survey of hierarchical classification across different application domains. Data Min. Knowl. Discov., Kluwer Academic Publishers, Hingham, MA, USA, v.22, n.1-2, pp.31–72, 2011. DOI: 10.1007/s10618-010-0175-9

[Stojanova, 13] Stojanova, D.; Ceci, M.; Malerba, D.; Dzeroski, S. Using PPI network autocorrelation in hierarchical multi-label classification trees for gene function prediction. BMC Bioninformatics, v. 14, 285, 2013. https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-14-285

[Van Der Maaten, 09] Van Der Maaten, L.; Postma, E.; Van Den Herik, J. Dimensionality reduction: a comparative. Journal of Machine Learning Research, v. 10, n. 66-71, p. 13, 2009. http://tsam-fich.wdfiles.com/local--files/apuntes/TPAMI_Paper.pdf

[Vens, 08] Vens, C. et al. Decision trees for hierarchical multi-label classification. Machine learning, Springer, v. 73, n. 2, pp. 185, 2008. https://link.springer.com/article/10.1007/s10994-008-5077-3

[Yan, 17] Yan, S.; Wong, K.-C. Elucidating high-dimensional cancer hallmark annotation via enriched ontology. Journal of Biomedical Informatics, Elsevier, v.73, pp. 84-94, 2017. DOI: 10.1016/j.jbi.2017.07.011