Systematic Literature Review on the symbolic knowledge extraction techniques

Giovanni Mormone[1,*], Fabián Aspée[1,*]

[1] University of Bologna, Cesena, Italy.

[*] Corresponding author: giovanni.mormone@studio.unibo.it, fabian.aspeeencina@studio.unibo.it

Abstract: The development in the field of the AI, with the good performance of the so called sub-symbolic algorithms, has given the rise to the problem of the opacity of this methods, that take decisions based on their internal structure in an hard to understand way. To address this problem researchers proposed various symbolic knowledge extraction techniques, to overcome the opacity problem and extract human-readable rules. In this work we are going to perform a Systematic Literature Review (SLR) on the field of symbolic knowledge extraction to summarize and explain approaches, techniques, evaluation metrics and motivations behind knowledge extraction. To answer the research questions, we grouped together all the ambiguous terms found in the literature, to make the answers richer and to clarify the study field.

Keywords: knowledge extraction, Systematic Literature Review, sub-symbolic, symbolic.

1. Introduction

In the artificial intelligence area, there is a problem that transcends various sub-fields, a problem that belongs to the category of algorithms known as *black box or sub-symbolic* (e.g. Support Vector Machine (SVM) [1], Artificial Neural Network (ANN) [2] etc.), known for their opacity. This opacity derives from the internal organization of the knowledge achieved during the training, opacity that makes the decision process of said technologies hard to understand by humans. The definition sub-symbolic is used to make an opposition between them and the algorithms that are known as *symbolic* (e.g. decision tree [3], etc.), algorithms whose internal state is easy to understand by humans thanks to the organization of the knowledge employed, knowledge that is exploited to create decision rules in a human-readable form.

However, with the development and the advancement in the artificial intelligence field, blackbox technologies are starting to move towards a wider utilization in multiple areas, moving away from research interest. The spread of the artificial intelligence in various areas contributed to increase the importance of the well known problem of the opacity, embedded in these technologies [4]. In spite of this problem, the good performances of algorithms like SVM or ANN makes them the best choice in a broad range of problems [5] [6], but in some applications named safety critical, such as airlines and power stations, is essential that an explanation capability should become an integral part of the functionality of a trained ANN [7] or trained SVM.

To address the opacity problem present in sub-symbolic algorithms, the researchers tried various approaches since the 80s [8], approaches that were later categorized in different categories, e.g decompositional [9], pedagogical [9] or eclectic [7], a key role is played by symbolic knowledge extraction techniques, with the aim to extract *symbolic knowledge* from sub-symbolic algorithms.

The aim of this work is to analyze and discuss the various approaches and techniques used in the literature to extract symbolic knowledge from sub-symbolic algorithms, in particular the most recent developments in the area. The analysis is done by the means of a Systematic Literature Review (SLR) regarding the topic proposed, in particular the work is articulated as follows: in section 2 we define the method that will be followed in the SLR, in section 3 we present the results of the research using the method defined in 2, in section 4 we discuss and analyze the results presented in section 3 to answer the research questions defined in 2, in section 5 we present the quality assessment of the works as defined in 3 and in section 6 the conclusions are presented.

2. Working Method

The first requirement to conduct an SLR is the definition and the clarification of the objective of the review. Making an SLR also requires the adoption of well defined and accepted guidelines regarding the area of the work. The objective of this SLR can therefore be defined as the discussion and analysis of the most recent techniques of knowledge extraction from sub-symbolic models to symbolic ones , where knowledge extraction is a broad definition that can be briefly summarized as the process that aims to extract symbolic rules from sub-symbolic approaches to make them understandable and transparent to human users [10].

To conduct this SLR we started defining a working method. The definition of a rigorous method is fundamental to reduce the bias of the writer towards the research topic and to make the work reliable and repeatable by the reader.

We based our method on the guidelines described in [11] [12], with the former being the base for SLR in software engineering and the latter being an extension of the classic SLR method that introduces the snowballing search-approach. The phases of our method can be synthesized in the following steps:

- Definition of the Research Questions.
- Definition of the Search Strategy.
- Definition of the Selection Criteria.
- Data Extraction and Refining.
- Clustering and Discussion.
- Quality Assessment.

2.1. Research Questions

The first and most important part of an SLR is the definition of the research questions, questions that will be the core of the review. As stated before, the main objective of this review is to analyze knowledge extraction from black box models in machine learning in the last years, as mentioned in section 1. In particular, the general direction of the SLR is aimed to answer to this broad questions:

- G1: Why we do knowledge extraction?
- G2: How we do knowledge extraction?
- G3: How we evaluate knowledge extracted?

To make the questions clearer and to remove their ambiguity, we refined them in this specific questions:

- Q1: What are the motivations that make knowledge extraction important? (G1)
- Q2: What knowledge extraction approaches are currently used? (G2)
- Q3: What knowledge extraction techniques are the more used currently? (G2)
- Q4: What are the quality metrics of the extracted knowledge? (G3)

This questions will influence how we are going to approach to the search process and to the analysis of the literature, being the main drive of the review and the first significant criteria taken in consideration when examining the research articles.

2.2. Search Strategy

To make the review reliable and repeatable, the definition of a search strategy is very important. The main search strategy used in this work is a research in digital archives via the utilization of keyword-based queries, integrated with the analysis of the bibliography cited and the works that cite the selected articles found by the initial research.

Initially our strategy considered the use of several search engines (*dbpl*, *ScienceDirect*, *SpringerLink*, *Google Scholar*) but, since in our opinion every search engine has a different definition of relevance, relevance that is not stated clearly (e.g. *SpringerLink*[1] *Google Scholar*[2]) or isn't stated at all (e.g. *dbpl*) reducing the reliability of the search process, we decided to use *ScienceDirect*.

The choice of *ScienceDirect* is motivated by the clear and simple definition of relevance[13], along with the option of defining powerful filters in the research, qualities that enhance the reliability and repeatability of the search process considering that the results of a research is less likely to change as time passes by.

Given the scope of this review and the research questions that are going to be addressed, we defined and used the following search queries[3]:

- Query1: ((rule extraction) OR (knowledge extraction)) AND (from (support vector machine) OR (SVM))
- Query2: ((Pedagogical Decompositional Eclectic) OR (symbolic knowledge extraction)) OR ((Pedagogical Decompositional Eclectic) AND (symbolic knowledge extraction))

[1] Relevance definition SpringerLink

[2] Relevance definition Google Scholar

[3] The keywords included in the research aren't case sensitive, the same condition is not equivalent for boolean operators. the brackets aren't mandatory but recommendable.

- Query3: ((rule extraction) OR (knowledge extraction)) AND (from (artificial neural networks) OR (ANN))

To get results that are more significant, the above queries were used in the search field labeled as *Title, abstract or author-specified keywords* using the advanced search option. To select only recent works, we decided to filter the results considering the *year of publication*, selecting only the works published between 2017 and 2021. The last filters used in the search engine were the selection of only *research articles*.

To further narrow the search, we also defined additional inclusion/exclusion criteria, described in the next section, with the objective to get a selection of relevant works. The last step in the study search phase was the application of a snowballing technique, where we analyzed the citations of the selected works and the works that cited them, to select works that could be used in the review.

2.3. Selection Criteria

To restrict the number of papers selected and only analyze relevant ones, we defined some additional selection criteria, considering that the time constraint and subject area filter defined in the above section could be described as selection criteria.

The inclusion criteria[4] that we added are the following:

- Language: we only included works in English.
- Citations: we included works that have cited 25 other works.
- Publication Journal: we included works publish in a journal with at least Q2 quartiles rank in the field of the review[5].

The exclusion criteria[6] that we added are the following:

- Pertinence of the keywords: we excluded works that in the keywords do not mention knowledge extraction, rule extraction or some equivalent term.
- Pertinence of the abstract: we excluded works that in the abstract do not state as objective the aim of our work[7]
- Work availability: we excluded works not openly accessible or accessible trough university.
- Short paper: we excluded works that are short versions of other papers.
- Repeated Papers: we excluded works that have already been selected in other queries.

2.4. Data extraction and refining

After the search phase we'll have a set of papers potentially useful for the review. The main objective of this phase is to deeply analyze the search set, to extract all the information that can answer one or more of the research questions defined above. The second goal of this phase is to remove the papers that, after a deep analysis, do not have information regarding the research questions, acting as a last filter phase.

[4]For a work to be selected it has to meet all the inclusion criteria.
[5]The considered quartiles are going to be searched on Scimago
[6]For a work to be excluded is sufficient to not meet one of the exclusion criteria
[7]e.g. papers not related to knowledge extraction techniques etc.

2.5. Clustering and Discussion

After the extraction of the useful data done in the above phase, we will group together the information extracted, making clusters of information regarding each research question. This is done to help us organize the concepts to better answer the questions. After the clustering of the information, we are going to answer each of the research questions.

2.6. Quality Assessment

After the discussion phase of the papers to be analyzed in the review, we are going to make a quality assessment of the works.

The quality assessment aims to define the quality of the works using defined criteria, criteria used to assign a numerical score to the works selected. The score is meant to give a quick overview of the importance of a given study in relation with the others, as well as to classify the papers considering their weight.

To asses the quality of the paper analyzed, in terms of a numerical score, we used the following criteria:

- Number of citations: we considered the numbers of works that reference each paper, given that as a rule of thumb we can say that a good and impactful work on the research area is going to be more referenced[8].
- Usability of the work: we considered the release state of the work (i.e. pseudo-code, source code and library availability).
- Type of work: we considered the type of work proposed in the paper (i.e. new proposal, extension of an algorithm or practical application of an algorithm).
- Impact factor of the publishing journal: we considered the impact factor of the publishing journal [14], that reflects the yearly average number of citations of articles published in the last two years in a given journal[9].
- Number of questions answered: we considered the number of questions that each work has given us information of, to asses how much a work is relevant for this survey.

[8] A self reference will be counted as half reference

[9] The considered impact factor is going to be searched on Scimago

3. Results of the Research

In this section we are going to present the results of the papers research, research done using the method described in the above section 2.

3.1. Search process

The search process can be divided in 3 stages:

- First stage: application of the search queries and filters 2.2.
- Second stage: application of the selection criteria 2.3.
- Third stage: snowballing on the papers that are the result of the second stage. The papers selected during the snowballing must comply with the filters defined for the first two stages, with the exception of the time limit, limit that will be raised to include work potentially useful related to the selected ones.

3.1.1. First stage: The first stage of the search process was the application of the queries defined in section 2.2, applying the basic filters provide by the search engine used. The results of the first stage are the following:

Query Id	Date	Total Article
Query1	June 12, 2021	93
Query2	June 12, 2021	10
Query3	June 12, 2021	59

Table 1 Results of the search process

3.1.2. Second stage: In the second stage we analyzed each paper found in the first stage and applied the selection criteria defined in 2.3. The result for each query, after the application of the selection criteria are the following 2:

Query Id	Starting Papers	Excluded Papers	Total Article
Query1	93	89	4
Query2	10	5	5
Query3	59	55	4

Table 2 Results of the search process second stage

The set of selected papers will be utilized as initial set of the snowballing phase and will be presented in 3.1.3.

3.1.3. Third stage: During the snowballing stage, we analyzed the references related to the works selected in the second stage. We searched the papers cited by our selected works as well as the papers that cited them, paper that had to comply with the searching criteria applied during 3.1.1 and 3.1.2 To select potentially useful works related to the selected ones, we raised the time limit for the referenced works, shifting the starting year to 2010. The snowballing process is iterative, meaning that it starts from the start set than, for each work it analyzes cited and citing works to search additional related papers, doing it in an iterative way until no additional paper could be found[10].

Start Set: The starting set, selected as a result of the first two stages and used to apply the snowballing technique, is the following:

- C1: Singh, N., Singh, P., & Bhagat, D. (2019). A rule extraction approach from support vector machines for diagnosing hypertension among diabetics. Expert Systems with Applications, 130, 188-205.
- C2: Baker, H., Hallowell, M. R., & Tixier, A. J. P. (2020). Automatically learning construction injury precursors from text. Automation in Construction, 118, 103145.
- C3: Jezewski, M., Czabanski, R., Leski, J. M., & Jezewski, J. (2019). Fuzzy classifier based on clustering with pairs of ξ-hyperballs and its application to support fetal state assessment. Expert Systems with Applications, 118, 109-126.
- C4: Moro, S., Rita, P., & Coelho, J. (2017). Stripping customers' feedback on hotels through data mining: The case of Las Vegas Strip. Tourism management perspectives, 23, 41-52.
- C5: Lima, R., Espinasse, B., & Freitas, F. (2019). A logic-based relational learning approach to relation extraction: the OntoILPER system. Engineering Applications of Artificial Intelligence, 78, 142-157.
- C6: Yu, J., & Liu, G. (2021). Extracting and inserting knowledge into stacked denoising auto-encoders. Neural Networks, 137, 31-42.
- C7: Yu, J., & Liu, G. (2020). Knowledge-based deep belief network for machining roughness prediction and knowledge discovery. Computers in Industry, 121, 103262.
- C8: Yedjour, D., & Benyettou, A. (2018). Symbolic interpretation of artificial neural networks based on multiobjective genetic algorithms and association rules mining. Applied Soft Computing, 72, 177-188.
- C9: Confalonieri, R., Weyde, T., Besold, T. R., & del Prado Martín, F. M. (2021). Using ontologies to enhance human understandability of global post-hoc explanations of black-box models. Artificial Intelligence, 296, 103471.
- C10: He, C., Ma, M., & Wang, P. (2020). Extract interpretability-accuracy balanced rules from artificial neural networks: A review. Neurocomputing, 387, 346-358.
- C11: Vidyarthi, V. K., & Jain, A. (2020). Knowledge extraction from trained ANN drought classification model. Journal of Hydrology, 585, 124804.

[10] In our work we stopped after the first iteration

- C12: Harder, F., & Besold, T. R. (2018). Learning Łukasiewicz logic. Cognitive Systems Research, 47, 42-67.
- C13: Chan, V. K., & Chan, C. W. (2019). Towards explicit representation of an artificial neural network model: Comparison of two artificial neural network rule extraction approaches. Petroleum.

Snowballing Iteration: In the snowballing iteration we considered the start set, looking for potential papers useful for our review. Analyzing every paper we made two search steps:

- Backward step: We analyzed all the papers cited by the start set found in the second stage of the search process.
- Forward step: We analyzed all the papers that cited one paper of the start set found in the second stage of the search process.

After the snowball iteration, we found 24 more papers related to the topic studied in our review. The final count of works that we used is the following:

Number of Papers in the start set	Number of additional papers found in the snowballing set	Total number of Papers
13	24	37

Table 3 Results of the search process in the snowballing

- C14: Pourpanah, F., Lim, C. P., & Saleh, J. M. (2016). A hybrid model of fuzzy ARTMAP and genetic algorithm for data classification and rule extraction. Expert Systems with Applications, 49, 74-85.
- C15: Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2018). A survey of methods for explaining black box models. ACM computing surveys (CSUR), 51(5), 1-42.
- C16: de Fortuny, E. J., & Martens, D. (2015). Active learning-based pedagogical rule extraction. IEEE transactions on neural networks and learning systems, 26(11), 2664-2677.
- C17: Wang, S., Wang, Y., Wang, D., Yin, Y., Wang, Y., & Jin, Y. (2020). An improved random forest-based rule extraction method for breast cancer diagnosis. Applied Soft Computing, 86, 105941.
- C18: Tran, S. N., & Garcez, A. S. D. A. (2016). Deep logic networks: Inserting and extracting knowledge from deep belief networks. IEEE transactions on neural networks and learning systems, 29(2), 246-258.
- C19: Panigutti, C., Perotti, A., & Pedreschi, D. (2020, January). Doctor XAI: an ontology-based approach to black-box sequential data classification explanations. In Proceedings of the 2020 conference on fairness, accountability, and transparency (pp. 629-639).
- C20: Shinde, S., & Kulkarni, U. (2016). Extracting classification rules from modified fuzzy min–max neural network for data with mixed attributes. Applied Soft Computing, 40, 364-378.

- C21: Yedjour, D. (2020). Extracting Classification Rules from Artificial Neural Network Trained with Discretized Inputs. Neural Processing Letters, 52(3), 2469-2491.
- C22: Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-cam: Visual explanations from deep networks via gradient-based localization. In Proceedings of the IEEE international conference on computer vision (pp. 618-626).
- C23: Zhang, Q., Wu, Y. N., & Zhu, S. C. (2018). Interpretable convolutional neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 8827-8836).
- C24: Zhang, Q., Yang, Y., Ma, H., & Wu, Y. N. (2019). Interpreting cnns via decision trees. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 6261-6270).
- C25: Chan, V., & Chan, C. (2017). Learning from a carbon dioxide capture system dataset: Application of the piecewise neural network algorithm. Petroleum, 3(1), 56-67.
- C26: Luo, G. (2016). Automatically explaining machine learning prediction results: a demonstration on type 2 diabetes risk prediction. Health information science and systems, 4(1), 1-9.
- C27: Hayashi, Y., Setiono, R., & Azcarraga, A. (2016). Neural network training and rule extraction with augmented discretized input. Neurocomputing, 207, 610-622.
- C28: Zeng, Q., Huang, H., Pei, X., Wong, S. C., & Gao, M. (2016). Rule extraction from an optimized neural network for traffic crash frequency modeling. Accident Analysis & Prevention, 97, 87-95.
- C29: Han, L., Luo, S., Yu, J., Pan, L., & Chen, S. (2014). Rule extraction from support vector machines using ensemble learning approach: an application for diagnosis of diabetes. IEEE journal of biomedical and health informatics, 19(2), 728-734.
- C30: Ribeiro, M. T., Singh, S., & Guestrin, C. (2016, August). " Why should i trust you?" Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining (pp. 1135-1144).
- C31: Özbakır, L., Baykasoğlu, A., & Kulluk, S. (2010). A soft computing-based approach for integrated training and rule extraction from artificial neural networks: DIFACONN-miner. Applied Soft Computing, 10(1), 304-317.
- C32: de Fortuny, E. J., & Martens, D. (2012, December). Active learning based rule extraction for regression. In 2012 IEEE 12th International Conference on Data Mining Workshops (pp. 926-933). IEEE.
- C33: Stoean, R., & Stoean, C. (2013). Modeling medical decision making by support vector machines, explaining by rules of evolutionary algorithms with feature selection. Expert Systems with Applications, 40(7), 2677-2686.
- C34: Augasta, M. G., & Kathirvalavakumar, T. (2012). Reverse engineering the neural networks for rule extraction in classification problems. Neural processing letters, 35(2), 131-150.

- C35: Augasta, M. G., & Kathirvalavakumar, T. (2012, March). Rule extraction from neural networks—A comparative study. In International Conference on Pattern Recognition, Informatics and Medical Engineering (PRIME-2012) (pp. 404-408). IEEE.
- C36: Farquad, M. A. H., Ravi, V., & Raju, S. B. (2010). Support vector regression based hybrid rule extraction methods for forecasting. Expert Systems with Applications, 37(8), 5577-5589.
- C37: Tsopze, N., Mephu-Nguifo, E., & Tindo, G. (2011, July). Towards a generalization of decompositional approach of rule extraction from multilayer artificial neural network. In The 2011 International Joint Conference on Neural Networks (pp. 1562-1569). IEEE.

3.2. Data extraction and refining

In the data extraction phase we have deeply analyzed all the potentially useful works found during the above phases. Our main goal was to extract all the information regarding one or more of the research questions. The second objective of this phase was to exclude all the works that, while meeting the inclusion and exclusion criteria, did not provide useful insights about the review topic.

During this phase we excluded the following works:

- **C2**, **C3**, **C4**, **C5**, **C22**, **C23**, **C26**.

The final amount of works selected for our review can be seen in table 4

Starting papers after snowballing	Number of papers excluded during data analysis	Final number of Papers
37	7	30

Table 4 Final number of works

4. Clustering and Discussion

4.1. Clustering

To cluster the information needed to answer the research questions, we used an iterative approach. We started reading the information extracted in the extraction phase looking for info regarding the first question, grouping the useful information together and deleting it from the extracted data. Then we repeated the process for each question, having less information to analyze in each step. This helped us to maintain focus and to look only to useful info, removing data already used that we did not need anymore.

4.2. Discussion

After the above phase we are now going to answer the research questions. As mentioned in 2.1 the main objective of this review is to answer this broad questions:

- Why we do knowledge extraction?
- How we do knowledge extraction?
- How we evaluate knowledge extracted?

To properly answer this question, we defined four more specific research questions, Q1, Q2, Q3 and Q4. In the following we are going to answer each specific question, using the information retrieved in the earlier phases of the review described in 3.

4.2.1. Q1 - What are the motivations that make knowledge extraction important?: In the literature analyzed we identified three main motivations that make knowledge extraction important. To next paragraphs are the explanation of this motivations.

Safety critical system: One of the main reasons for rule extraction is the need, in safety critical systems, to explain why a decision is made. In the medical field [15–25], although the performance of blackbox classifiers are usually very good, the users are reluctant to accept them because the explanation of decision for classification of the disease is an important requirement.

Another field which requires an explanation to blackbox is credit scoring [16, 17, 21, 24], where this requirement is even a legal one[17, 21, 24].

Additional examples of systems where an explanation is needed are fault detection [16, 18, 19], traffic crash-control [16–18, 21, 26], insurance [17], process control [18], audit mining [21], air traffic control [21, 24], power plants [24], nuclear [25], self-driving cars and robotic assistants [20].

In safety critical systems the justification of the decisions taken by the ANN is even more important than its performance [16].

To make the blackbox model user understandable: Blackbox models have good performance in their application fields and are vastly used. However they make decisions in an opaque way, using internal logic that the users cannot understand. In this regard, to make the systems accepted by the users [15, 17, 18, 21, 23, 24, 27–29] and to expand their utilization [15, 21, 22, 24, 30] it is important to give a way to extract and explain and verify their decision process [15–20, 24–26, 29–38]

Improve performance and robustness of the system: When applying blackbox models we are utilizing algorithms that make decisions autonomously, basing their process on functions and criteria that are opaque, not visible by the users. Although the performance of the classifiers are usually very good, they can make wrong assumptions that lead to errors. To address this problem knowledge extraction techniques aims to extract information to improve performance and robustness of this systems [15, 17, 18, 20, 21, 24]

4.2.2. Q2 - What knowledge extraction approaches are currently used?: In the analyzed literature, we found various definitions regarding the approaches used for the knowledge extraction. The approaches,also defined as translucency [20, 21, 24, 35], found can be summarized in three big categories:

- Decompositional
- Pedagogical
- Eclectic

Decompositional Approach: In the decompostional [15, 18, 20, 21, 24, 25, 27, 32, 35, 37–40] approach, also named transparent [15],local [16, 25, 41], the aim is to utilize the internal structure of the blackbox to directly extract symbolic rules [15, 16, 20, 21, 24, 32, 35, 38]. In particular if the blackbox is an SVM the decompositional approach uses the Support Vectors to extract the knowledge [15, 40] In the case of ANNs the rules are extracted at using the single neurons that compose the network [15, 16, 18, 19, 25, 27, 32, 35, 37, 38, 41, 42] Decompostional approaches are best suited for function approximation type of problems [27] Some of the problems of this approach are that these methods become inefficient when the underlying model become large [16, 41], that cannot deal with ANNs that have more than three hidden layers [18], that the complexity of the extracted rules increases too much [18, 37, 38], are specifically tailored to extract rule from a defined method and are not portable to other types of blackbox models [16, 21]

Pedagogical Approach: In the pedagogical [15, 18–21, 24, 25, 32, 35, 37–40, 42], also named learning-based [15],global [16, 25, 41], black-box [19, 41] the extracted rules are obtained without reference to the internal structure of the underlying model, using it as a blackbox and generating the rules operating directly on the input (or predictor) - output (or predicted) relationship [15, 16, 18–21, 24, 25, 27, 32, 35, 37, 38, 40–42] The pedagogical approach has the advantage that it is not specifically tied to a certain blackbox model and can be applied to all types of algorithms [16, 21, 24, 41] The pedagogical approach can be faster than the decompositional [37, 38] and, if the performance of a pedagogical approach is the same of a decompositional one, the pedagogical approach is to prefer [24]. One drawback of the pedagogical approach is that hey are some what less likely to accurately capture all of the valid rules describing the network's behavior [37, 38].

Eclectic Approach: The eclectic approach [15, 18, 20, 25, 32, 35, 37, 38, 41, 42], also named hybrid [15, 32, 42], is the combination of both the decompositional and pedagogical approach [15, 18, 25, 35, 37–39, 41, 42], e.g. analyzing the internal structure of the blackbox like the decompositional approach and extracting global rules like the pedagogical [15, 42]. The need for the definition of the eclectic approach rises because not all rule extraction techniques map neatly into the pedagogical or decompositional category [42]

4.2.3. Q3 - What knowledge extraction techniques are the more used currently?:

Knowledge extraction is a field vastly explored in the literature. In the recent past many techniques have been proposed for the knowledge extraction, both new proposal or extensions of previous works. In the following paragraphs we are going to group the techniques using the information extracted in the question Q2 and we will shortly explain how they work.

Decompositional: The decompositional approach is the one with the biggest number of techniques proposed.

ANN

RxRen A remarkable technique is rule extraction by reverse engineering from neural network (RxRen). The proposed algorithm consists of two phases. In the first phase, it removes the insignificant input neurons from the trained neural network and finds the mandatory data range of each significant input neuron for classifying the given testing data as in particular class. The second phase constructs the classification rules for each class using the data ranges obtained in phase1 and refines the generated rules by the process of rule pruning and rule update [37].

MLP + MGA is a technique aimed to extract rules from an MLP using Multi-objective genetic algorithm (MGA). After the training of the MLP an MGA searches in the space of the rule set trying to satisfy a maximum of fidelity and coverage and a minimum of complexity. Rules whose support and confidence and lift values are greater than certain thresholds are selected and sorted by support in descending order; The resulting set only comprises the best rules [16].

PWL-ANN piece-wise linear (PWL) algorithm is responsible for approximating the hidden activation functions of a trained neural network with piece-wise linear equations. It then replaces the hidden layer of the network with the approximated equations, with the new network called piece-wise linear neural network model (or PWL-ANN model). It then can extract rules in the form of multiple linear regression equations from the PWL-ANN models [32].

Rule-INF to extract rules from stacked denoising auto-encoders (SDAE) or from restrict Boltzmann machines (RBMs). The method can be briefly divided in a phase where the model is pre-trained based on the data and, after this phase the rules are extracted from the trained network using the rule extraction algorithm, called Rule-INF [30, 43].

Another proposal is the application of a layer-wise approach to extract knowledge from RBMs and DBNs, with the repeated application of a tailored algorithm [33].

Another case is the proposal to extract knowledge from a CNN in the form of a decision tree, with the application of a specific algorithm [23].

A proposal that can extract knowledge from neural networks(NN) with augmented discretized variables(ADV), employs an algorithm that can extract classification rules from pruned networks in the form of propositional rules, MofN[11] rules or rule sets with disjoint rule conditions for discrete and continuous attribute [44].

A proposal that can extract the same type of rules as the above technique, uses a new method able to extract one intermediate structure (called generators list) from an ANN [25].

Trying to integrate rule extraction with Łukasiewicz logic [45] an algorithm that extracts all minimal activating and all maximal non-activating inputs for each output neuron of the network, which can then be used to compute the logical rules generating this activation [31].

RF

[11] M of a set of N conditions

IRFRE is a proposal that aims to extract rules from a Random Forest using multi-objective evolutionary algorithms (MOEA). Improved Random Forest-based rule extraction (IRFRE) can be divided in two main steps:

- As for the Random Forest-based rule generation, Random Forest is first constructed with Classification and Regression Tree (CART) and then IF-THEN rules are then detached from the trained trees by tracking the path from the root node to each leaf node in each tree, which are combined into a rule set [22].
- As for the MOEA-based rule extraction, a multi-objective evolutionary algorithm is employed to identify optimal combination of rules [22].

Eclectic:

SVM

SVM+XGboost aims to develop an hybrid approach to extract rules from SVM using XGBoost, which is a powerful ensemble gradient boosting method for prediction Gradient boosting is a learning technique that constructs a strong learner by integrating various weak learners to provide a more regularized model formulation that controls overfitting [15]. SVM+XGboost uses a two step approach, where in the first step they train an SVM and extract SVs that are then classified by the SVM to generate a modified training set. In the second step the modified dataset is given to the XGBoost to generate the best rule set.

SVM + RF is aimed towards SVM using RF rule induction technique to develop an affordable and feasible assessment rule. They extract SVs from a trained SVM, predict the class label for the SVs extracted, creating a new dataset. Finally they give the dataset created to the RF to extract rules [39].

SVM + CC is the combination of SVM with evolutionary algorithms, in particular they utilize Cooperative Coevolution (CC), an approach that models an ecosystem consisting of two or more species[46]. SVM + CC first models the output of the model data s relabeled samples or by support vectors then The CC will train on these data and output the rules of classification in an IF-THEN format[40].

A proposal regarding rule extraction from SVR works in a similar way as the above proposal, with an additional case. In the first step of the method they extract SVs. In the second phase they extract the rules feeding the SVs to a rule generator, that can be CART[12], ANFIS[13] and DENFIS[14]

ANN

QFAM + GA is a two step technique that aims to use Genetic Algorithms to extract rules from a Fuzzy ARTMAP(FAM) network. The proposed work equips FAM with a reinforcement learning technique(Q-learning) to produce an incremental learning classifier known as QFAM. After the training of QFAM they use a GA to extract a set of if-then rules [19].

ANN + MGA + AR is a proposed technique to extract rules from an ANN using multi-objective genetic algorithm and the association rules. The ANN is first trained with discretized attributes and then used by the rule extraction algorithm that transforms the ANN classification into comprehensible rules. The rule extraction module is composed of two sub-modules: multi-objective genetic algorithm (MGA) and association rules mining (AR) [41].

[12] Classification and regression tree
[13] Adaptive network based fuzzy inference system
[14] Dynamic evolving neuro-fuzzy inference system

The last found eclectic proposal is an algorithm aimed to extract knowledge from a modified fuzzy min–max neural network (MFMMN) [42].

Pedagogical: The pedagogical is the approach with the smaller number of techniques proposed, in the analyzed literature.

A notable proposed technique is the ALPA Algorithm, an algorithm based on active learning, where they define active learning as *we do not have to limit ourselves to the original data: since we are using an oracle, we can generate new artificial data points and their predictions (provided by the oracle) without restrictions* [21]. The central idea of ALPA is to train a whitebox algorithm using the prediction of the blackbox analyzed and not the original target values of the dataset. They conclude ALPA is the first to be applicable to any black-box model while using advanced algorithmic concepts [21]. The ALPA algorithm can be used, with some modifications, to cope with regression problems, using what the authors define ALPA-R [24].

Another technique that can explain any blackbox is LIME, an algorithm that can explain the predictions of any classifier or regressor in a faithful way, by approximating it locally with an interpretable model [28].

Doctor XAI is a novel explainability technique able to deal with multi-labeled, sequential, ontology linked data. It first generates local synthetic neighborhood around the selected data exploiting the semantic information encoded in the ontology and uses the black-box model to label it. Then it transforms the data into a format suitable to train a decision tree. This transformation allows taking the sequential nature of the data into account. Finally, Doctor XAI trains a decision tree on the labeled synthetic neighborhood, and it extracts an explanation in the form of a decision rule [34].

Trepan Reloaded is a proposal based on Trepan [47]. The aim of Trepan Reloaded is to explain ANNs by means of decision trees, objective obtained extending Trepan to take into account ontologies in the tree extraction process [17]. Another proposed technique is a methodology for extraction of knowledge from the trained ANN forecasting model, where the the inputs and outputs(target) of the ANN model were used as the inputs and the targets for decision tree induction [27].

4.2.4. Q4 - What are the quality metrics of the extracted knowledge?: Given that the knowledge extraction techniques have the goal to extract symbolic rules from blackboxes, the evaluation of the extracted rules is an important requirement to estimate the quality of the extraction method, because good extracted rules are directly related to the motivations addressed by Q1. In the literature there are many ways used to asses the quality of the rules, the more utilized being the following:

- Fidelity: fidelity is the ability to extract rules from a blackbox so that they mimic accurately the model from which they are extracted [16, 20, 21, 24, 32, 35, 37, 38, 41, 42]. Some examples of the fidelity assessment are, in the case of regression the use of the Root Mean Squared Deviation to measure agreement between the rules and the black-box [24] and, in the case of classification, the use of F1-score with respect to the outcome of the blackbox [20]

- Accuracy: accuracy is used to express how well the extracted rules are able to make accurate predictions on new cases [16, 18, 20, 21, 24, 35, 37, 38, 41, 42, 44]. Some examples of the accuracy assessment are, in the case of regression the use of Root Mean Squared Error of the true value versus the predicted value [24] and, in the case of classification, the use of F1-score [20]

- Comprehensibility: the comprehensibility [16, 20, 21, 24, 32, 37, 41, 42], also named interpretability [18, 20, 22], is the description of how the data extracted from the model are human-comprehensible [16, 18, 20, 24, 32, 37, 38], i.e. a rule must be comprehensible for it to be interpretable [24].

  Although vastly utilized in the literature, it is a subjective metric [18]. Some measures used for the model comprehensibility are the number of rules and the number of premises [16, 18, 41, 42], or the definition of a Global comprehensibility and Individual comprehensibility [37, 38] or the consideration of the the number of refined rules, the active number of features, and the coverage of rules in the system [22].

Some other metrics defined in the analyzed literature are the confidence [15, 30, 43], that is used to measure the precision of extracted rules [15], the support [15], utilized to eliminate uninteresting rules [15], effectiveness [30], sensitivity and specificity [36]

5. Quality Assessment

Using the criteria defined in section 2.6, in this phase we assessed the quality of the works found during the above phases. The score associated to each criteria is the following

- Number of citations: if a paper has less than 15 citations it's given 0.35 points, if its between 15 and 43 it's given 0.70 points and if it has more it's given 1 point.
- Usability of the work: if the source code (SC) is available, the paper is given 0.5 points. If it has pseudo-code (PC), the paper is given 0.25 points. If there is a library (L) available, the paper is given 0.25 points. The points in this criteria can be added one together.
- Type of work: if the paper proposes a new algorithm (NA) it is given 0.5 points. If the algorithm is an extension (EX) of a previous work, it is given 0.3 points. If there is a practical application (PA) of the algorithm, it is given 0.2 points. The points in this criteria can be added one together.
- Impact factor of the publishing journal: if the journal has an impact factor less than 0.97 it's given 0.35 points, if it's between 0.97 and 1.67 it's given 0.7 points and if it has more it's given 1 point.
- Number of questions answered: each paper it's given 0.25 point for each research question it has information of.

Using the above definition, each paper has the possibility to reach a total of 5 points. Table 5 presents the score for each paper.

Id	Number of citations	Usability of the work			Type of work			Impact factor	Number of questions answered				Total
		SC	PC	L	NA	EX	PA		Q1	Q2	Q3	Q4	
C1	0.35	0	0	0	0.5	0	0.2	0.7	0.25	0.25	0.25	0.25	2.75
C6	0.35	0	0.25	0	0.5	0	0.2	0.7	0.25	0	0.25	0.25	2.75
C7	0.35	0	0.25	0	0.5	0	0.2	0.7	0.25	0	0.25	0.25	2.75
C8	0.35	0	0.25	0	0.5	0	0.2	0.7	0.25	0.25	0.25	0.25	3.00
C9	0.35	0	0.25	0	0	0.3	0.2	0.35	0.25	0	0.25	0	1.85
C10	0.35	0	0	0	0	0	0	0.7	0.25	0.25	0	0.25	1.80
C11	0.35	0	0	0	0.5	0	0.2	1	0.25	0.25	0.25	0	2.80
C12	0.35	0	0.25	0	0.5	0	0	0.35	0.25	0	0.25	0	2.20
C13	0.35	0	0.25	0	0.5	0.3	0.2	0.35	0.25	0.25	0.25	0.25	2.95
C14	1	0	0	0	0	0.3	0.2	0.7	0.25	0.25	0.25	0	2.95
C15	1	0	0	0	0	0	0	1	0.25	0.25	0	0.25	2.75
C16	0.7	0	0.25	0.25	0.5	0	0.2	1	0.25	0.25	0.25	0.25	3.90
C17	0.35	0	0.25	0	0.5	0	0.2	0.7	0.25	0	0.25	0.25	2.75
C18	1	0	0.25	0	0.5	0	0.2	1	0.25	0	0.25	0	3.45
C19	0.35	0	0	0	0.5	0	0.2	0.7	0.25	0.25	0.25	0	2.50
C20	0.7	0	0	0	0	0.3	0.2	0.7	0	0.25	0.25	0.25	2.65
C21	0.35	0	0.25	0	0	0.3	0.2	0.35	0	0.25	0.25	0.25	2.20
C24	1	0	0.25	0	0.5	0	0.2	0.35	0.25	0	0.25	0	2.80
C25	0.35	0	0	0	0	0	0.2	0.35	0.25	0.25	0	0.25	1.65
C27	0.35	0	0	0	0.5	0	0.2	0.7	0	0	0.25	0.25	2.25
C28	0.7	0	0	0	0	0.3	0.2	1	0.25	0	0.25	0	2.70
C29	1	0	0	0	0.5	0	0.2	0.7	0	0.25	0.25	0	2.90
C30	1	0	0.25	0	0.5	0	0.2	0.7	0.25	0	0.25	0	3.15
C31	0.7	0	0.25	0	0.5	0	0.2	0.7	0.25	0	0.25	0.25	3.20
C32	0.35	0	0.25	0.25	0.5	0.3	0.2	0.35	0.25	0.25	0.25	0.25	3.10
C33	1	0	0	0	0.5	0	0.2	0.7	0	0.25	0.25	0	2.90
C34	0.35	0	0.25	0	0.5	0	0.2	0.35	0.25	0.25	0.25	0.25	2.65
C35	0.7	0	0	0	0	0	0.2	1	0.25	0.25	0	0.25	2.65
C36	0.7	0	0	0	0.5	0	0.2	0.7	0.25	0	0.25	0	2.60
C37	0.35	0	0.25	0	0	0.3	0.2	1	0.25	0.25	0.25	0	2.85

Table 5 Table with score by paper

6. Conclusion

The realization of an SLR regarding a specific field, is an important and useful, although difficult, task to perform. Making a It is useful and important because, often, the literature regarding a topic is very broad, with ambiguous terms and definitions. Systematically reviewing the literature can help summarize the information and to overcome the ambiguity found in each work. However the subjectivity of the metrics used to evaluate the literature, along with the workload needed to perform an SLR, can make this type of works daunting.

In this work we analyzed the literature searching information about knowledge extraction, with the aim to clarify and make a summary of this research field.

References

[1] C. Cortes, V. Vapnik, Support-vector networks, Machine learning 20 (3) (1995) 273–297.

[2] G. E. Hinton, How neural networks learn from experience, Scientific American 267 (3) (1992) 144–151.

[3] J. Quinlan, Decision trees as probabilistic classifiers, in: P. Langley (Ed.), Proceedings of the Fourth International Workshop on MACHINE LEARNING, Morgan Kaufmann, 1987, pp. 31–37. doi:https://doi.org/10.1016/B978-0-934613-41-5.50007-6.
URL https://www.sciencedirect.com/science/article/pii/B9780934613415500076

[4] R. Guidotti, A. Monreale, S. Ruggieri, F. Turini, F. Giannotti, D. Pedreschi, A survey of methods for explaining black box models, ACM Comput. Surv. 51 (5) (Aug. 2018). doi:10.1145/3236009.
URL https://doi.org/10.1145/3236009

[5] K. I. Qazi, H. Lam, B. Xiao, G. Ouyang, X. Yin, Classification of epilepsy using computational intelligence techniques, CAAI Transactions on Intelligence Technology 1 (2) (2016) 137–149. doi:https://doi.org/10.1016/j.trit.2016.08.001.
URL https://www.sciencedirect.com/science/article/pii/S2468232216300142

[6] A. Zakaryazad, E. Duman, A profit-driven artificial neural network (ann) with applications to fraud detection and direct marketing, Neurocomputing 175 (2016) 121–131. doi:https://doi.org/10.1016/j.neucom.2015.10.042.
URL https://www.sciencedirect.com/science/article/pii/S0925231215015015

[7] R. Andrews, J. Diederich, A. B. Tickle, Survey and critique of techniques for extracting rules from trained artificial neural networks, Knowledge-Based Systems 8 (6) (1995) 373–389, knowledge-based neural networks. doi:https://doi.org/10.1016/0950-7051(96)81920-4.
URL https://www.sciencedirect.com/science/article/pii/0950705196819204

[8] K. Krawiec, R. Słowiński, I. Szcześniak, Pedagogical method for extraction of symbolic knowledge, in: International Conference on Rough Sets and current Trends in Computing, Springer, 1998, pp. 436–443.

[9] M. W. Craven, J. W. Shavlik, Using sampling and queries to extract rules from trained neural networks, in: In Proceedings of the Eleventh International Conference on Machine Learning, Morgan Kaufmann, 1994, pp. 37–45.

[10] R. Calegari, G. Ciatto, A. Omicini, On the integration of symbolic and sub-symbolic techniques for xai: A survey, Intelligenza Artificiale 14 (1) (2020) 7–32.

[11] B. Kitchenham, S. Charters, Guidelines for performing systematic literature reviews in software engineering (2007).

[12] C. Wohlin, Guidelines for snowballing in systematic literature studies and a replication in software engineering, in: Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering, EASE '14, Association for Computing Machinery, New York, NY, USA, 2014. doi:10.1145/2601248.2601268.
URL https://doi.org/10.1145/2601248.2601268

[13] ScienceDirect, What does "Relevance" mean in sciencedirect? (2019).
URL https://service.elsevier.com/app/answers/detail/a_id/28453/supporthub/sciencedirect/~/what-does-relevance-mean-in-sciencedirect%3F/

[14] M. E. McVeigh, Open access journals in the ISI citation databases: analysis of impact factors and citation patterns: a citation study from Thomson Scientific, Thomson Scientific Philadelphia, 2004.

[15] N. Singh, P. Singh, D. Bhagat, A rule extraction approach from support vector machines for diagnosing hypertension among diabetics, Expert Systems with Applications 130 (2019) 188–205. doi:https://doi.org/10.1016/j.eswa.2019.04.029.
URL https://www.sciencedirect.com/science/article/pii/S0957417419302593

[16] D. Yedjour, A. Benyettou, Symbolic interpretation of artificial neural networks based on multiobjective genetic algorithms and association rules mining, Applied Soft Computing 72 (2018) 177–188. doi:https://doi.org/10.1016/j.asoc.2018.08.007.
URL https://www.sciencedirect.com/science/article/pii/S1568494618304551

[17] R. Confalonieri, T. Weyde, T. R. Besold, F. Moscoso del Prado Martín, Using ontologies to enhance human understandability of global post-hoc explanations of black-box models, Artificial Intelligence 296 (2021) 103471. doi:https://doi.org/10.1016/j.artint.2021.103471.
URL https://www.sciencedirect.com/science/article/pii/S0004370221000229

[18] C. He, M. Ma, P. Wang, Extract interpretability-accuracy balanced rules from artificial neural networks: A review, Neurocomputing 387 (2020) 346–358. doi:https://doi.org/10.1016/j.neucom.2020.01.036.
URL https://www.sciencedirect.com/science/article/pii/S0925231220300801

[19] F. Pourpanah, C. P. Lim, J. M. Saleh, A hybrid model of fuzzy artmap and genetic algorithm for data classification and rule extraction, Expert Systems with Applications 49 (2016) 74–85. doi:https://doi.org/10.1016/j.eswa.2015.11.009.
URL https://www.sciencedirect.com/science/article/pii/S0957417415007691

[20] R. Guidotti, A. Monreale, S. Ruggieri, F. Turini, F. Giannotti, D. Pedreschi, A survey of methods for explaining black box models, ACM Comput. Surv. 51 (5) (Aug. 2018). doi:10.1145/3236009.
URL https://doi.org/10.1145/3236009

[21] E. Junqué de Fortuny, D. Martens, Active learning-based pedagogical rule extraction, IEEE Transactions on Neural Networks and Learning Systems 26 (11) (2015) 2664–2677. doi:10.1109/TNNLS.2015.2389037.

[22] S. Wang, Y. Wang, D. Wang, Y. Yin, Y. Wang, Y. Jin, An improved random forest-based rule extraction method for breast cancer diagnosis, Applied Soft Computing Journal 86, cited By 11 (2020). doi:10.1016/j.asoc.2019.105941.
URL https://www.scopus.com/inward/record.uri?eid=2-s2.0-85076576728&doi=10.1016%2fj.asoc.2019.105941&partnerID=40&md5=a4a462f40ec291c7d6ad164cb6a8b623

[23] Q. Zhang, Y. Yang, H. Ma, Y. N. Wu, Interpreting cnns via decision trees, in: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 6254–6263. doi:10.1109/CVPR.2019.00642.

[24] E. J. de Fortuny, D. Martens, Active learning based rule extraction for regression, in: 2012 IEEE 12th International Conference on Data Mining Workshops, 2012, pp. 926–933. doi:10.1109/ICDMW.2012.13.

[25] N. Tsopze, E. Mephu-Nguifo, G. Tindo, Towards a generalization of decompositional approach of rule extraction from multilayer artificial neural network, in: The 2011 International Joint Conference on Neural Networks, 2011, pp. 1562–1569. doi:10.1109/IJCNN.2011.6033410.

[26] Q. Zeng, H. Huang, X. Pei, S. Wong, M. Gao, Rule extraction from an optimized neural network for traffic crash frequency modeling, Accident Analysis & Prevention 97 (2016) 87–95. doi:https://doi.org/10.1016/j.aap.2016.08.017.
URL https://www.sciencedirect.com/science/article/pii/S0001457516303037

[27] V. K. Vidyarthi, A. Jain, Knowledge extraction from trained ann drought classification model, Journal of Hydrology 585 (2020) 124804. doi:https://doi.org/10.1016/j.jhydrol.2020.124804.
URL https://www.sciencedirect.com/science/article/pii/S002216942030264X

[28] M. T. Ribeiro, S. Singh, C. Guestrin, ”why should i trust you?”: Explaining the predictions of any classifier, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD ’16, Association for Computing Machinery, New York, NY, USA, 2016, p. 1135–1144. doi:10.1145/2939672.2939778.
URL https://doi.org/10.1145/2939672.2939778

[29] M. Farquad, V. Ravi, S. B. Raju, Support vector regression based hybrid rule extraction methods for forecasting, Expert Systems with Applications 37 (8) (2010) 5577–5589. doi:https://doi.org/10.1016/j.eswa.2010.02.055.
URL https://www.sciencedirect.com/science/article/pii/S0957417410000941

[30] J. Yu, G. Liu, Extracting and inserting knowledge into stacked denoising auto-encoders, Neural Networks 137 (2021) 31–42. doi:https://doi.org/10.1016/j.neunet.2021.01.010.
URL https://www.sciencedirect.com/science/article/pii/S0893608021000186

[31] F. Harder, T. R. Besold, Learning Łukasiewicz logic, Cognitive Systems Research 47 (2018) 42–67. doi:https://doi.org/10.1016/j.cogsys.2017.07.004.
URL https://www.sciencedirect.com/science/article/pii/S1389041717300451

[32] V. K. Chan, C. W. Chan, Towards explicit representation of an artificial neural network model: Comparison of two artificial neural network rule extraction approaches, Petroleum 6 (4) (2020) 329–339, sI: Artificial Intelligence (AI), Knowledge-based Systems (KBS), and Machine Learning (ML). doi:https://doi.org/10.1016/j.petlm.2019.11.005.
URL https://www.sciencedirect.com/science/article/pii/S2405656119301051

[33] S. N. Tran, A. S. d'Avila Garcez, Deep logic networks: Inserting and extracting knowledge from deep belief networks, IEEE Transactions on Neural Networks and Learning Systems 29 (2) (2018) 246–258. doi:10.1109/TNNLS.2016.2603784.

[34] C. Panigutti, A. Perotti, D. Pedreschi, Doctor xai: An ontology-based approach to black-box sequential data classification explanations, in: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, FAT* '20, Association for Computing Machinery, New York, NY, USA, 2020, p. 629–639. doi:10.1145/3351095.3372855.
URL https://doi.org/10.1145/3351095.3372855

[35] V. Chan, C. Chan, Learning from a carbon dioxide capture system dataset: Application of the piecewise neural network algorithm, Petroleum 3 (1) (2017) 56–67, carbon Capture and Storage (CCUS). doi:https://doi.org/10.1016/j.petlm.2016.11.004.
URL https://www.sciencedirect.com/science/article/pii/S2405656116301997

[36] L. Özbakır, A. Baykasoğlu, S. Kulluk, A soft computing-based approach for integrated training and rule extraction from artificial neural networks: Difaconn-miner, Applied Soft Computing 10 (1) (2010) 304–317. doi:https://doi.org/10.1016/j.asoc.2009.08.008.
URL https://www.sciencedirect.com/science/article/pii/S1568494609001318

[37] M. Augasta, T. Kathirvalavakumar, Reverse engineering the neural networks for rule extraction in classification problems, Neural Processing Letters (04 2011). doi:10.1007/s11063-011-9207-8.

[38] M. G. Augasta, T. Kathirvalavakumar, Rule extraction from neural networks — a comparative study, in: International Conference on Pattern Recognition, Informatics and Medical Engineering (PRIME-2012), 2012, pp. 404–408. doi:10.1109/ICPRIME.2012.6208380.

[39] L. Han, S. Luo, J. Yu, L. Pan, S. Chen, Rule extraction from support vector machines using ensemble learning approach: An application for diagnosis of diabetes, IEEE Journal of Biomedical and Health Informatics 19 (2) (2015) 728–734. doi:10.1109/JBHI.2014.2325615.

[40] R. Stoean, C. Stoean, Modeling medical decision making by support vector machines, explaining by rules of evolutionary algorithms with feature selection, Expert Systems with

Applications 40 (7) (2013) 2677–2686. doi:https://doi.org/10.1016/j.eswa.2012.11.007.
URL https://www.sciencedirect.com/science/article/pii/S0957417412012171

[41] D. Yedjour, Extracting classification rules from artificial neural network trained with discretized inputs, Neural Processing Letters 52 (3) (2020) 2469–2491, cited By 0. doi:10.1007/s11063-020-10357-x.
URL https://www.scopus.com/inward/record.uri?eid=2-s2.0-85092385087&doi=10.1007%2fs11063-020-10357-x&partnerID=40&md5=af7df9621bf13e17b83a91f133c1bba1

[42] S. Shinde, U. Kulkarni, Extracting classification rules from modified fuzzy min–max neural network for data with mixed attributes, Applied Soft Computing 40 (2016) 364–378. doi:https://doi.org/10.1016/j.asoc.2015.10.032.
URL https://www.sciencedirect.com/science/article/pii/S1568494615006729

[43] J. Yu, G. Liu, Knowledge-based deep belief network for machining roughness prediction and knowledge discovery, Computers in Industry 121 (2020) 103262. doi:https://doi.org/10.1016/j.compind.2020.103262.
URL https://www.sciencedirect.com/science/article/pii/S0166361520304966

[44] Y. Hayashi, R. Setiono, A. Azcarraga, Neural network training and rule extraction with augmented discretized input, Neurocomputing 207 (2016) 610–622. doi:https://doi.org/10.1016/j.neucom.2016.05.040.
URL https://www.sciencedirect.com/science/article/pii/S0925231216303927

[45] L. Borkowski, et al., On three-valued logic, Selected Works by Jan Łukasiewicz; North–Holland: Amsterdam, The Netherlands (1970) 87–88.

[46] M. A. Potter, K. A. D. Jong, Cooperative coevolution: An architecture for evolving coadapted subcomponents, Evolutionary computation 8 (1) (2000) 1–29.

[47] M. Craven, J. Shavlik, Extracting tree-structured representations of trained networks, Advances in neural information processing systems 8 (1995) 24–30.