Computer Modeling in
Engineering & Sciences

Tech Science Press

DOI: 10.32604/cmes.2023.023897

REVIEW

# Explainable Rules and Heuristics in AI Algorithm Recommendation Approaches—A Systematic Literature Review and Mapping Study

**Francisco José García-Peñalvo*, Andrea Vázquez-Ingelmo and Alicia García-Holgado**

GRIAL Research Group, Computer Science Department, University of Salamanca, Salamanca, 37008, Spain

*Corresponding Author: Francisco José García-Peñalvo. Email: fgarcia@usal.es

Received: 17 May 2022 Accepted: 08 October 2022

**ABSTRACT**

The exponential use of artificial intelligence (AI) to solve and automated complex tasks has catapulted its popularity generating some challenges that need to be addressed. While AI is a powerful means to discover interesting patterns and obtain predictive models, the use of these algorithms comes with a great responsibility, as an incomplete or unbalanced set of training data or an unproper interpretation of the models' outcomes could result in misleading conclusions that ultimately could become very dangerous. For these reasons, it is important to rely on expert knowledge when applying these methods. However, not every user can count on this specific expertise; non-AI-expert users could also benefit from applying these powerful algorithms to their domain problems, but they need basic guidelines to obtain the most out of AI models. The goal of this work is to present a systematic review of the literature to analyze studies whose outcomes are explainable rules and heuristics to select suitable AI algorithms given a set of input features. The systematic review follows the methodology proposed by Kitchenham and other authors in the field of software engineering. As a result, 9 papers that tackle AI algorithm recommendation through tangible and traceable rules and heuristics were collected. The reduced number of retrieved papers suggests a lack of reporting explicit rules and heuristics when testing the suitability and performance of AI algorithms.

**KEYWORDS**

SLR; systematic literature review; artificial intelligence; machine learning; algorithm recommendation; heuristics; explainability

## Nomenclature

| | |
|---|---|
| AI | Artificial intelligence |
| SLR | Systematic literature review |
| MQ | Mapping question |
| RQ | Research question |
| IC | Inclusion criteria |
| EC | Exclusion criteria |
| C5T | Quinlan's C5.0 decision tree |
| C5R | Quinlan's C5.0 rule inducer |
| MLP | Multilayer perceptron |

| | |
|---|---|
| RBF | Radial basis function network |
| LDS | J. Gama's linear discriminant |
| LTR | J. Gama's linear tree |
| IBL | Instance-based learner |
| NB | Naïve Bayes |
| RIP | Ripper |
| KD | Kernel density |
| SVM | Support vector machine |
| NN | Neural networks |
| ANN | Artificial neural network |
| LR | Linear regression |
| KNN | K-nearest neighbors |
| DT | Decision tree |
| RF | Random forest |
| FCBF | Fast correlation based filter |
| mRMR | Minimum redundancy maximum relevance |
| IG | Information gain |
| GR | Gain ratio |
| DR | Dimensionality reduction |
| MLP | Multi layer perceptron |
| DSL | Domain specific language |

## 1 Introduction

The increasing use of artificial intelligence (AI) to tackle a wide range of problems has opened this discipline to many people interested in solving or automating tasks through AI algorithms. Although this field has expanded over the years, applying these algorithms is not straightforward and requires expert knowledge in different senses: (1) knowledge regarding the input data, i.e., the data used to feed the AI algorithms, (2) knowledge regarding AI methods to get the most out of their application, and (3) knowledge to understand and explain the outcomes from AI models. Having this knowledge is crucial to obtaining valuable results from AI. Otherwise, the outputs could lead to wrong conclusions, losses, discrimination, and even negligence [1–4].

However, several non-expert users could benefit from applying AI to their domain problems [5–7]. These non-experts would understand the domain and relationships of the input data but lack the skills to use or select the proper algorithm for their tasks. For these reasons, several frameworks and tools have arisen to assist and give suggestions to novice users in the journey of applying and interpreting AI algorithms [8–14].

But while these tools provide robust support to non-expert users, they can lack a didactic dimension that could enrich the experience and yield more benefits in the medium-long term. AI algorithm recommendation could be an obscure realm in which several (and powerful) black-box approaches are provided but without clear feedback regarding traceability and explainability of the rules followed in choosing a proper model.

As stated at the beginning, to avoid wrong conclusions, it is important to understand the implications of selecting different algorithms or at least to understand why an algorithm is better in certain domains than others. For these reasons, relying on readable heuristics or explainable rules can improve transparency and bridge the gap between domain and AI experts.

This work presents a systematic literature review (SLR) [15] on heuristics and rules obtained from approaches focused on tackling the AI algorithm selection or recommendation problem. Our main goal is to discuss the literature landscape of heuristics related to AI algorithm recommendations and to explore transparent recommendation methodologies to assist and educate non-expert users in selecting the right AI algorithm for their data. The SLR provides a traceable methodology to identify existing works that tackle this issue, allowing us to analyze gaps and challenges within this context.

To sum up, the main contribution of this work is:

- The identification and analysis of existing AI selection algorithms that explain their internal rules or follow readable heuristics

Presenting the internal decisions of an AI algorithm is not only beneficial for building more reliable systems but also for visually understanding the complex procedure of selecting an AI algorithm. To sum up, the main contribution of this paper is a summary and discussion of the use of explainable heuristics when selecting AI algorithms.

The rest of this paper is structured as follows. Section 2 presents the SLR methodology proposed by Kitchenham et al. [16,17]. Section 3 outlines the review planning, while Section 4 details the review process protocol. Sections 5 and 6 present the systematic literature mapping and review results, respectively. Finally, these results are discussed in Section 7, following the limitations of the study (Section 8) and the conclusions derived from this work (Section 9).

## 2 Methodology

We followed a systematic process for the present review; the systematic literature review (SLR) methodology by Kitchenham et al. [16,17], complemented with the approach by García-Holgado et al. [18]. The SLR has been complemented with a quantitative study by carrying out a systematic mapping of the literature following the methodology proposed in [19].

The SLR follows a clearly defined protocol using transparent and well-defined criteria, allowing replicability of the results. Every outcome from the different steps is accessible and can be consulted through the resources shared in the following subsections. The main goal is to answer previously defined research questions by identifying, selecting, and evaluating existing research.

This section provides every detail of the protocol used to carry out the SLR to make the obtained results properly traceable. We conducted the SLR following three main phases [16,17]: planning, conducting, and reporting of the study.

The planification of the present SLR started after verifying that no recent SLRs regarding AI algorithm selection rules and heuristics were previously conducted. This verification was performed by searching through different electronic databases (Scopus and Web of Science) terms related to the methodology ("SLR", "systematic literature review", etc.) and the target of the review ("heuristics", "rule-based", etc., along with the term "AI algorithm selection OR recommendation"). Specifically, the following search strings were employed:

1) SCOPUS

*TITLE-ABS-KEY (("SLR" OR "systematic literature review") AND ("heuristic∗" OR "rule-based") AND ("algorithm∗ recommend∗" OR "algorithm∗ config∗" OR "algorithm∗ selection" OR "algorithm∗ selector")) AND NOT DOCTYPE(cr)*

2) WEB OF SCIENCE

*TS=(("SLR" OR "systematic literature review") AND ("heuristic∗" OR "rule-based") AND ("algorithm∗ recommend∗" OR "algorithm∗ config∗" OR "algorithm∗ selection" OR "algorithm∗ selector"))*

Some SLRs about algorithm recommendation and meta-learning were found; however, none of these were focused on addressing tangible rules or heuristics for their selection, justifying the execution of this SLR.

## 3 Review Planning

In this section, the basic aspects of the review are defined: the research questions, the protocol followed, and every detail to make this review traceable.

### *3.1 Research Questions*

We posed the following research questions to be answered by the selected papers:

- **RQ1.** Which methods have been applied to support AI selection algorithms?
- **RQ2.** What categories and instances of AI are available to choose in selection processes?
- **RQ3.** What factors determine the selection of an AI algorithm?
- **RQ4.** How is the transparency of the selection process managed?
- **RQ5.** How are the selection processes evaluated?

The first question aims at answering how the AI selection process is performed technically speaking. RQ2 is focused on the applicability of the methods regarding the type of specific algorithms that it supports and the type of problem it tackles, respectively. RQ3 and RQ4 are related to the traceability of the selection process and the decisions (rules or heuristics) followed to recommend a specific algorithm. Finally, RQ5 is posed to answer how these methods have been evaluated.

As mentioned before, the SLR has been complemented with a quantitative analysis through a literature mapping. The following mapping questions (MQs) were posed:

- **MQ1.** How many studies were published over the years?
- **MQ2.** Who are the most active authors in the area?
- **MQ3.** What type of papers are published?
- **MQ4.** Which are the factors that condition the algorithm recommendation process?
- **MQ5.** Which methods have been used for algorithm recommendation?
- **MQ6.** Which AI problems are tackled?
- **MQ7.** How many studies have tested their proposed solutions?

While the mapping provides a quantitative overview of the research area, the SLR results involve the analysis and interpretation of the selected works [20] to answer the previous research questions.

Finally, we followed the PICOC method proposed by Petticrew et al. [21] to define the review scope.

- **Population (P):** Software solutions
- **Intervention (I):** Provide support to recommending AI algorithms in a transparent way

- **Comparison (C):** No comparison intervention in this study, as the primary goal of the present SLR is to analyze existing approaches regarding AI recommendation processes and gain knowledge about them
- **Outcomes (O):** AI recommendation processes proposals
- **Context (C):** Environments related to AI or in which AI can be applied

### *3.2 Inclusion and Exclusion Criteria*

We defined a series of inclusion (IC) and exclusion criteria (EC) to filter relevant works that could answer the research questions. If a work does not meet every inclusion criterion or does meet any exclusion criterion, it will be dismissed from the review.

- **IC1.** The paper describes an ML, DL, or AI algorithm recommendation approach AND
- **IC2.** The solution supports or addresses the algorithm selection problem AND
- **IC3.** The solution is not limited to a highly specific problem AND
- **IC4.** The papers are written in English AND
- **IC5.** The papers are published in peer-reviewed Journals, Books, or Conferences AND
- **IC6.** The publication is the most recent or complete of the set of related publications regarding the same study

On the other hand, the exclusion criteria are the opposite as the inclusion criteria as their opposite.

- **EC1.** The paper does not describe an ML, DL, or AI algorithm recommendation approach OR
- **EC2.** The solution does not support or addresses the algorithm selection problem OR
- **EC3.** The solution is limited to a highly specific problem OR
- **EC4.** The papers are not written in English OR
- **EC5.** The papers are not published in peer-reviewed Journals, Books, or Conferences OR
- **EC6.** The publication is not the most recent or complete of the set of related publications regarding the same study

The IC1 and IC2 ensure that the paper is related to AI algorithm recommendation processes, while IC3 refers to the application domain. We decided to include solutions that solve a wider range of problems to obtain abstract heuristics rather than domain-specific heuristics. Although domain-specific heuristics are crucial, we wanted to extract heuristics or rules that provide a generic starting point for selecting AI algorithms. The remaining inclusion criteria are focused on other details such as the language, publication type, and completeness of the research.

### *3.3 Search Strategy*

The first step to define the search strategy is to identify relevant databases within the search domain in which the queries will be performed. This way, we will obtain outcomes aligned with the research context. In this case, two electronic databases were selected: Scopus and Web of Science (WoS). These databases were chosen according to the following criteria:

- It is a reference database in the research scope.
- It is a relevant database in the research context of this literature review.

- It allows using similar search strings to the rest of the selected databases as well as using Boolean operators to enhance the outcomes of the retrieval process.

Regarding the search query, we included the following terms:

- Machine learning, deep learning, and artificial intelligence. Although machine learning and deep learning are artificial intelligence approaches, the latter term might not be explicitly used in field-specific works.
- Terms related to algorithm selection: recommendation, configuration, selection, and selector along with the term "algorithm", which is the target of the study.

### 3.4 Query Strings

The search strings for each database were composed through terms derived from the PICOC methodology outcomes, connected by Boolean AND/OR operators, and using the wildcard (∗) to enclose the singular and plural tense of each term. Although this SLR is focused on heuristic-driven and rule-based selections, these terms were not included in the search to avoid dismissing relevant works whose outcomes provide heuristics or rules but do not necessarily include these terms in their abstracts or titles.

3) SCOPUS

*TITLE-ABS-KEY (("machine learning" OR "deep learning" OR "artificial intelligence") AND ("algorithm∗ recommend∗" OR "algorithm∗ config∗" OR "algorithm∗ selection" OR "algorithm∗ selector")) AND NOT DOCTYPE(cr)*

4) WEB OF SCIENCE

*TS=(("machine learning" OR "deep learning" OR "artificial intelligence") AND ("algorithm∗ recommend∗" OR "algorithm∗ config∗" OR "algorithm∗ selection" OR "algorithm∗ selector"))*

### 3.5 Quality Criteria

While the inclusion and exclusion criteria are crucial to avoid including unrelated works into the literature review, they do not address the retrieved papers' quality in terms of their potential capability to answer the research questions. For these reasons, the following criteria have been defined to evaluate the retrieved works' quality before including them in the final analysis. Each criterion can be scored with three values: 1 (the paper meets the criterion), 0.5 (the paper partially meets the criterion), and 0 (the paper does not meet the criterion).

1. The research goals of the work are focused on addressing the selection of AI algorithms

- *Partial: not every research goal tries to address the selection of AI algorithms*

2. A set of rules, heuristics, or a traceable framework for selecting a proper AI algorithm is clearly exposed

- *Partial: a set of rules, heuristics or a traceable framework is described but not detailed, i.e., the methods are described, but their details are not further explained*

3. The context or domain of application is generic

- *Partial: the application context refers to a comprehensive set of problems framed in the same domain (e.g., an approach applied to the clinical context but not limited to a specific clinical disease or an approach only focused on selecting classification algorithms)*

4. The proposed solution has been tested

- *Partial: the selection or recommendation solution has been tested in terms of functionality but not in other dimensions*

5. Issues or limitations regarding the proposed solution are identified

- *Partial: issues or limitations are mentioned but not detailed*

Following these criteria, each paper can score 5 points regarding its quality following this methodology. This 0-to-5 score was mapped into a 0-to-10 scale, and the seven value was chosen as the threshold for including a paper into the final synthesis. If on a 0-to-10 scale, a paper obtains a score of fewer than seven points, it will be dismissed from the review as it did not meet a minimum quality to answer the research questions. With this threshold, a paper is limited to a maximum of one criterion with a 0 score to reach the next phase, ensuring that most criteria are always fully or almost entirely met.

## 4 Review Process

The data extraction phase of this review has been divided into different steps in which various activities are carried out. Once the search was performed–on April 13, 2021, for the **first version** and on September 22, 2022, for the **updated version**, the paper selection process was carried out through the following procedure:

1. The raw results (i.e., the records obtained from each selected database) were collected, uploaded into a GIT repository[1] and arranged into a spreadsheet[2]. A total of 1346 (940 + 406) papers were retrieved: 586 (396 + 190) from Web of Science and 760 (544 + 216) from Scopus.
2. After organizing the records, duplicate works were removed. Specifically, 521 records were excluded, leaving 825 works (61.29% of the raw records) for the next phase.
3. The candidate papers were analyzed by applying the inclusion and exclusion criteria to their titles, abstracts, and keywords 717 papers were discarded as they did not meet the criteria, retaining 109 papers (13.21% of the unique papers retrieved) for the next phase.
4. These last papers were read in detail and further analyzed. Each article was associated with a score regarding their quality to answer the research questions using the quality assessment checklist described in the previous section.
5. After applying the quality criteria, a total of 9 papers (1.09% of the unique papers retrieved and 8.82% of the full text assessed papers) were selected for the present review.

[1] https://github.com/AndVazquez/slr-ai-heuristics/tree/main/21_09_22_update.

[2] https://bit.ly/3SZ842N.

The PRISMA flow diagram has been employed to carry out the data extraction process. The PRISMA 2009 [22] flow diagram was used for the first version of the SLR (Fig. 1). In the case of the updated version, the PRISMA 2020 [23,24] guidelines were applied. Fig. 2 shows the PRISMA 2020 flow diagram for the updated version of the SLR and the previous paper selection procedure.

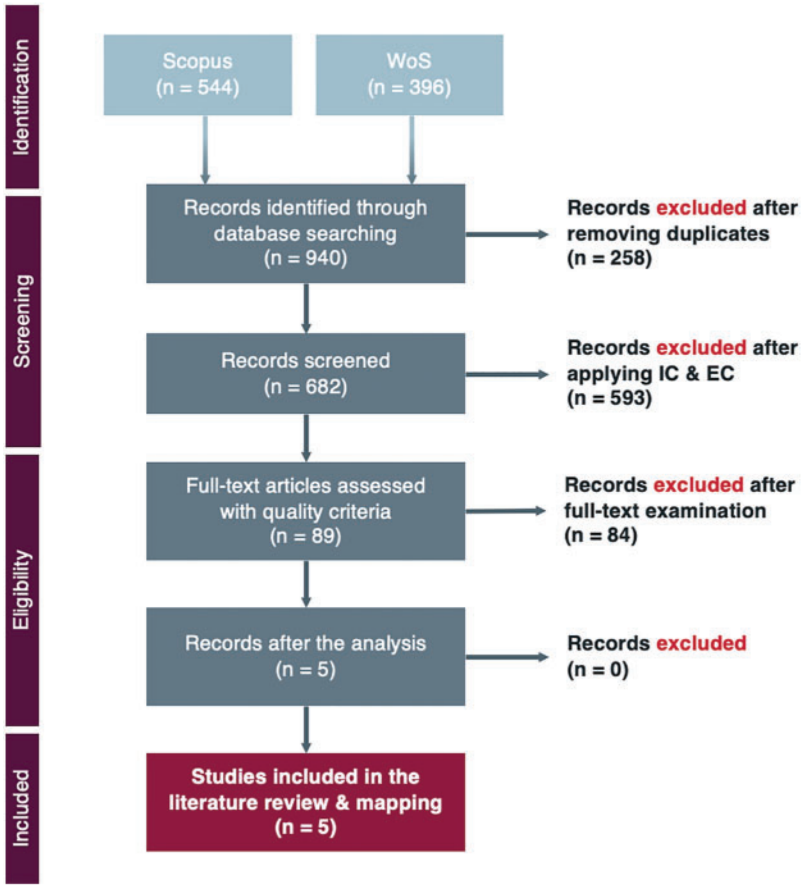


**Figure 1:** PRISMA flow diagram [22] of the data extraction process of the first review. Source: own elaboration

## 5 Systematic Literature Mapping Results

***MQ1. How many studies were published over the years?***

The number of selected papers per year is visualized in Fig. 3. From the nine selected papers, two of them belong to the early 2000s period [25,26], while the remaining belong to the last six years interval [27–33].

***MQ2. Who are the most active authors in the area?***

The following authors were retrieved, each with one paper among the selected ones, so there is not a specific author that stand out in this field: Hilario, M.; Ali, S.; Smith, K.A.; Parmezan, A.R.S.; Lee, H.D.; Wu, F.C.; Kumar, C.; Käppel, M.; Schützenmeier, N.; Eisenhuth, P.; Jablonski, S.; Martínez-Rojas, A.; Jiménez-Ramírez, A.; Enríquez, J.G.

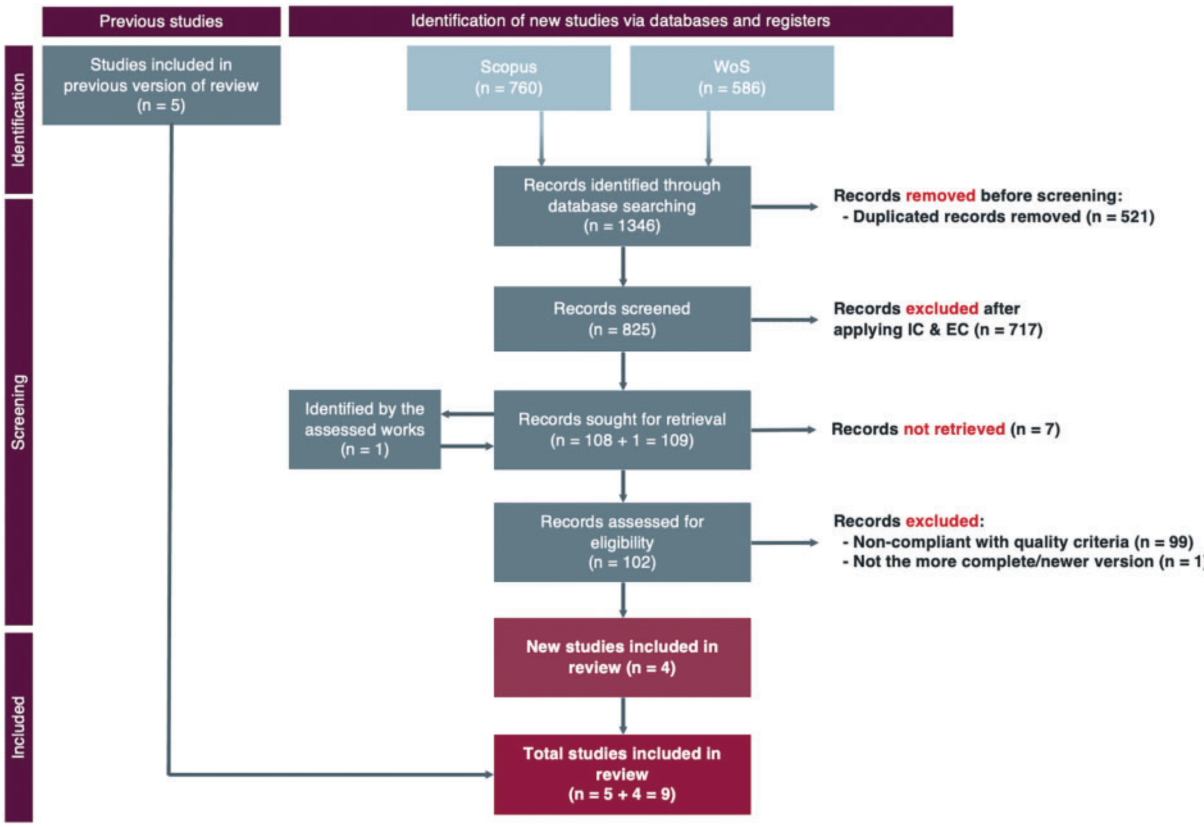


**Figure 2:** PRISMA flow diagram [23,24] of the data extraction process of the updated review. Source: own elaboration

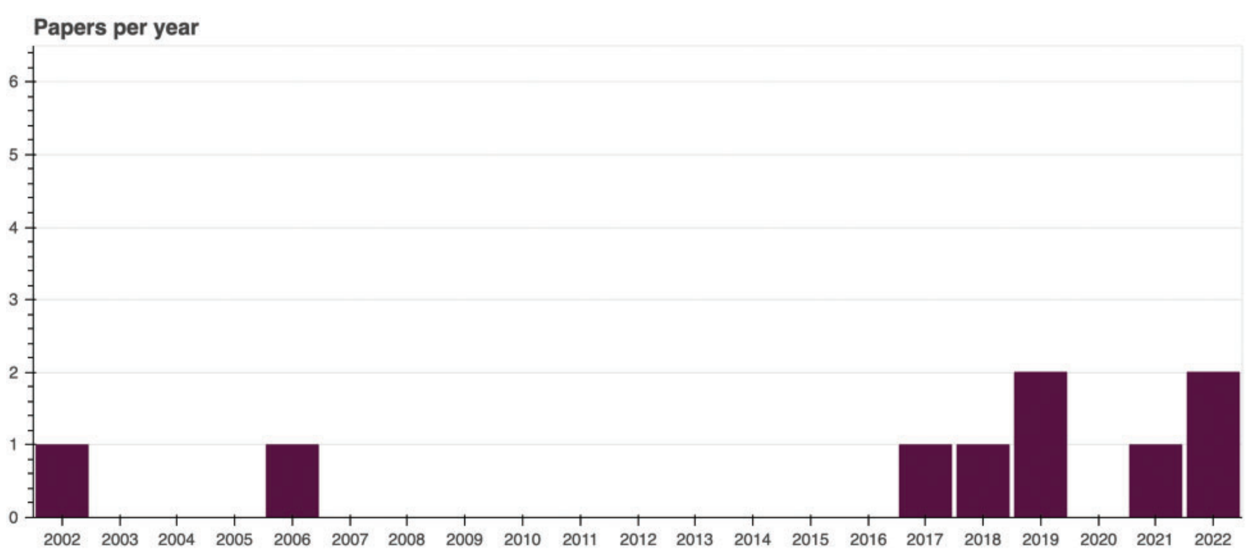


**Figure 3:** Number of papers per year. Source: own elaboration

***MQ3. What type of papers are published?***

We analyzed the metadata provided by the electronic databases to answer this mapping question. According to our inclusion and exclusion criteria, only peer-reviewed papers (either in journals, conferences, books, or workshops) were included. The complete list of types regarding the selected works can be consulted in Table 1.

**Table 1:** Papers grouped by type of publication. Source: own elaboration

| Type | Total | Papers |
|---|---|---|
| Conference paper | 5 | [25,28,29,32,33] |
| Article | 4 | [26,27,30,31] |

***MQ4. Which are the factors that condition the algorithm recommendation process?***

AI algorithm selection approaches could employ different aspects that influence the recommendation process. In this case, from the nine selected papers, we identified two main factors: dataset characteristics (present in every selected work) but also the task to be carried out through AI, present in [29,32], or the algorithm requirements and characteristics [28,31]. Table 2 summarizes the results.

**Table 2:** Papers grouped by recommendation process. Source: own elaboration

| Factor | Total | Papers |
|---|---|---|
| Data | 9 | [25–33] |
| Algorithm characteristics | 2 | [28,31] |
| Task | 2 | [29,32] |

***MQ5. Which methods have been used for algorithm recommendation?***

We identified two main methodologies regarding the implementation of explainable rules and heuristics for AI algorithm selection (Fig. 4). The first, meta-learning, is employed by 4 solutions [25–27,30]. This methodology carries out the recommendation process by training another AI model (meta-learner) to learn from dataset characteristics (meta-features) and selecting the best performant AI algorithm (meta-target).

The second methodology involves literature reviews and analysis of the application of AI algorithms to different context to obtain a set of structured guidelines and heuristics [28,29,31,32]. Finally, authors in [33] created an ontology to support the algorithm recommendation process. Table 3 shows the results of this mapping question.

***MQ6. Which AI problems are tackled?***

As can be seen in Fig. 5, the target AI algorithms of the recommendation process are mostly framed in the "classification" category [25,26,28,30–32]. But there are some papers that also provide heuristics for clustering [28], regression [28,32], and feature selection problems [27,33].

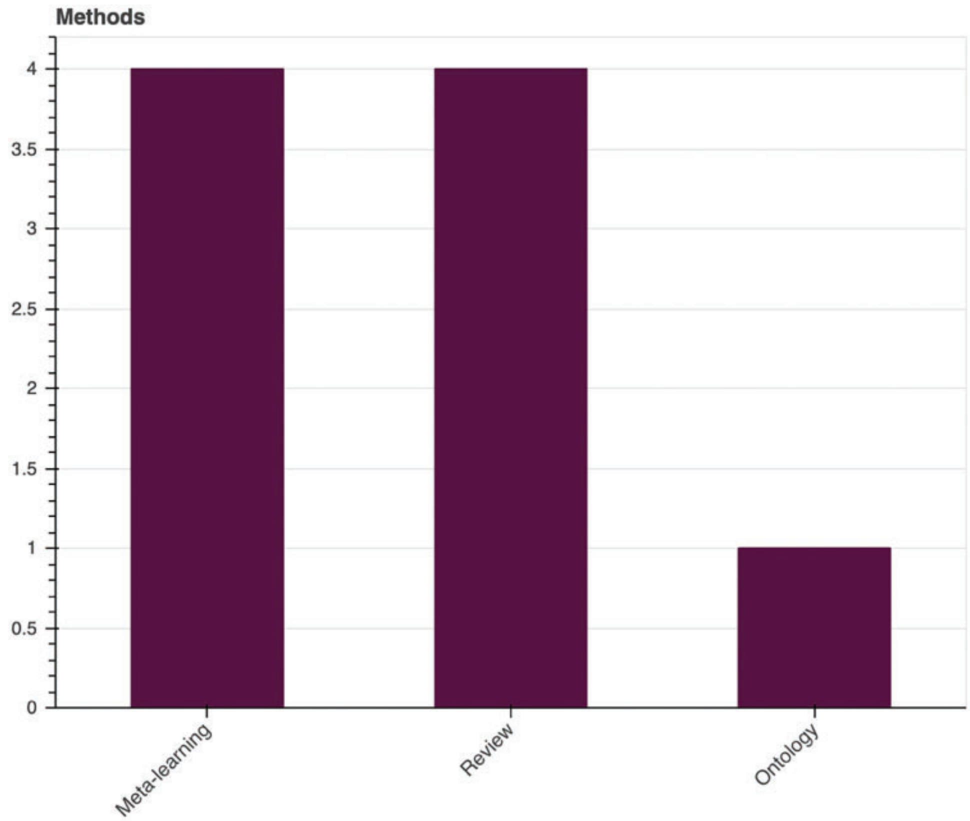


**Figure 4:** Methods employed for algorithm recommendation. Source: own elaboration

**Table 3:** Papers grouped by methodology. Source: own elaboration

| Method | Total | Papers |
|---|---|---|
| Meta-learning | 4 | [25–27,30] |
| Review | 4 | [28,29,31,32] |
| Ontology | 1 | [33] |

In addition, authors in [31] provided heuristics for an important task within AI pipelines; the encoding of categorical variables during the preprocessing tasks. The detailed categorization of the selected records can be consulted in Table 4.

***MQ7. How many studies have tested their proposed solutions?***

Only one of the papers has not been explicitly tested regarding its functionality, as they present a theoretical proposal for the unification of ML knowledge [29]. The remaining papers have tested their solutions whether in terms of performance or in terms of functionality. Table 5 presents the results.

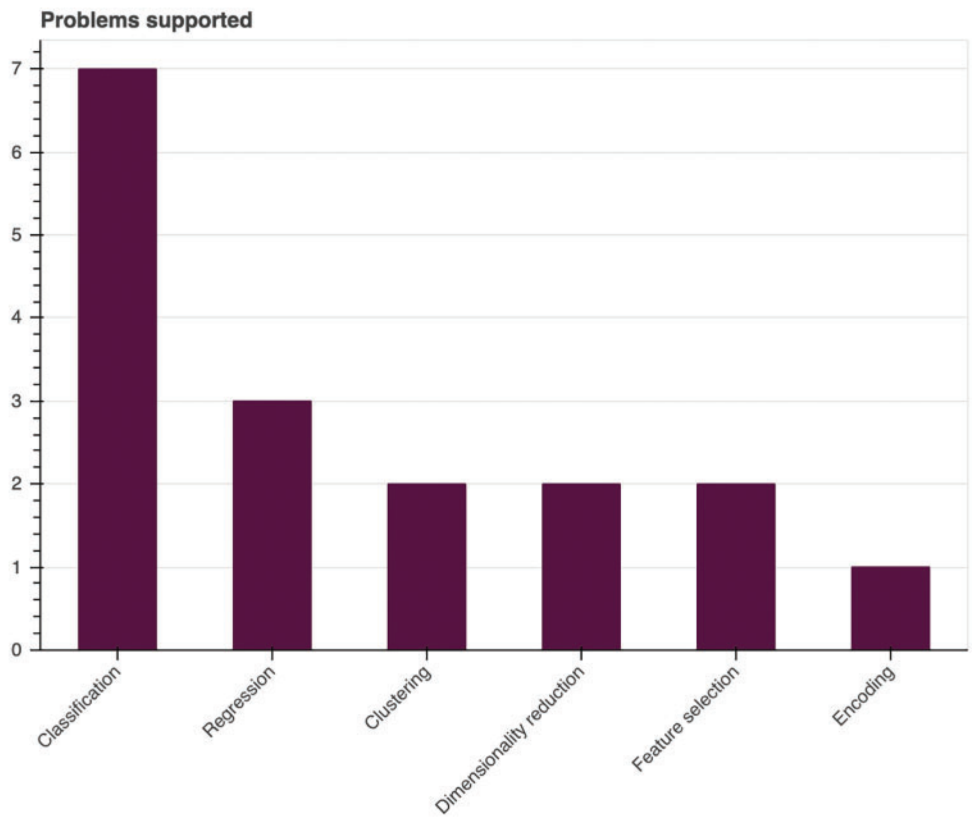


**Figure 5:** Types of AI problems addressed in the selected solutions. Source: own elaboration

**Table 4:** Papers grouped by the AI problem they solve. Source: own elaboration

| AI problem | Total | Papers |
|---|---|---|
| Classification | 7 | [25,26,28–32] |
| Regression | 3 | [28,29,32] |
| Clustering | 2 | [28,29] |
| Feature selection | 2 | [27,33] |
| Dimensionality reduction | 2 | [29,32] |
| Encoding | 1 | [31] |

**Table 5:** Papers grouped by testing maturity. Source: own elaboration

| Tested? | Total | Papers |
|---|---|---|
| Yes | 7 | [25–27,30–33] |
| No | 2 | [28,29] |

## 6 Systematic Literature Mapping Results

***RQ1. Which methods have been applied to support AI selection algorithms?***

This question aims at answering how the algorithm selection rules or heuristics have been obtained, that is, which methodologies are most widely used to extract and capture explicit rules for selecting suitable algorithms. Most of the solutions rely on meta-learning [34–40], which is also the most common method found during every step in the review process and will be further discussed in Section 7. However, the selected solutions detail the internal rules of the meta-learner, providing valuable heuristics to select AI algorithms.

For example, in [25], the author blendsed the model and instance selection problems. First, she experiments with parameter optimization to subsequently apply meta-learning and obtain rules for identifying which algorithm outperforms the rest under specific circumstances. In this case, she applied the **C5.0 algorithm** [41–43] as a meta-learner and employed 20 dataset characteristics as meta-features along with the performance of each algorithm with specific parameters.

The **C5.0 algorithm** was also applied as a meta-learner in [26]. In this work, authors first carried out an experimental study to test the performance of different classification algorithms over 100 different datasets. These results were then used as an input, along with each dataset characteristics for a decision tree classifier (C5.0) to seek for relationships between the data characterization and the algorithm performance.

The remaining meta-learning-based solution presented in [27] is applied to feature selection algorithms. The selected meta-learner was the **J48 algorithm** [44–46], and three meta-bases were derived to tackle the selection problem as binary: the first one tackle the approach (Ranking or Feature Subset Selection), and the second and third one the recommended algorithm within the approach (InfoGain or ReliefF, and CBF or CFS, respectively). This meta-learner, as well as the, also takes as an input the performance of FS algorithms as well as a set of dataset characteristics.

Similarly, authors in [30] presented a combination of **meta-learning** and a **genetic algorithm** to recommend the best classifier combinations. Their method, namely MEGA, extracts meta-features from datasets and discovers the best classifiers through a genetic algorithm.

Apart from meta-learning, literature reviews are also a powerful method for extracting rules and heuristics. This is the case of [28], in which authors exhaustively **reviewed the literature** and extracted common parameters, although the methodology for selecting reliable papers is not further detailed–however, they reference that they used "reliable sources".

The remaining work in this category combines a **review** of cheat sheets with expert knowledge to build a common language for ML algorithm recommendation [29], however, details about the implementation are undisclosed.

A similar approach is taken in [31] and [32], where authors analyzed a set of techniques (encoding techniques, in the case of [31]) and ML algorithms (classification, regression, or visualization) to obtain **heuristic guidelines**, thus allowing the systematic selection of ML methods.

Finally, an **ontology-based** approach was employed in [33], complemented with Semantic Web Rule Language (SWRL) rules to support the recommendation of AI algorithms based on the relationships of the ontology's entities.

***RQ2. What categories and instances of AI algorithms are available to choose in selection processes?***

This question focuses on the applicability of the algorithm selection, precisely, which kind of algorithms are the targets of the algorithm selection. All works focus on supervised ML techniques and, in general, the target algorithms of the selected works are framed in the "classification" category of algorithms. Table 6 summarizes the supported algorithms in each record.

**Table 6:** Papers grouped by problem and supported algorithms. Source: own elaboration

| Record | Problem(N) | Algorithms | | | |
|---|---|---|---|---|---|
| [25] | Classification (9) | C5T | RBF | IBL | |
| | | C5R | LDS | NB | |
| | | MLP | LTR | RIP | |
| [26] | Classification (8) | IBK | KD | SVM | |
| | | C4.5 | NB | NN | |
| | | PART | OneR | | |
| [27] | Feature selection (4) | CBF | InfoGain | | |
| | | CFS | ReilefF | | |
| [28] | Classification (11) | LR | KNN | RF | XGBoost |
| | | NB | SVM | AdaBoost | LightGBM |
| | | ANN | DT | CatBoost | |
| | Clustering (16) | k-Means | BIRCH | DBSCAN | Wave |
| | | PAM | CURE | OPTICS | CLIQUE |
| | | CLARA | ROCK | DENCLUE | SOM |
| | | CLARANS | Chameleon | STING | SLINK |
| | Regression (17) | LR | Elastic Net | MARS | Extreme Learning |
| | | Quantile | Ordinal | Additive | Boosted |
| | | Bayesian | Poisson | RF | |
| | | LASSO | SVR | XGBoost | |
| | | Ridge | Spline | Ensemble | |
| [33] | Feature selection (6) | FCBF | IG | Relief | |
| | | mRMR | GR | ReliefF | |
| [31] | Encoding (8) | One-hot | Contrast | James Stein | Target |
| | | Ordinal | Hash | Leave-one-out | Label |
| | Classification (6) | KNN | DT | ANN | |
| | | NB | RF | SVM | |
| [32] | Classification (15) | SVM | NN | RF | Mini Batch KMeans |
| | | NB | KMeans | VGBMM | Spectral Clustering |
| | | KNeighbor | Ensemble | KModes | Kernel Approx. |
| | | GMM | Mean shift | DR | |
| | Regression (5) | SGD | Linear SVR | Elastic net | |
| | | Ridge | Lasso | | |

(Continued)

**Table 6 (continued)**

| Record | Problem(N) | Algorithms | | | |
|---|---|---|---|---|---|
| | Dimensionality reduction (8) | Isomap | GA | NMF | Kernel Approx. |
| | | LLE | PCA | Autoencoders | Spectral Embedding |
| [30] | Classification (14) | J48<br>REP<br>Stump<br>Random tree | RF<br>JRIP<br>OneR<br>PART | NB<br>Bayes net<br>KNN<br>LWL | Decision table<br>MLP |

Finally, authors in [29] do not specify any set of algorithms, although they refer ML supervised algorithms and mention some like Linear SVC, SDG Classifier, Naïve Bayes or KN Classifier. However, they do contemplate 4 main categories of algorithms to model a knowledge source of the sci-kit learn library [47]: classification, regression, clustering, and dimensionality reduction.

***RQ3. What factors determine the selection of AI algorithm?***

There is not a one-size-fits-all algorithm that fits perfectly for any domain or dataset. The performance and suitability of AI models are tied to their specific context of application. For this reason, it is important to identify the factors or circumstances that influence the efficacy of an AI algorithm, how to measure them, and how to structure them to define heuristics.

In [25], 20 dataset characteristics are considered to feed the C5.0 meta-learner. Specifically, she mentions the following as an example: "# of instances, classes, and explanatory variables; proportion of missing values, average entropy of predictive variables, class entropy, mutual information between predictive and class variables", but the full set of dataset characteristics are not detailed.

Dataset characteristics are also the meta-features contemplated in the meta-learning solution from [26]. Three categories of characteristics are measured for each dataset (Table 7).

**Table 7:** Meta-features considered in [26]. Source: own elaboration

| Type (N) | Characteristics | | |
|---|---|---|---|
| Simple (8) | # of attributes | % of majority class | % of continuous variables |
| | # of samples<br>% of minority class | % of binary variables<br>% of discrete variables | % of missing values |
| Statistical (18) | Geometric mean<br>Harmonic mean<br>Trim mean | Percentile<br>Interquartile range<br>Max. and min. eigenvalue | Kurtosis<br>Skewness<br>Correlation coefficient |
| | Mad | Canonical correlation | Z-score |

(Continued)

**Table 7 (continued)**

| Type (N) | Characteristics | | |
|---|---|---|---|
| | Variance | Index of dispersion | Normal cumulative distribution test |
| | Standard deviation | Center of gravity | Chi-square test |
| Information theoretical (5) | Entropy of classes | Equivalent # of variables | Mean mutual entropy of class and variables |
| | Noise–signal ratio | Mean entropy of variables | |

A similar approach is taken in [27], in which simple, statistical and information theoretical characteristics are also considered, in addition to complexity measures (Table 8).

**Table 8:** Meta-features considered in [27]. Source: own elaboration

| Type (N) | Characteristics | | |
|---|---|---|---|
| Simple (5) | # of qualitative attributes | # of attributes | # of examples |
| | # of quantitative attributes | # of classes | |
| Statistical (6) | Avg. correlation between attributes | Avg. attributes asymmetry | Dataset balancing |
| | Avg. attribute variation coefficient | Avg. attributes kurtosis | Majority class error |
| Information theoretical (6) | Avg. mutual information between classes/attributes | Avg. conditional entropy between classes/attributes | Signal/noise ratio |
| | Avg. attributes entropy | Equivalent # of attributes | Class entropy |
| Complexity (4) | Fractal dimension of the dataset | Fisher's discriminant | |
| | Volume of the overlapping region | Dispersion of the data set | |

In [30], authors collected a total of 60 meta-features that were used in previous works. These meta-features were finally screened based on their number of citations in the literature, obtaining the final set displayed in Table 9.

**Table 9:** Meta-features considered in [30]. Source: own elaboration

| Type (N) | Characteristics | | |
|---|---|---|---|
| General (11) | # of classes | # of categorical features | # of lines with missing values |
| | # of samples | # of numerical features | % of lines with missing values |
| | # of features | # of missing values | # of data with noise |
| | Dimension | % of missing values | |
| Statistical (18) | Mean coefficient | Mean of mean | Mean skewness |
| | Max coefficient | Max of mean | Max skewness |
| | Min coefficient | Min of mean | Min skewness |
| | Mean covariance | Mean STD | Mean kurtosis |
| | Max covariance | Max STD | Max kurtosis |
| | Min covariance | Min STD | Min kurtosis |
| Information theoretical (5) | Class entropy | Min. attribute entropy | Mean mutual information |
| | Max attribute entropy | Mean attribute entropy | |

The above dataset characteristics are the meta-features of their respective meta-learning approaches. However, similar metrics are used in the ontology-based approach presented in [33], where authors employ a reduced set of simple, statistical and information theoretical values to characterize the datasets (Table 10).

**Table 10:** Meta-features considered in [33]. Source: own elaboration

| Type (N) | Characteristics | |
|---|---|---|
| Simple (3) | # of classes | # of instances |
| | # of features | |
| Statistical (2) | Avg. correlation of feature attributes | Avg. asymmetry of feature attributes |
| Information theoretical (3) | Class entropy | Equivalent # of attributes |
| | Signal-noise ratio | |

In the case of works that have captured heuristics from literature reviews, the data measurements vary. For example, in [28], authors define 12 parameters to guide the recommendation process, including: type of data, dimension of the dataset, outliers, feature noise, item noise, record noise, handling of overfitting, size of training data, size of the dataset, multicollinearity, based, and the shape of model. As can be observed, these parameters not only consider explicit measurements of the input dataset, but also other characteristics specific to the algorithms, such as their performance when preventing overfitting.

On the other hand, in [31], authors also used different criteria related to the requirements of the problem (in this case, classification problems) and the dataset structure (Table 11).

**Table 11:** Factors considered in [31]. Source: own elaboration

| Type (N) | Characteristics | | |
|---|---|---|---|
| Requirements (8) | Accuracy | Scalability | Amount of parameter tuning |
| | Performance/speed | Parallelization | Parametric/non-parametric |
| | Training | Extrapolation/interpolation | |
| Dataset (5) | # of features | Structure | Missing values |
| | Size | Multi-class labels | |

Also in this work, a flowchart for selecting a categorical encoding technique (before training the models) is presented [31]. This flowchart is based on three criteria: the dataset characteristics ("are the categories ordered?", "is the cardinality of the dataset less than 20?", etc.), the type of algorithm (tree-based or linear) and the required accuracy.

Regarding the remaining heuristic-based platform, Czako et al. also considered the problem type and requirements ("does the user want to predict a category?", "does the user want to predict a quantity?", etc.) when selecting or recommending an algorithm, in addition to some data-related attributes focused on the size of the dataset ("are there less than 100K examples?", "is the number of categories in the dataset known?", etc.) [32].

Finally, the remaining work [29] also reviewed other works to define the algorithm selection rules. Although the specific set of parameters is not detailed, they rely on a cheat sheet from the Sci-Kit Learn Library [47] to describe the examples of their Domain Specific Language (DSL), so both dataset characteristics and the task to be carried out are contemplated in the knowledge store.

***RQ4. How is the transparency of the selection process managed?***

This question is focused on how the selected papers convey the rules and heuristics for selecting the right AI algorithm. Conveying this information properly is crucial, because these rules are what make the algorithm selection process outcomes transparent and traceable. In general, the preferred method for summarizing this information is through matrices or decision trees.

Using a decision tree meta-learner when following a meta-learning approach eases the communication of the learned rules, because decision tree algorithms are explainable and readable. For example, in her work, Hilario conveys the results of the meta-learner (C5.0, a decision tree algorithm) through a table grouped by categories [25]. The same applies to [26], which also employed the C5.0 algorithm, although, in this case, the learned rules were detailed through pseudo-code.

A similar approach was employed in [30], in which an a priori algorithm [48] is employed to discover relationships among the different combination of recommended classifiers, along with a multi-label [49] decision tree (MEKA [50]) to identify interpretable and legible rules which also are described through pseudo-code.

In the remaining meta-learning approach, authors also selected a decision tree meta-learner (J48 algorithm) [27]. The resulting rules are directly visualized through three different decision trees: the first one to select the feature selection approach (Ranking or FSS), and the other two to select the specific feature selection model within the category recommended by the first decision tree.

Regarding the work presented in [33], authors used an ontology and SWRL rules to recommend the feature selection algorithms. The rule base is built by evaluating the different feature selection methods with symbolic, instance, statistical, and connectionist classifiers, and the feature selection algorithms are ranked by the accuracy of the model and the time required during the feature selection phase.

In [31], authors explained thoroughly the heuristics to select encoding algorithms and classification algorithms through flowcharts, tables, and pseudocode, while in [32] a decision tree is presented.

On the other hand, in one of the literature reviews [28], authors chose to explain the captured heuristics through matrices with different parameters. Depending on the problem to solve, users can check through the matrices which algorithms fit better their requirements and dataset characteristics.

Finally, the last work proposed a meta-modeling approach to model ML expert knowledge [29]. In this case, rules are visualized through node-link relationships in the form of decision trees to assist users in the selection of a proper ML algorithm.

***RQ5. How are the selection processes evaluated?***

The goal of answering this question is to check if the learned rules obtained from the algorithm selection approaches were validated and if so, through which methods.

Every work related to meta-learning involved validations of their outcomes. For example, in [25], the dataset coverage and confidence (correctly covered cases) values was provided along with the rules.

On the other hand, in [26], the quality of each rule was measured through support and confidence values after a 10-fold cross-validation. The support is defined as "number of dataset that match dataset conditions and best algorithm prediction/total number of datasets" and the confidence as "number of datasets that match best algorithm prediction/number of datasets that match dataset conditions".

Finally, in [27] authors measured their models' quality through their predictive performance average using a $10 \times 10$-fold cross-validation, while in [30] they compare the F-measure of their method with respect to other standard meta-learning approaches that focus on combining classifiers, outperforming them all.

Regarding the ontology-based approach [33], an experiment was carried out to test the approach, and also to compare the accuracy of the models with the recommended feature selection method and the non-recommended feature selection methods.

In the two heuristics-based approaches, a set of case studies were carried out to test the performance of the selected methods in two different domains: in the context of anomaly detection [32], and in the context of domestic fire injuries [31].

Although the remaining review works mentioned the implementation of the rules and knowledge bases into functional systems, they do not allude any testing [28,29].

### *6.1 Summary of the Results*

We found two main methodologies that tackle the problem of recommending AI algorithms: meta-learning and reviews/analyses of the literature and/or existing resources. While meta-learning is a powerful methodology, the selected meta-learning-based works only cover two main categories of

AI problems: feature selection and classification. However, review works shed light on more categories, such as regression, clustering, or dimensionality reduction (Fig. 6).

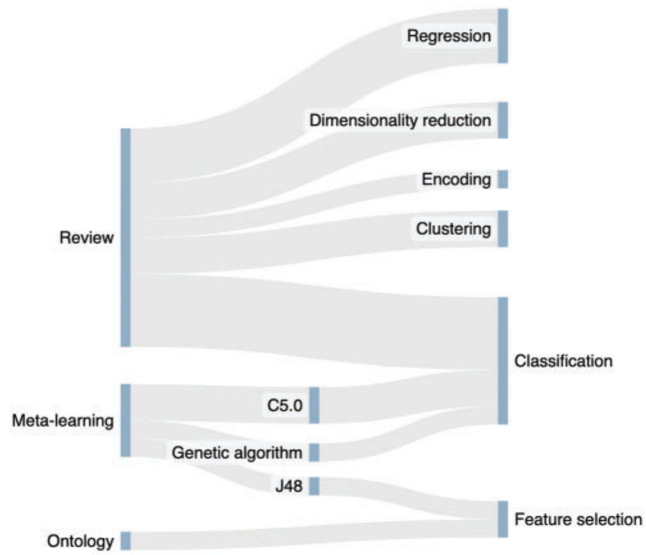


**Figure 6:** Summary of AI problems covered by the selected works. Most works focus on classification algorithms. Source: own elaboration

Regarding the factors that are considered to drive the recommendation or selection process, data characteristics are the preferred aspect. However, review works also consider other dimensions such as algorithm characteristics and tasks to build their recommendation criteria (Fig. 7).


**Figure 7:** Summary of aspects considered to drive the recommendation processes. In general, data characteristics are the preferred method to select an AI algorithm. Source: own elaboration

On the other hand, we found three types of methods to convey heuristics and rules among the selected works (Fig. 8). However, they rely on the same foundations: to detail the specific thresholds that make an algorithm the most suitable under certain conditions. Meta-learning approaches, in fact, benefit from the use of decision-tree algorithms, which are inherently explainable and human-readable [51].

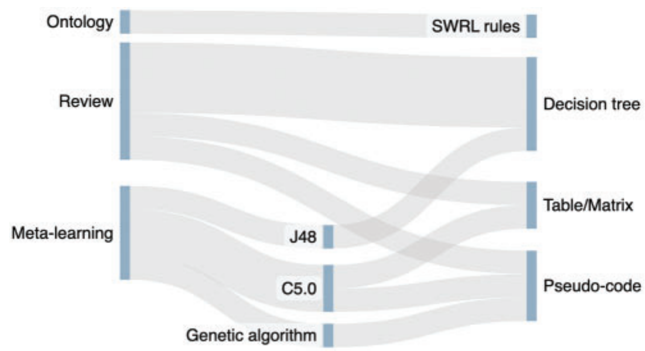


**Figure 8:** Summary of methods to convey the AI selection rules and heuristics. Source: own elaboration

Finally, in terms of evaluation, meta-learning solutions seem more robust referring to this matter, as the four selected works pertaining to this category tested the quality of their meta-learners' obtained rules. On the other hand, the ontology-based and heuristics-based approaches are also tested in terms of performance and functionality. The remaining review approaches, although they rely on validated works, don't contemplate any formal evaluation of their selected criteria (Fig. 9).

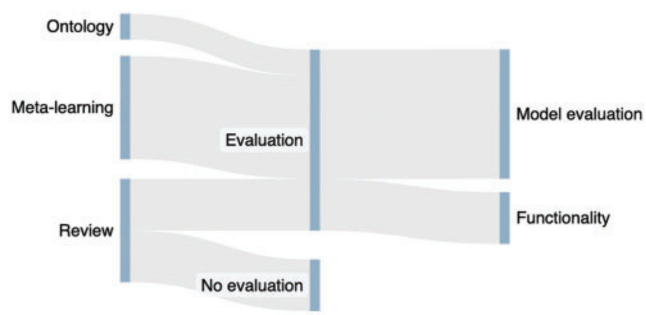


**Figure 9:** Summary of evaluation methods. The method used to evaluate the approaches only focus on the AI model evaluation. Source: own elaboration

## 7 Discussion

While several methods can be employed to recommend and select AI algorithms in different domains, we focused on those whose outcomes yielded explainable and straightforward rules to implement as a set of heuristics. This factor was a determiner when discarding works that didn't explain through specific thresholds the resulting selection or recommendation criteria.

The reviewed literature proposes very powerful methods to tackle the algorithm selection problem but mainly focuses on reporting performance, leaving the internal decisions of the meta-learners or the employed methodologies undisclosed (even if the methodologies allow explainability through decision trees or by other means).

The reduced number of selected papers (1.09% of the unique retrieved papers) is another outcome of the review itself. It implies that although several papers address the problem of recommending the best AI approach depending on the data, only a few of them report the internal rules followed to obtain the recommendation outputs.

This analysis was not focused on evaluating the performance or utility of these methods, as they are clearly powerful and useful for selecting suitable algorithms and are widely backed up by the literature. The focus of this research is on how the inherent knowledge obtained from meta-learning or other AI algorithm selection approaches has been transformed into readable and easy-to-follow rules that non-expert users can learn or apply to their own datasets.

In fact, the usefulness of having tangible heuristics is pointed out in some of the selected works. In [31], authors highlighted how previous works have noticed that using heuristics can lead to greater insights and engage users [52]. Also in [30], authors remarked that one of their main contributions is related to the detection of interpretable rules, because "it makes the achieved classification human-understandable".

Following the SLR results, we have seen that most strategies employed to select AI algorithms involve automatic methods such as meta-learning, which yields yet another AI model that predicts the most performant algorithm. In this case, reporting the underlying rules of explainable models could improve the transparency and traceability of the decisions and provide foundations to non-experts to understand how the AI algorithm selection takes place.

Not providing these heuristics may obscure the inspection, transferability, and adaptation of this knowledge to other contexts, in which the algorithm selection process might vary depending on the domain data. Fig. 10 visualizes the difference between a black-box approach and an explainable approach when using meta-learning to recommend proper AI algorithms.

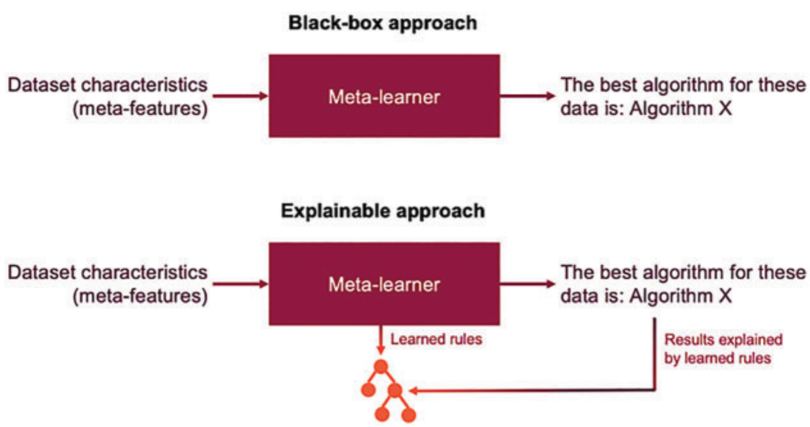


**Figure 10:** Black-box approach *vs*. explainable approach. Source: own elaboration

One of the retrieved works [29] pointed out the necessity of unifying this knowledge into information stores. These strategies could imply several benefits, from improving the traceability of the knowledge sources to the representation of expert knowledge through common language and graphical means, which could also derive a good way of learning how to select AI algorithms.

Relying on structured expert knowledge can also leverage the storage and application of different heuristics depending on the context, a previously mentioned challenge in this field.

The obtained results suggest that current AI selection research does not focus on explaining the internal rules of their approaches. Moreover, the selected works only cover a relatively small set of problems (mostly ML classification problems), which set open paths to tackle heuristics and rules for other AI fields such as DL, or other ML problems like clustering, regression, optimization, etc.

However, although only a small section of ML is tackled in the review works, there is a wide variety of algorithms to choose, not to mention those not covered, which makes the unification of heuristics and rules very complex. If these explicit rules are disseminated through research studies, it could be easier to gather them and build knowledge bases, as proposed in [29].

On the other hand, every selected work uses data characteristics as the determining factor for selecting a proper algorithm. This fact makes sense, as heuristics and rules must rely on explicit and measurable aspects of the problem space. As outlined in RQ3, using data characteristics as meta-features provide a powerful manner to understand both the dataset itself (which is crucial to interpret the outcomes of AI models) and the performance of AI algorithms over them.

We also found that the preferred method to convey the heuristics is to structure them through visual methods (like decision trees, tables, or cheat sheets) or pseudo-code. In fact, three works involving meta-learning employed a decision tree algorithm, increasing the explainability and readability of the results [51]. The selection of an explainable meta-learner is important for this matter because black-box algorithms could be powerful. Still, the inherent knowledge they hold cannot be transferred to lay users.

However, as outlined in RQ5, none of these works have tested their solutions with non-expert users. Although meta-learning approaches tested the quality of their models or the functionality of the methods, it is still a challenge to transfer expert knowledge through user-friendly platforms that not only assist lay users but also to teach them through readable rules how and when to use specific AI approaches.

### *7.1 Clarifications on the Excluded Records*

During the second phase of the SLR, a total of 109 works complied with the inclusion and exclusion criteria. However, the full-text assessment along with the quality criteria provided a more detailed view of the 109 works, which resulted in the selection of the 9 analyzed works in the present SLR.

The discarded works did not comply with the quality criteria for a series of different reasons. Some papers, for example, were more focused on hyperparameter optimization (out of the scope of this review) than in selecting an appropriate algorithm given the context of application [53–58].

On the other hand, other works focused on analyzing and creating frameworks for meta-features, without diving deep into the algorithm selection problem [59–63].

However, as discussed at the beginning of this section, the main reason for excluding works at this stage was related to the lack of explicit explanations regarding the rules or thresholds of the presented approaches for recommending AI algorithms (including meta-learning, decision frameworks, and similar methods) [64–70].

## 8 Threats to Validity

This type of review is prone to some limitations. One of these limitations is the bias that can be introduced during the data extraction. Quality criteria were employed to reduce the effects of bias in the inclusion phase of the SLR. All authors were involved in the review planning to identify and avoid any early issues regarding the study design. In this case, the first author was the lead reviewer, while the remaining reproduced each phase to check the validity of the results. On the other hand, data related to the entire process is provided and available to make the SLR reproducible.

Another limitation is that it is not guaranteed that every relevant work related to AI algorithm is retrieved. To mitigate this issue, we selected two relevant electronic databases in the field of computer science. The exclusion of Google Scholar from this review is justified by the necessity of considering only databases that index quality contrasted contents.

Finally, AI algorithm recommendation not only covers the selection of a certain model, as these models have parameters that also need to be tuned and require more specific expert knowledge. This challenge was out of the scope of this review, but it is important to bear in mind this matter, as the performance of a model can be improved by properly configuring its hyperparameters.

## 9 Conclusions

A systematic literature review (SLR) and systematic literature mapping have been carried out to analyze the state-of-the-art of AI algorithm selection in terms of its transparency: how the underlying expert knowledge has been materialized through heuristics or rules in the literature. The present review addresses relevant aspects of the solutions, including the factors that drive the selection process, the target algorithms, or the methods to convey the underlying rules obtained from the selection process.

1346 papers were retrieved from two electronic databases. The number of papers was reduced to nine after applying inclusion and exclusion criteria and a quality assessment to keep only relevant works for the scope of the research. The reduced number of retrieved papers suggests a lack of reporting explicit rules and heuristics when testing the suitability and performance of AI algorithms.

The systematic literature review and mapping provide a summary of existing approaches and works that tackle the automatic selection of AI models transparently. Also, it encourages researchers to explicitly report the results of their AI recommendation methodologies.

Future work will involve the application of the gained knowledge to implement platforms [71–75] to teach and assist non-expert users in learning how to properly apply AI to their domain problems, and to manage and explain heuristics in a more comprehensible manner [76].

**Funding Statement:** This research was partially funded by the Spanish Government Ministry of Economy and Competitiveness through the DEFINES Project Grant No. (TIN2016-80172-R) and the Ministry of Science and Innovation through the AVisSA Project Grant No. (PID2020-118345RB-I00). This research work has been supported by the Spanish *Ministry of Education and Vocational Training* under an FPU Fellowship (FPU17/03276).

**Conflicts of Interest:** The authors declare that they have no conflicts of interest to report regarding the present study.

## References

1. Weyerer, J. C., Langer, P. F. (2019). Garbage in, garbage out: The vicious cycle of AI-based discrimination in the public sector. *Proceedings of the 20th Annual International Conference on Digital Government Research*, pp. 509–511. Dubai, United Arab Emirates.
2. Ferrer, X., van Nuenen, T., Such, J. M., Coté, M., Criado, N. (2021). Bias and discrimination in AI: A cross-disciplinary perspective. *IEEE Technology and Society Magazine, 40(2),* 72–80. DOI 10.1109/MTS.2021.3056293.
3. Hoffman, S. (2021). The emerging hazard of AI-related health care discrimination. *Hastings Center Report, 51(1),* 8–9. DOI 10.1002/hast.1203.

4. Wachter, S., Mittelstadt, B., Russell, C. (2021). Why fairness cannot be automated: Bridging the gap between EU non-discrimination law and AI. *Computer Law & Security Review, 41,* 105567. DOI 10.1016/j.clsr.2021.105567.
5. Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F. et al. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis, 42(13),* 60–88. DOI 10.1016/j.media.2017.07.005.
6. González Izard, S., Sánchez Torres, R., Alonso Plaza, Ó., Juanes Méndez, J. A., García-Peñalvo, F. J. (2020). Nextmed: Automatic imaging segmentation, 3D reconstruction, and 3D model visualization platform using augmented and virtual reality. *Sensors, 20(10),* 2962. DOI 10.3390/s20102962.
7. Izard, S. G., Juanes, J. A., García Peñalvo, F. J., Estella, J. M. G., Ledesma, M. J. S. et al. (2018). Virtual reality as an educational and training tool for medicine. *Journal of Medical Systems, 42(3),* 50. DOI 10.1007/s10916-018-0900-2.
8. Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A. et al. (2016). TensorFlow: A system for large-scale machine learning. *12th USENIX Symposium on Operating Systems Design and Implementation*, pp. 265–283. Savannah, GA.
9. Anil, R., Capan, G., Drost-Fromm, I., Dunning, T., Friedman, E. et al. (2020). Apache mahout: Machine learning on distributed dataflow systems. *Journal of Machine Learning Research, 21(127),* 1–6.
10. Bisong, E. (2019). *Google colaboratory, in building machine learning and deep learning models on google cloud platform: A comprehensive guide for beginners*. Berkeley, CA: Apress.
11. Frank, E., Hall, M., Holmes, G., Kirkby, R., Pfahringer, B. et al. (2009). *Weka–A machine learning workbench for data mining, in data mining and knowledge discovery handbook*. Boston, MA: Springer.
12. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P. et al. (2009). The WEKA data mining software: An update. *ACM SIGKDD Explorations Newsletter, 11(1),* 10–18. DOI 10.1145/1656274.1656278.
13. Bjaoui, M., Sakly, H., Said, M., Kraiem, N., Bouhlel, M. S. (2020). Depth insight for data scientist with RapidMiner «an innovative tool for AI and big data towards medical applications». *Proceedings of the 2nd International Conference on Digital Tools & Uses Congress*, Virtual Event, Tunisia.
14. Berthold, M. R., Cebron, N., Dill, F., Gabriel, T. R., Kötter, T. et al. (2009). KNIME—the Konstanz information miner: Version 2.0 and beyond. *SIGKDD Explorations Newsletter, 11(1),* 26–31. DOI 10.1145/1656274.1656280.
15. García-Peñalvo, F. J. (2022). Developing robust state-of-the-art reports: Systematic literature reviews. *Education in the Knowledge Society, 23,* e28600.
16. Kitchenham, B. (2004). Procedures for performing systematic reviews. *Keele University Technical Report TR/SE-0401 and NICTA Technical Report 0400011T.1*.
17. Kitchenham, B., Charters, S. (2007). *Guidelines for performing systematic literature reviews in software engineering. Version 2.3*. School of Computer Science and Mathematics, Keele University and Department of Computer Science, Durham, UK.
18. García-Holgado, A., Marcos-Pablos, S., García-Peñalvo, F. J. (2020). Guidelines for performing systematic research projects reviews. *International Journal of Interactive Multimedia and Artificial Intelligence, 6(2),* 136–144.
19. Kitchenham, B., Budgen, D., Brereton, O. P. (2011). Using mapping studies as the basis for further research-a participant-observer case study. *Information Software Technology, 53(6),* 638–651. DOI 10.1016/j.infsof.2010.12.011.
20. Napoleão, B., Felizardo, K. R., de Souza, É. F., Vijaykumar, N. L. (2017). Practical similarities and differences between systematic literature reviews and systematic mappings: A tertiary study. *29th International Conference on Software Engineering and Knowledge Engineering (SEKE 2017)*, pp. 1–6. Pittsburgh, CA, USA.
21. Petticrew, M., Roberts, H. (2008). *Systematic reviews in the social sciences: A practical guide*. Hoboken, NJ, USA: John Wiley & Sons.

22. Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *The Prisma Group, 6(7),* e1000097.
23. Page, M. J., Moher, D., Bossuyt, P. M., Boutron, I., Hoffmann, T. C. et al. (2021). PRISMA, 2020 explanation and elaboration: Updated guidance and exemplars for reporting systematic reviews. *BMJ, 372,* n160. DOI 10.1136/bmj.n160.
24. Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C. et al. (2021). The PRISMA, 2020 statement: An updated guideline for reporting systematic reviews. *BMJ, 372,* n71.
25. Hilario, M. (2002). Model complexity and algorithm selection in classification. *Proceedings of the 5th International Conference on Discovery Science*, pp. 113–126. Lubeck, Germany.
26. Ali, S., Smith, K. A. (2006). On learning algorithm selection for classification. *Applied Soft Computing, 6(2),* 119–138. DOI 10.1016/j.asoc.2004.12.002.
27. Parmezan, A. R. S., Lee, H. D., Wu, F. C. (2017). Metalearning for choosing feature selection algorithms in data mining: Proposal of a new framework. *Expert Systems with Applications, 75(10),* 1–24. DOI 10.1016/j.eswa.2017.01.013.
28. Kumar, C., Käppel, M., Schützenmeier, N., Eisenhuth, P., Jablonski, S. (2019). A comparative study for the selection of machine learning algorithms based on descriptive parameters. *Proceedings of the 8th International Conference on Data Science, Technology and Applications*, vol. 1, pp. 408–415. Prague, Czech Republic.
29. Martínez-Rojas, A., Jiménez-Ramírez, A., Enríquez, J. (2019). Towards a unified model representation of machine learning knowledge. *Proceedings of the 15th International Conference on Web Information Systems and Technologies*, pp. 470–476. Vienna, Austria.
30. Golshanrad, P., Rahmani, H., Karimian, B., Karimkhani, F., Weiss, G. (2021). MEGA: Predicting the best classifier combination using meta-learning and a genetic algorithm. *Intelligent Data Analysis, 25(6),* 1547–1563. DOI 10.3233/IDA-205494.
31. Reilly, D., Taylor, M., Fergus, P., Chalmers, C., Thompson, S. (2022). The categorical data conundrum: Heuristics for classification problems—A case study on domestic fire injuries. *IEEE Access, 10,* 70113–70125. DOI 10.1109/ACCESS.2022.3187287.
32. Czako, Z., Sebestyen, G., Hangan, A. (2018). Evaluation platform for artificial intelligence algorithms. *Proceedings of the 10th International Joint Conference on Computational Intelligence*, pp. 39–46. Seville, Spain.
33. Nayak, A., Božić, B., Longo, L. (2022). An ontological approach for recommending a feature selection algorithm. *ICWE 2022: Web Engineering*, pp. 300–314. Bari, Italy.
34. Hospedales, T., Antoniou, A., Micaelli, P., Storkey, A. (2021). Meta-learning in neural networks: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(9),* 5149–5169. DOI 10.1109/TPAMI.2021.3079209.
35. Wang, J. X. (2021). Meta-learning in natural and artificial intelligence. *Current Opinion in Behavioral Sciences, 38,* 90–95. DOI 10.1016/j.cobeha.2021.01.002.
36. Vanschoren, J. (2019). *Meta-learning, in automated machine learning*. Cham: Springer.
37. Vanschoren, J. (2018). Meta-learning: A survey. arXiv preprint arXiv:1810.03548.
38. Smith-Miles, K. A. (2009). Cross-disciplinary perspectives on meta-learning for algorithm selection. *ACM Computing Surveys, 41(1),* 1–25. DOI 10.1145/1456650.1456656.
39. Schweighofer, N., Doya, K. (2003). Meta-learning in reinforcement learning. *Neural Networks, 16(1),* 5–9. DOI 10.1016/S0893-6080(02)00228-9.
40. Vilalta, R., Drissi, Y. (2002). A perspective view and survey of meta-learning. *Artificial Intelligence Review, 18(2),* 77–95. DOI 10.1023/A:1019956318069.

41. Rajeswari, S., Suthendran, K. (2019). C5.0: Advanced decision tree (ADT) classification model for agricultural data analysis on cloud. *Computers and Electronics in Agriculture, 156(2),* 530–539. DOI 10.1016/j.compag.2018.12.013.
42. Pandya, R., Pandya, J. (2015). C5. 0 algorithm to improved decision tree with feature selection and reduced error pruning. *International Journal of Computer Applications, 117(16),* 18–21. DOI 10.5120/20639-3318.
43. Bujlow, T., Riaz, T., Pedersen, J. M. (2012). A method for classification of network traffic based on C5.0 machine learning algorithm. *2012 International Conference on Computing, Networking and Communications (ICNC)*, pp. 237–241. Hawaii, USA.
44. Sahu, S., Mehtre, B. M. (2015). Network intrusion detection system using J48 decision tree. *2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pp. 2023–2026. Kerala, India.
45. Kaur, G., Chhabra, A. (2014). Improved J48 classification algorithm for the prediction of diabetes. *International Journal of Computer Applications, 98(22),* 13–17. DOI 10.5120/17314-7433.
46. Bhargava, N., Sharma, G., Bhargava, R., Mathuria, M. (2013). Decision tree analysis on j48 algorithm for data mining. *Proceedings of International Journal of Advanced Research in Computer Science and Software Engineering, 3(6),* 1114–1119.
47. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B. et al. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research, 12,* 2825–2830.
48. Kotsiantis, S. B., Kanellopoulos, D. N. (2006). Association rules mining: A recent overview. *GESTS International Transactions on Computer Science and Engineering, 32(1),* 71–82.
49. Zhang, M. L., Zhou, Z. H. (2014). A review on multi-label learning algorithms. *IEEE Transactions on Knowledge and Data Engineering, 26(8),* 1819–1837. DOI 10.1109/TKDE.2013.39.
50. Read, J., Reutemann, P., Pfahringer, B., Holmes, G. (2016). Meka: A multi-label/multi-target extension to weka. *Journal of Machine Learning Research, 17(21),* 1–5.
51. Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S. et al. (2020). Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion, 58,* 82–115.
52. Lu, Z., Yin, M. (2021). Human reliance on machine learning models when performance feedback is limited: Heuristics and risks. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. Yokohama, Japan.
53. Muravyov, S. B., Efimova, V. A., Shalamov, V. V., Filchenkov, A. A., Smetannikov, I. B. (2019). Automatic hyperparameter optimization for clustering algorithms with reinforcement learning. *Nauchno-Tekhnicheskii Vestnik Informatsionnykh Tekhnologii, Mekhaniki i Optiki, 19(3),* 508–515.
54. Cachada, M., Abdulrhaman, S. M., Brazdil, P. (2017). Combining feature and algorithm hyperparameter selection using some metalearning methods. *Automatic Machine Learning (AutoML) 2017*, pp. 69–83. Skopje, Macedonia.
55. Marques, R. Z. N., Coutinho, L. R., Borchartt, T. B., Vale, S. B., Silva, F. J. S. (2015). An experimental evaluation of data mining algorithms using hyperparameter Optimization. *2015 Fourteenth Mexican International Conference on Artificial Intelligence (MICAI)*, pp. 152–156. Cuernavaca, Mexico.
56. Donoghue, J. O., Roantree, M. (2015). A framework for selecting deep learning hyper-parameters. *British International Conference on Databases (BICOD 2015)*, pp. 120–132. Edinburgh, UK.
57. Feurer, M., Springenberg, J., Hutter, F. (2015). Initializing bayesian hyperparameter optimization via meta-learning. *Proceedings of the AAAI Conference on Artificial Intelligence, 29(1),* 1128–1135. DOI 10.1609/aaai.v29i1.9354.
58. Guo, X., van Stein, B., Bäck, T. (2019). A new approach towards the combined algorithm selection and hyper-parameter optimization problem. *2019 IEEE Symposium Series on Computational Intelligence (SSCI)*, pp. 2042–2049, Xiamen, China.

59. Pinto, F., Soares, C., Moreira, J. (2014). A framework to decompose and develop metafeatures. *Proceedings of the International Workshop on Meta-Learning and Algorithm Selection Co-Located with 21st European Conference on Artificial Intelligence (ECAI 2014)*, pp. 32–36. Prague, Czech Republic.
60. Ler, D., Teng, H., He, Y., Gidijala, R. (2018). Algorithm selection for classification problems via cluster-based meta-features. *2018 IEEE International Conference on Big Data (Big Data)*, pp. 4952–4960. Seattle, WA, USA.
61. Shah, R., Khemani, V., Azarian, M., Pecht, M., Su, Y. (2018). Analyzing data complexity using metafeatures for classification algorithm selection. *2018 Prognostics and System Health Management Conference (PHM-Chongqing)*, pp. 1280–1284. Chongqing, China.
62. Rivolli, A., Garcia, L. P. F., Soares, C., Vanschoren, J., de Carvalho, A. C. P. L. F. (2022). Meta-features for meta-learning. *Knowledge-Based Systems, 240(2),* 108101. DOI 10.1016/j.knosys.2021.108101.
63. Pimentel, B. A., de Carvalho, A. C. P. L. F. (2018). Statistical versus distance-based meta-features for clustering algorithm recommendation using meta-learning. *2018 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8. Brazil: Rio de Janeiro.
64. Ali, R., Lee, S., Chung, T. C. (2017). Accurate multi-criteria decision making methodology for recommending machine learning algorithm. *Expert Systems with Applications, 71(5),* 257–278. DOI 10.1016/j.eswa.2016.11.034.
65. Luo, G. (2016). A review of automatic selection methods for machine learning algorithms and hyper-parameter values. *Network Modeling Analysis in Health Informatics and Bioinformatics, 5(1),* 18. DOI 10.1007/s13721-016-0125-6.
66. Hilario, M., Kalousis, A., Nguyen, P., Woznica, A. (2009). A data mining ontology for algorithm selection and meta-mining. *European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases, Third Generation Data Mining: Towards Service-Oriented Knowledge Discovery, SoKD 2009*, pp. 76–87. Bled, Slovenia.
67. Dyrmishi, S., Elshawi, R., Sakr, S. (2019). A decision support framework for AutoML systems: A meta-learning approach. *2019 International Conference on Data Mining Workshops (ICDMW)*, pp. 97–106. Beijing, China.
68. Sala, R., Pirola, F., Pezzotta, G., Corona, M. (2021). The machine learning algorithm selection model: Test with multiple datasets. *Proceedings of the Summer School Francesco Turco*. Bergamo, Italy, University of Bergamo, Virtual Event.
69. Taratukhin, O., Muravyov, S. (2021). Meta-learning based feature selection for clustering. *Intelligent Data Engineering and Automated Learning—IDEAL 2021*, pp. 548–559. Manchester, UK.
70. Patil, P. S., Kappuram, K., Rumao, R., Bari, P. (2022). Development of AMES: Automated ML expert system. *2022 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COM-IT-CON)*, pp. 208–213. Faridabad, India.
71. Vázquez-Ingelmo, A., Alonso-Sánchez, J., García-Holgado, A., García-Peñalvo, F. J., Sampedro-Gómez, J. et al. (2021). Bringing machine learning closer to non-experts: Proposal of a user-friendly machine learning tool in the healthcare domain. *Ninth International Conference on Technological Ecosystems for Enhancing Multiculturality (TEEM'21)*, pp. 324–329. Barcelona, Spain.
72. García-Peñalvo, F. J., Vázquez-Ingelmo, A., García-Holgado, A., Sampedro-Gómez, J., Sánchez-Puente, A. et al. (2021). Application of artificial intelligence algorithms within the medical context for non-specialized users: The CARTIER-IA platform. *International Journal of Interactive Multimedia & Artificial Intelligence, 6(6),* 46–53.
73. García-Holgado, A., Vázquez-Ingelmo, A., Alonso-Sánchez, J., García-Peñalvo, F. J., Therón, R. et al. (2021). User-centered design approach for a machine learning platform for medical purpose. In: *Human-computer interaction*, pp. 237–249. Sao Paulo, Brazil.
74. Vázquez-Ingelmo, A., Alonso, J., García-Holgado, A., García-Peñalvo, F. J., Sampedro-Gómez, J. et al. (2021). Usability study of CARTIER-IA: A platform for medical data and imaging management. In:

*Learning and collaboration technologies: New Challenges and learning experiences*, pp. 374–384. Washington DC, USA, Virtual event.

75. Vázquez-Ingelmo, A., Sampedro-Gómez, J., Sánchez-Puente, A., Vicente-Palacios, V., Dorado-Díaz, P. I. et al. (2020). A platform for management and visualization of medical data and medical imaging. *Eighth International Conference on Technological Ecosystems for Enhancing Multiculturality*, pp. 518–522, Salamanca, Spain.
76. Vázquez-Ingelmo, A., García-Holgado, A., García-Peñalvo, F. J., Therón, R. (2022). Proof-of-concept of an information visualization classification approach based on their fine-grained features. *Expert Systems,* 1–16. DOI 10.1111/exsy.12872.