Safety Science 138 (2021) 105238

Contents lists available at ScienceDirect

Safety Science

journal homepage: www.elsevier.com/locate/safety

ProMetaUS: A proactive meta-learning uncertainty-based framework to select models for Dynamic Risk Management

Elena Stefana [a], Nicola Paltrinieri [b,*]

[a] Department of Mechanical and Industrial Engineering, University of Brescia, via Branze 38, 25123 Brescia, Italy
[b] Department of Mechanical and Industrial Engineering, Faculty of Engineering, Norwegian University of Science and Technology NTNU, S.P. Andersens vei 5, 7491 Trondheim, Norway

ARTICLE INFO

Keywords:
Algorithm ranking
Model recommendation
Risk assessment
Uncertainty quantification
Probability
Machine learning

ABSTRACT

Safety managers, practitioners, and researchers can employ different models for estimating and assessing hazards, consequences, likelihoods, risks, and/or mitigation measures in the safety field. The selection of a specific model may depend on the uncertainty associated with its estimation and its impact on the safety-related decision-making process. The recognition of this issue as an example of Algorithm Selection Problem (ASP) allows investigating the applicability of meta-learning principles that are scarcely adopted in the risk and safety literature. Consequently, we propose a novel meta-learning inspired framework to proactively rank a set of candidate models for Dynamic Risk Management (DRM) based on desired uncertainty conditions. We denominate this framework ProMetaUS (Proactive Meta-learning and Uncertainty-based Selection for dynamic risk management). To achieve this purpose, our meta-learning system acquires knowledge that relates the characteristics extracted both directly and indirectly from datasets (e.g. data-based, domain-based, simple and fast uncertainty-based, simple and fast sensitivity-based meta-features) to some performance measures of the models. Performance measures include confidence information, shape measurable quantities, safety decision criteria and threshold limits, and sensitivity analysis outputs. We tested the proposed framework in a case study about Oxygen Deficiency Hazard (ODH) assessment by means of @RISK. For each of the five datasets, single-performance measure rankings and a final ranking of the three models are generated. Such rankings are aggregated to obtain the global recommended ranking.

1. Introduction

In the safety field, several models able to estimate and assess hazards, consequences, likelihoods, risks, and/or mitigation measures have been developed. A model is a simplified representation of the real system (Cullen and Frey, 1999; Nilsen and Aven, 2003), which reflects the causal relations that produce the events focused on by the decision-makers (Nilsen and Aven, 2003). The complexity of a model is governed by several factors, e.g. the complexity of the system, existing knowledge about the system, amount of information the decision-makers consider a sufficient basis for making the decision in question, and available resources (Nilsen and Aven, 2003). In going from simple to more complex models, the uncertainty due to the model structure may be reduced, but the uncertainty due to the larger number of inputs tends to increase (Cullen and Frey, 1999). According to Chen and Ma (2007), for a model to be helpful, the associated uncertainty should be limited. Uncertainty is the lack of knowledge about the true value of a quantity, regarding which of several alternative model representations best describes a mechanism of interest, or about which of several alternative probability density functions should represent a quantity of interest (Cullen and Frey, 1999; IPCS, 2008). In addition, uncertainty analysis is an essential component and integral part of hazard and risk management, and risk assessment to quantify the degree of confidence in the estimate of risk (IAEA, 1989; Thompson and Warmink, 2017). The relevance of uncertainty during risk assessment is also recognised by Arunraj et al. (2013), according to which modelling uncertainty is a vital component for effective decision-making.

In such uncertain context, safety managers, practitioners, and researchers may wonder which model(s) should be selected for making decisions objectively about the hazard, consequence, likelihood, risk, and/or mitigation measure under analysis. Since there is no a single algorithm that performs better than the others for all the possible

* Corresponding author.
E-mail addresses: elena.stefana@unibs.it (E. Stefana), nicola.paltrinieri@ntnu.no (N. Paltrinieri).

https://doi.org/10.1016/j.ssci.2021.105238
Received 21 October 2020; Received in revised form 12 January 2021; Accepted 22 February 2021
Available online 8 March 2021

problems, scenarios, and/or datasets, as stated by No Free Lunch (NFL) Theorem (Wolpert, 1996; Wolpert and Macready, 1995, 1997), suitable models have to be selected for each problem (Kück et al., 2016). This type of issue is known in the literature as Algorithm Selection Problem (ASP), which is originally described by Rice (1976). ASP endeavours to select and apply the best algorithm(s) for a given task (Abdulrahman et al., 2015, 2018), and is composed by: (1) problem space, (2) feature space, (3) algorithm space, and (4) performance space (Rice, 1976).

One effective approach to solve ASP and thus recommend the most adequate algorithm for a certain task and for a new dataset, and automatically provide guidance on the best alternatives is provided by meta-learning (also known as "meta-level learning", "learning to learn", or "learning about learning") (Abdulrahman and Brazdil, 2014; Bhatt et al., 2012; Brazdil et al., 2009, 2017; Cohen-Shapira et al., 2019; Cunha et al., 2018; de Souto et al., 2008; Filchenkov and Pendryak, 2015; Khan et al., 2020; Kozielski and Łaskarzewski, 2019; Kück et al., 2016; Makmal et al., 2016; Muñoz et al., 2013; Pimentel and de Carvalho, 2019a, 2019b; Pinto et al., 2014, 2016; Pise and Kulkarni, 2016; Prudêncio and Ludermir, 2004, 2012; Prudêncio et al., 2011a, 2011c, 2011d; Reif et al., 2012; Ren et al., 2020; Romero et al., 2013; Rossi et al., 2014, 2017; Santos et al., 2012; Shahoud et al., 2020; Smith-Miles, 2008a; Soares et al., 2009; Sousa et al., 2016; Vanschoren, 2018, 2019; Vilalta et al., 2004, 2009; Zorrilla and García-Saiz, 2014). An overview of main meta-learning definitions can be found in Stefana and Paltrinieri (2020).

Meta-learning can be viewed as an important feature of self-adaptive systems (Brazdil et al., 2017), an understanding and adaptation of learning itself (Lemke et al., 2015; Ren et al., 2020). It is concerned with understanding the learning mechanism, and the process of exploiting and learning from experience; this previous knowledge is gained during the application of various learning algorithms on different kinds of data to offer an automatic selection, recommendation, or support for a future task (Brazdil et al., 2009; Dyrmishi et al., 2019; Kanda et al., 2016; Pise and Kulkarni, 2016; Reif et al., 2012; Rossi et al., 2012, 2017; Smith-Miles, 2008a). Meta-learning systems assist (non-expert) users in the process of algorithm selection by mapping a particular task to a suitable model (or combination of models) and by acquiring knowledge from the application of a set of algorithms on different problems (Brazdil et al., 2009; de Souto et al., 2008; Giraud-Carrier et al., 2004; Lemke et al., 2015; Pise and Kulkarni, 2016; Prudêncio and Ludermir, 2004; Rossi et al., 2014; Santos et al., 2012; Soares et al., 2009; Vilalta and Drissi, 2002; Vilalta et al., 2004, 2009; Zorrilla and García-Saiz, 2014, 2015). Specifically, meta-learning aims to predict an algorithm or a set of algorithms suitable for a specific problem under study by learning the relationship between the meta-features extracted from the datasets and the algorithms performance applied on them (Bhatt et al., 2012; Brazdil et al., 2009; Cohen-Shapira et al., 2019; de Souto et al., 2008; Lemke et al., 2015; Prudêncio and Ludermir, 2012; Prudêncio et al., 2011c; Ren et al., 2020; Smith-Miles, 2008b; Zhu et al., 2018; Zorrilla and García-Saiz, 2015). In the literature, several proposals on algorithm selection via meta-learning in different domains can be found: some examples are reported in Table 1.

However, to the best of our knowledge, there is limited evidence about the use of meta-learning potential in the risk and safety literature: Kozielski (2016) gives a description of a meta-learning approach for predicting methane concentration in a coal mine, Paltrinieri et al. (2020) attempt to generalise and model the risk analysis learning process by considering the case study of a drive-off scenario involving an oil and gas drilling rig, for which a risk assessment approach based on machine learning is developed, Stefana and Paltrinieri (2020) recently describe the introductory aspects of a preliminary meta-learning framework for ranking models estimating a safety risk in uncertain conditions, and Brocal et al. (2021) qualitatively address the applicability of meta-learning lessons for the selection of strategies for emerging risk management assuming uncertainty as the main decision variable in industrial context. The objective of this paper is to define a novel proactive framework based on meta-learning concepts and ASP, ultimately enabling effective Dynamic Risk Management (DRM). It will support safety managers, practitioners, and researchers, despite not being experts in all the existing models and data mining techniques, in the selection of models for the assessment of DRM core elements (hazard, consequence, likelihood, risk, or mitigation measures) based on desired uncertainty conditions. Indeed, since each model is characterised by different types and levels of uncertainty, also depending on the problem under study, analysts should establish which factors are particularly relevant for a specific assessment: e.g. for the problem under study, is it better to use a model with a narrower or more conservative confidence interval / is a model minimising the tail to the right or the tail to the left preferable?

Table 1
Examples of proposals on algorithm selection via meta-learning in different domains.

Author(s) (Year)	Brief description of proposal
Prudêncio et al. (2011a)	Ranking meta-learning approaches in time series forecasting and clustering of gene expression data
Rossi et al. (2012)	Meta-learning approach for periodic algorithm selection in time-changing environments where data flow continuously
Cui et al. (2016b)	Building Energy Model Recommendation system for short term load forecasting
Kück et al. (2016)	Meta-learning approach to select time series forecasting models
Kozielski and Łaskarzewski (2019)	Automated approach to Liquefied Petroleum Gas consumption prediction in a short term horizon
Shahoud et al. (2020)	Methodology for characterising the behaviour of time series datasets with meta features to achieve a more accurate model selection for time series energy load forecasting

To our knowledge, this is the first framework based on meta-learning concepts to select models for estimating and assessing DRM core elements that considers uncertainty in the safety field. It is a flexible and dynamic tool that permits continuously incorporating new models, and additional evidence and information when available, which represents itself a key aspect for reducing uncertainty. The paper is partly an extension of Stefana and Paltrinieri (2020). Readers interested in DRM process can refer to Bucelli et al. (2020), Paltrinieri et al. (2019), and Paltrinieri and Khan (2016).

The framework is named ProMetaUS (Proactive Meta-learning and Uncertainty-based Selection for dynamic risk management) after the Greek mythological figure Prometheus (signifying "forethought"). His myth represents the pursuit of knowledge, as he fixed Epimetheus's mistake (signifying "afterthought"). While assigning positive traits to every animal, Epimetheus ran out of options when it came to humans, but Prometheus's effort ultimately allowed man to acquire skills and technology.

The remainder of this paper is organised as follows. Section 2 specifies details about the methods employed for the research, while Section 3 summarises the cornerstones of meta-learning architecture for ASP. The ProMetaUS framework is presented in Section 4. Its application to a case study related to Oxygen Deficiency Hazard (ODH) assessments and the main results are described in Section 5 and Section 6, respectively. Discussion about the case study results, the main properties and limitations of the proposed framework is stated in Section 7, while concluding remarks are provided in the final section.

2. Methods

To achieve our objective, we implemented and followed the strategy reported in Fig. 1.

The first step of our strategy regarded an examination of the literature about meta-learning and ASP concepts and applications. We searched for scientific publications by means of various combinations of

Fig. 1. Strategy for designing the ProMetaUS framework (Abbreviations used: ASP = Algorithm Selection Problem; DRM = Dynamic Risk Management).

the following keywords: (1) meta-learning, and (2) algorithm or model combined with recommend*, rank*, choice/choose, compar*, select*, evaluat*. Note that such groups of keywords are merged in different search strings by means of Boolean operators and focusing on title, abstract, and keywords fields in several electronic (bibliographic) databases of scientific publications (e.g. Scopus, Science Direct, and Web of Science). To consider all potentially relevant papers, a starting date was not established. We took into account different document types written in English, such as articles, conference papers, reviews, and book chapters. We included studies based on the relevance of titles, abstracts, and full-texts. The list of references in each study was checked through a manual examination for identifying any additional relevant articles. Studies explained general notions and structures of meta-learning were also considered in results. We examined in depth papers proposing meta-learning architectures and approaches for ASP and presenting real case studies about their applications. Papers detailing a specific constituent part of a meta-learning system for ASP (e.g. meta-features) were also included. Documents solely on particular meta-learning approaches as stacked generalisation or boosting were excluded. The relevant documents obtained by means of this literature review permitted us to outline the principal components and steps of a typical meta-learning process for ASP, which are summarised in the following section.

In order to define a framework as generable as possible for models estimating and assessing DRM core elements, we consulted a wide variety of literature sources on models. This step had the purpose to identify their typical structures, common inputs and outputs, and assumptions. Because one of key aspect of the ProMetaUS framework regards the description and treatment of uncertainty, we also investigated the different types of uncertainty sources involved in the assessments, such as parameter and model uncertainty. Such investigation was followed by a deep review for analysing the available literature about techniques for mathematical model evaluation and comparison, uncertainty and sensitivity definitions and methods, and uncertainty management and treatment. We defined several queries through the following keyword categories: (1) uncertainty and/or sensitivity and/or probabil* in combination with estimate*, measur*, indicator*, metric*, performance, rank*, quantif*, statistic*, analys*, (2) safety, hazard*, or risk* combined with assess*, manage*, or analys*. To ensure to capture as many significant documents as possible, we paid particular attention on the identification of all the possible synonyms that can be utilised to express the two categories. The search queries were used to interrogate the electronic databases of scientific publications, and were applied to all the fields available for each database (i.e. title, abstract, keywords, and full-text). In addition to scientific publications, we also analysed English-language technical reports, books, and specialised guidelines. We rated their relevance by reading the full-text. The list of references in each study was checked through a manual examination to identify any additional relevant documents. For the uncertainty and sensitivity topics, we focused on the safety field, but we also dedicated our attention on other domains, such as industrial practice (e.g. de Rocquigny et al., 2008), environmental modelling, assessments, and decision-making (e.g. Loucks et al., 2005), risk analysis and assessment for decision-making (e.g. Zio and Aven, 2013), atmospheric and dispersion modelling (e.g. Chang and Hanna, 2004). We were particularly interested in publications dealing with the probabilistic approach and epistemic uncertainty, expressed in the form of probability distribution. Documents were excluded if exclusively focused on Bayesian approaches, evidence theory and fuzzy approach, stochastic response surface or bootstrap methods. Also details of advanced mathematical methods and techniques related to sensitivity analysis were neglected.

The information and knowledge gathered by means of these literature reviews helped us to compare the principal ASP elements with inputs and outputs of uncertainty and sensitivity analyses of models. By means of several brainstorming sessions between the authors, we obtained a preliminary and shared definition of meta-features and performance measures, we supposed a reasonable functioning of the meta-learner, and we decided a well-established method for recommending models. For testing our assumptions, we performed different preliminary applications of this version of the framework that highlighted a set of aspects to be improved. We iterated the testing and improvement steps until we achieved the current version of the ProMetaUS framework.

3. Meta-learning for algorithm selection problem

The ASP is addressed by meta-learning as a supervised learning task, whose aim is to learn a model that captures the relationship between the properties of the datasets (or the characteristics of learning problems) and the algorithms, in particular their performance (Abdulrahman et al., 2015, 2018; Brazdil et al., 2009; Cunha et al., 2018; Filchenkov and Pendryak, 2015; Khan et al., 2020; Prudêncio et al., 2011b, 2011d; Rossi et al., 2012; Shahoud et al., 2020; Smith-Miles, 2008a; Sousa et al., 2016). This model can then be used to predict the most suitable algorithm for a given new dataset (Abdulrahman et al., 2015, 2018). The readers can think the datasets as observations and measurements collected with a specific frequency by some equipment installed in a working environment, their properties as the minimum, maximum, median of each datasets, while the performance measures as criteria to measure the performance of a particular algorithm for a particular problem (Rice, 1976), such as accuracy and runtime.

The main steps of a typical meta-learning process for ASP is depicted in Fig. 2. Several authors describe the meta-learning system functioning in terms of training (or off-line) and use (or testing, on-line, prediction) phases: e.g. Bhatt et al., 2020; de Souto et al., 2008; Prudêncio et al., 2011a; Prudêncio and Ludermir, 2004; Romero et al., 2013; Shahoud et al., 2020; Zorrilla and García-Saiz, 2015. Other authors (Bhatt et al., 2012; Vilalta et al., 2004, 2009) refer to the following two modes of operation of a meta-learning system: (knowledge) acquisition mode and advisory mode.

During the acquisition mode, the main result is meta-knowledge (or meta-data). Meta-knowledge is the knowledge extracted from the learning process (Brazdil et al., 2009), and obtaining it is a crucial step for the success of a meta-learning system (Castiello et al., 2005; Ferrari and de Castro, 2015). It consists of meta-features and meta-target or performance of the algorithms (Abdulrahman et al., 2015; Bhatt et al., 2012; Brazdil et al., 2017; Cohen-Shapira et al., 2019; Ferrari and de Castro, 2015; Khan et al., 2020; Pimentel and de Carvalho, 2019b; Pinto et al., 2014, 2016; Prudêncio et al., 2011b, 2011c, 2011d; Zhu et al., 2018). Note that meta-target (also called target meta-feature) corresponds to the type of output that the system produces in the form of estimated relative performance of candidate algorithms for any given problem (Ferrari and de Castro, 2015; Khan et al., 2020). In other words, meta-knowledge "is stored as an object composed of meta-attributes, which characterize the problems, and the ranking, which indicates the performance of the algorithms" (Ferrari and de Castro, 2015), which can be simply represented in a tabular form.

The generation and extraction of informative and useful meta-features (e.g. simple, information-theoretic, model-based, landmarking) are important and challenging parts of the algorithm selection process, and constitute critical aspects for its success (Brazdil et al., 2003, 2009; Castiello et al., 2005; Dyrmishi et al., 2019; Filchenkov and Pendryak, 2015; Kanda et al., 2016; Khan et al., 2020; Ler et al., 2018; Pimentel and de Carvalho, 2019a; Pinto et al., 2014, 2016; Ren et al., 2020; Rossi et al., 2017; Vilalta et al., 2004, 2009). Meta-features (also known as meta-attributes, data(set) characteristics/features, characteristics from a dataset, characteristics of datasets, data characterisation, or domain characteristics) can be defined as follows:

- common characteristics of several problems and tasks (Brazdil et al., 2009; Ferrari and de Castro, 2015; Filchenkov and Pendryak, 2015);
- features describing and extracted from the problem (Brazdil et al., 2009; de Souto et al., 2008; Ferrari and de Castro, 2015; Prudêncio et al., 2011a; Prudêncio and Ludermir, 2012);
- an abstraction of knowledge extracted from the dataset (Cui et al., 2016a; das Dôres et al., 2016).

The set of meta-features suitable for different meta-learning problems may vary substantially, and depends on the task, the datasets, and the algorithms (Brazdil et al., 2009; Ler et al., 2018).

The algorithm that models the relationship between meta-features and performance of candidate algorithms is a meta-learner (also known as meta-algorithm, learning algorithm, machine learning algorithm, or meta-level algorithm) (Khan et al., 2020; Prudêncio and Ludermir, 2012; Smith-Miles, 2008b; Sousa et al., 2016). Therefore, meta-learner is the algorithm used for the meta-learning (Brazdil et al., 2009), and its effectiveness increases as it accumulates meta-knowledge (Vilalta et al., 2004, 2009). The output of the meta-learner is the meta-model or meta-learning model (Brazdil et al., 2009), which is built to predict the target value for a new dataset (Reif, 2012).

The meta-model is used to recommend algorithm(s) for a specific dataset (Cohen-Shapira et al., 2019; Pinto et al., 2016). The form of recommendation generated by the meta-learning system determines the type of meta-target to learn (Brazdil et al., 2009). Brazdil et al. (2009) report the following four different types of meta-targets or types of algorithm recommendation methods:

1. the best algorithm in a set: for each dataset, the recommendation consists of a single base-algorithm;
2. a subset of algorithms: suggestion of a subset of algorithms that are expected to perform well, in relative terms, on the given problem;
3. ranking of algorithms: provision of an ordered set of algorithms;
4. estimated performance of algorithms: recommendations in the form of a value indicating the performance that each algorithm is expected to achieve; a set of estimates concerning the performance of the base-

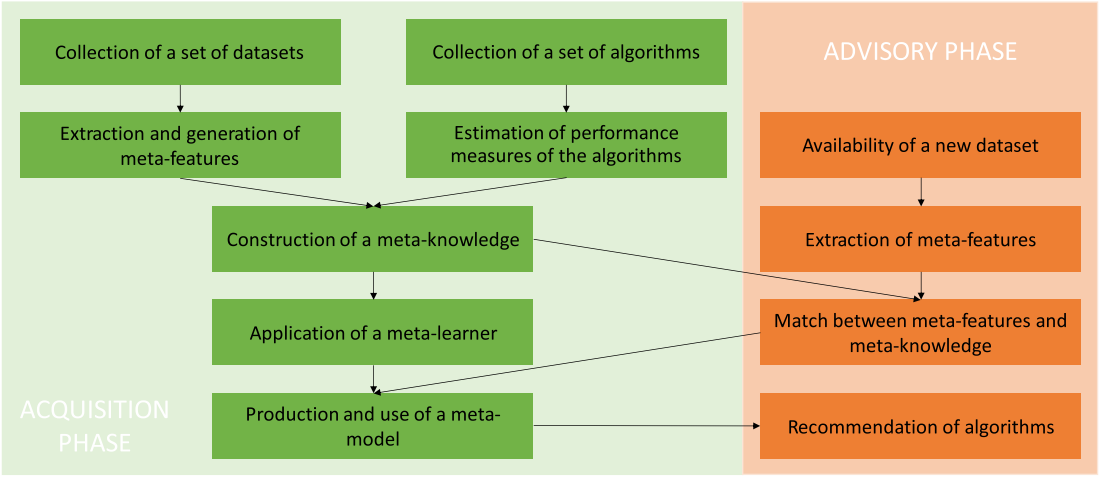

Fig. 2. Typical meta-learning process.

algorithms can be transformed into the other forms of recommendation.

4. The ProMetaUS framework

The plethora of available models for DRM to estimate and assess a hazard, consequence, likelihood, risk, or mitigation measure (indicated in the rest of the paper as "quantity of interest"), and the relevance of uncertainty analysis in such models motivate the interest in developing a meta-learning based approach to proactively rank them based on desired uncertainty conditions. The uncertainty conditions are evaluated by uncertainty and sensitivity analyses in order to completely characterise the state-of-knowledge affecting the assessment and decision.

An uncertainty analysis is a methodology that considers domain knowledge and its limitations in qualifying and/or quantifying the uncertainty in the structure of a scenario, structure of a model, model inputs, and model outputs (IPCS, 2008). Uncertainty analysis can ensure that decision processes are informed and transparent, and can help decision-makers define their confidence in model results and evaluate the utility of investing in reducing uncertainty, where feasible (Thompson and Warmink, 2017). This can be supplemented with sensitivity analysis to identify key sources of uncertainty for prioritising activities that could reduce uncertainty (IPCS, 2008). Sensitivity analysis is the study of how uncertainty in the model output can be apportioned to different sources of uncertainty in the model inputs (Saltelli et al., 2004). It aids to identify the scenarios and model inputs that are most responsible for the uncertainty in the variables of interest (de Rocquigny et al., 2008).

The ProMetaUS framework is of assistance to safety managers, practitioners, and researchers (indicated in the rest of the paper as "analysts") for identifying and predicting which model(s) is (are) the most suitable for the task under investigation, despite not being experts in all the existing models and data mining techniques. Such framework is a flexible and dynamic tool because allows continuously incorporating new models, and additional evidence and information when available. The gathering of further and updated data and information for the development of refined datasets and/or models also represents an approach to reduce uncertainty (Cullen and Frey, 1999).

The proposed framework is outlined in Fig. 3, and is based on the following general assumptions:

- the focus is on epistemic uncertainty, which refers to the lack of knowledge about the properties and conditions of the phenomena underlying the behaviour of the systems (Zio and Pedroni, 2012);
- the contribution of variability (defined as the heterogeneity of values over time, space, or different members of a population, and described as the inherent randomness of the natural system by Cullen and Frey, (1999), IPCS (2008), and Thompson and Warmink (2017)) is neglected;
- the available data are constituted by the model inputs for estimating the quantity of interest, no measurement outputs of the quantity of interest are required, and thus the validation of the models is supposed to be already performed; for these reasons, the comparisons between the measured and predicted quantity of interest are ignored in terms of meta-features and performance measures; consequently, our sets of meta-features and algorithm performance metrics do not include statistical performance measures, such as fractional bias, normalised mean square error, geometric mean bias, or geometric mean variance reported for instance in Chang and Hanna (2004), and Ivings et al. (2016);
- the probabilistic approach is applied for analysing uncertainty and thus the uncertainty is described by means of probability distributions (in accordance with the suggestion by Morgan and Henrion (1990), and the statement by Verma et al. (2010): "the most common approach used to represent uncertainty regarding a quantity is to use probability distributions").

4.1. Dataset characterisation and meta-features

The first element of the framework is constituted by datasets. We take inspiration from Santos et al. (2012): a dataset is a collection of data organised in a certain format, containing more than one instance in a specific domain. An instance is a row in a specific dataset describing an observation of a known event in the past in that particular domain (Santos et al. 2012).

We suppose that real-world and comparable datasets are derived from the organisations and regard the outcomes of measurements, expert judgments, and/or previous estimates of the quantity of interest by feeder models. Particularly, each dataset contains the parameters (or variables) and their related values that are relevant for the *quantity of interest* under investigation. Data can be derived from sensors and

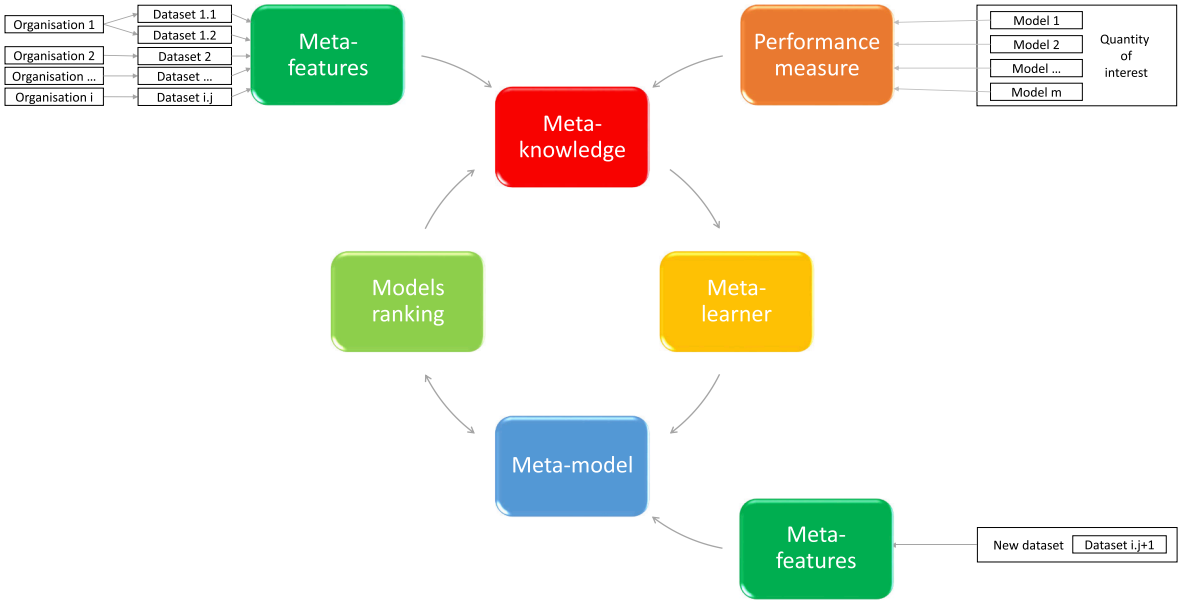

Fig. 3. The ProMetaUS framework to select models for DRM.

measuring devices, literature sources, and/or equipment designers and suppliers. In addition to these values, other pieces of information can be used, such as types of sensors and sensor placement, the department or process unit where each instance is collected, and the indication of the type of industry. The outcome or dependent variable is represented by the parameters specific for the quantity of interest: the knowledge of the outcome variable is required for implementing a strategy of supervised learning (i.e. the training data contain inputs and explicit outputs, and the model can be trained until it produces the correct output for a given input, as underlined by Smith-Miles (2008b)). The available datasets are split into training and testing data, as stated by Pise and Kulkarni (2016). Training data contain a known output (related to the meta-learning process, not simply obtained by the use of the considered base-models) from which the model learns and generalises, while test data have the goal to test the model prediction. Reasonable number of datasets is required that can appropriately map the feature space into the performance space and in order to make the model of the learning process more predictive (Khan et al., 2020; Prudêncio et al., 2011c; Reif et al., 2012).

To characterise datasets, a combination of meta-features that are representative of their problem domain and good predictors of the relative performance of algorithms should be identified and extracted (Bhatt et al., 2012; Brazdil et al., 2003, 2009; Lemke et al., 2015; Pinto et al., 2014, 2016). Our set of meta-features consists of a combination of properties computed directly and indirectly from datasets. Because the task under analysis regards the prediction of a ranking of models for estimating and assessing a quantity of interest under uncertain conditions, the following two categories are proposed:

1. meta-features to describe the input datasets and understand how the input uncertainty affects the model performance: (a) data-based meta-features for characterising datasets, and (b) domain-based meta-features for defining the problem under investigation and indicating the specific source of each dataset;
2. meta-features concerning the outcomes of simple and fast uncertainty and sensitivity analyses performed for each model on the datasets considered for that specific quantity of interest: (a) simple and fast uncertainty-based meta-features, and (b) simple and fast sensitivity-based meta-features.

Examples of meta-features belonging to the first category are the probability distributions, minimum, most likely, maximum, mean/median, and standard deviation values for each model input. This category can also include any dependencies and correlations among inputs. The second category of meta-features was inspired from landmarking concept (e.g. Bensusan and Giraud-Carrier, 2000; Pfahringer et al., 2000): estimations obtained by running models with common sampling techniques and a small number of iterations and simulations are expected to provide valuable information about properties affecting model performance. This type of information should also be useful for understanding which input parameters are not significantly subject to uncertainty and thus may be set to a fixed and constant value.

4.2. Models and performance measures

The problem space is composed by all the candidate models able to estimate and assess the quantity of interest under analysis. Such models should be selected in such a way to provide a wide range of characteristics and give some generality to the results (Prudêncio et al., 2011a), be comparable (e.g. time-dependent vs time-independent models), and be verified and validated to guarantee sufficiently realistic descriptions of the investigated phenomenon.

To obtain the performance measures, the considered models should be executed over each dataset. Due to the fact that performance measure definition is based on the type of the task (Filchenkov and Pendryak, 2015) and, in our case, also on the purpose of the performed assessment, no single performance measure is comprehensive enough to capture all the aspects that should be taken into account by the analysts. Additionally, some metrics may be more appropriate than others depending on the task and assessment. Consequently, we propose different single performance measures, grouped into various sets, outlined in Table 2. We defined these performance measures taking inspiration from different references available in the literature (e.g. Abdo et al., 2017; de Rocquigny, 2009; Loucks, 2002; Loucks et al., 2005; Yegnan et al., 2002).

The meta-learning strategy for confidence-based, shape-based, and sensitivity-based performance measures can be common and valid for all the types of quantity of interest and models. On the contrary, the safety-based performance measures and thus meta-learning strategies are quite different depending on the objectives of the assessment, models, quantity of interest, and task under investigation.

Confidence-based performance measures involve the confidence interval and coefficient of variation. A model with a narrower confidence interval and/or lower value of coefficient of variation is able to estimate a more precise quantity of interest (if the probability distributions used to describe uncertainty are based on justifiable statistical data). On the

Table 2
Performance measures of models for DRM.

Set of performance measures	Single performance measures
Confidence-based performance measures	Confidence interval
	Coefficient of variation
Shape-based performance measure	Skewness
	Kurtosis
Safety-based performance measures in the presence of a threshold value (assuming that there are no uncertainties about this threshold value)	Probability not exceeding the threshold value
	Skewness
	Comparison between the maximum and threshold value
Safety-based performance measures in the presence of more than one threshold value (i.e. several categories of limits)	Probability that the quantity of interest is between the two threshold values
	Skewness
Safety-based performance measures in the presence of a minimum desired level	Probability exceeding the value of the minimum desired level
	Skewness
Safety-based performance measures in the absence of a threshold value, but in the presence of indications and suggestions about adverse negative effects	Probability that the quantity of interest is between two values suggested in the literature
	Skewness
Safety-based performance measures in the absence of values for comparing the predictions	Minimum
	Maximum
	Mean/Median
Sensitivity-based performance measures	Coefficients and indices of inputs obtained thanks to typical sensitivity methods (e.g. partial correlation coefficients, multiple linear regression analysis and coefficient of determination, rank correlation coefficients)
	Coefficients greater than a defined threshold/target
	Coefficients and indices of a specific input of interest

contrary, a model with a higher confidence interval is more conservative than others, and higher values of the coefficient of variation indicate a greater level of dispersion around the mean. Note that to compare the several models, a confidence interval equal for each one should be chosen.

Shape-based performance measures reflect the measurable quantities used for analysing the shape of a distribution (Cullen and Frey, 1999). These are skewness (i.e. representing the asymmetry of a distribution) and kurtosis (i.e. indicating the flatness or peakedness of a distribution) (Cullen and Frey, 1999; Lee et al., 2019).

Skewness also captures useful characteristics of the tail of a distribution that should be evaluated differently according to the type of assessment and quantity of interest. Indeed, if the quantity of interest should be compared with a threshold value, analysts should prefer a model minimising the tail to the right. On the contrary, if the comparison should be performed between the quantity of interest and a minimum desired level, a model that minimises the tail to the left should be preferred. Finally, when there are more than one threshold value or suggested category of limits for the quantity of interest, models characterised by a skewness near to 0 are more suitable because of the reduction of the values in the tails.

Besides skewness, safety-based performance measures include typical decision criteria in the safety-related decision-making (e.g. the compliance with a limit or threshold value). By inheriting a concept from Rao (2005), we believe that the definition of our safety-based performance measures could also be useful for understanding if some controls are advised or required. For instance:

- if the lower confidence limit is above a threshold value, then appropriate controls are probably needed;
- if the upper confidence limit is below the threshold value, a control is probably not required;
- if the upper confidence limit is above the threshold value but the 50th percentile is below it, further study on the parameters dominating the overall uncertainty should be recommended;
- if the 50th percentile is also above the threshold value, further study may be recommended, but cost-effective controls for risk reduction could be implemented (Paltrinieri et al., 2012).

Sensitivity-based performance measures identify the most influential model inputs and this information can assist in uncertainty reduction efforts. Indeed, if all the models agree that a specific input is one of the most contributing factor to uncertainty, the efforts should be addressed for reducing its uncertainty in order to reduce the overall uncertainty. On the contrary, if all the models agree that a specific input is not an influencing factor, the uncertainty of this factor is not relevant for the overall uncertainty and should be neglected. Sensitivity analyses also permit distinguishing the input(s) with coefficients greater than a specified target and so characterised by an excessive and no acceptable uncertainty level.

The above performance measures can be estimated in different time instants in time-dependent models. This allows assessing the temporal propagation and time evolution of uncertainties for the quantity of interest, and thus recommending the most suitable model(s) at various points in time.

4.3. Meta-knowledge and meta-learner

As displayed in Fig. 3, the combination of meta-features and performance of the models or meta-target represents meta-knowledge, whose objective is to capture certain relationships between the dataset characteristics and the performance of the models (Brazdil et al., 2003, 2009; Brazdil and Soares, 2000; Soares and Brazdil, 2002).

Similarly to what is described in the literature (Bhatt et al., 2012; Brazdil et al., 2017; Cohen-Shapira et al., 2019; Cunha et al., 2018; de Souto et al., 2008; Pimentel and de Carvalho, 2019a, 2019b; Pinto et al., 2016; Prudêncio and Ludermir, 2012; Prudêncio et al., 2011a, 2011b, 2011c, 2011d; Rossi et al., 2017; Soares et al., 2009; Sousa et al., 2016; Zorrilla and García-Saiz, 2015), in the ProMetaUS framework the meta-learner is applied to the meta-knowledge to suggest a model that associates meta-features to the uncertainty and sensitivity metrics, and acquires knowledge to predict the performance of the models for new problems.

4.4. Ranking

In the ProMetaUS framework, the method for recommending models is ranking, which produces an ordered list of models, sorted according to their expected performance measure(s) for the dataset of interest (Brazdil et al., 2009; das Dôres et al., 2016; Reif, 2012; Tripathy and Panda, 2017; Zhu et al., 2018). Such a method is an advantageous option because of the following reasons:

- to be a more flexible and informative option compared to the selection of the best algorithm because it suggests more options (Brazdil et al., 2003; de Souto et al., 2008; Ferrari and de Castro, 2015; Soares and Brazdil, 2000; Vilalta et al., 2004, 2009);
- to allow the user to select either a single algorithm or more than one in accordance with the available resources (dos Santos et al., 2004; Prudêncio et al., 2011a; Soares and Brazdil, 2000);
- to furnish alternative solutions to users who may wish to incorporate their own expertise or any other criterion into their decision-making process (Vilalta et al., 2004, 2009);
- to offer a next best alternative if the first algorithm seems to be suboptimal (Brazdil et al., 2017), e.g. due to computational times, its overall complexity, or training requirements for its proper use;
- to develop the meta-learning system without any information about how many base-algorithms the user will try out (Brazdil et al., 2009).

Since the performance measures are different from each other, we propose to produce single rankings for each performance measure for each model applied for each dataset (referred as “single-performance measure rankings”), and then estimate a single final score for each model (defined “final score”). By taking into account every dataset, the final model score is based on the model ranks in the single-performance measure rankings and calculated thanks to the proposal by Tripathy and Panda (2017). The values of such final metrics guide the determination of the final ranking of models for a specific dataset. These estimations are repeated for all datasets in order to produce a global ranking of all candidate models by means of one of the possible methods mentioned in the literature (e.g. Bhatt et al., 2012, 2020; Brazdil and Soares, 2000; Ferrari and de Castro, 2015; Soares and Brazdil, 2000; Tripathy and Panda, 2017): average ranking, score ranking, winner ranking, ideal ranking, relative ranking, percentage ranking, zoomed ranking.

When a new dataset becomes available, the ProMetaUS framework allows extracting the meta-features describing these data, and using the meta-model previously produced to predict the performance and ranking of the models for that dataset. In addition, such ranking is continuously updated when new meta-features are obtained. Indeed, as noted by Brazdil and Giraud-Carrier (2018), as tests proceed on a new dataset, the tests already carried out can be interpreted as also constituting meta-knowledge. In this sense, meta-knowledge is acquired in previous learning episodes on datasets and/or from different domains or problems (Lemke et al., 2015).

5. Case study

In order to preliminarily test the ProMetaUS framework for selecting and ranking models for DRM, we considered a case study about ODH assessments. ODH occurs when the indoor oxygen (O_2) content drops to a level that may expose workers to the asphyxiation risk (Stefana et al., 2015). In the literature there is no consensus about a safe Threshold

Limit Value (TLV) for the O_2 content in terms of concentration by volume and/or atmospheric partial pressure (Stefana et al., 2015). However, typical human body reactions due to exposures to O_2 deficient atmospheres, and correlations between O_2 concentrations and symptoms (Stefana et al., 2016, 2019b) can give some guidelines about the hazardousness of investigated scenarios.

Several causes can be responsible for an O_2 reduction in a working environment, such as combustion of flammable substances, consumption due to chemical reactions, overcrowding in the workplace, release of inert gases, evaporation of cryogenic liquids (Stefana et al., 2015, 2019b). ODH is a common hazard in different kinds of working environments: O_2 deficiency is a well-recognised cause of death in confined spaces (McManus and Haddad, 2015), and can produce severe adverse health effects also in laboratories, manufacturing firms, and process industries. In such working environments, inert gases (e.g. nitrogen and argon) are often present (Stefana et al., 2015) and their releases may lead to the displacement and the consequent reduction of O_2 in the air (Stefana et al., 2019b). Additionally, such substances are particularly insidious since they are odourless, colourless, tasteless, and so undetectable by the exposed people (EIGA, 2018).

Therefore, proper ODH assessments should be performed in order to evaluate the criticality of an O_2-deficient atmosphere and identify adequate risk reduction controls for minimising the individuals' exposure. Recent scientific contributions (e.g. Stefana et al., 2015, 2016, 2019b, 2021) emphasise the potential and usefulness of the application of predictive models for such purpose. The available predictive models for ODH assessments due to inert gas releases permit estimating the indoor O_2 level, mainly in terms of O_2 concentration by volume, as a constant value instantaneously achieved or as a function of time (Stefana et al., 2015, 2016). Such models consider various parameters: initial indoor conditions (working environment volume, air composition, temperature and pressure), outdoor variables (air composition, temperature, atmospheric pressure), ventilation aspects of the working environment (forced and/or natural ventilation systems, supply and/or return air sub-systems, number of ventilation systems, airflow rates), and inert gas releases (flow rates, types of releases). These parameters are differently defined and combined in the mathematical formulations of the models. Moreover, their assumptions are quite diverse: some of them are based on rather simplified hypotheses, whereas others introduce some refinements in order to furnish more precise estimations of the indoor O_2 levels (e.g. Stefana et al., 2017).

In such a context, analysts should consider a wide range of the properties of the models (e.g. characteristics, assumptions, initial and boundary conditions) to select the one(s) most suitable for conducting proper ODH assessments in the analysed scenarios based on desired uncertainty conditions. The ProMetaUS framework helps these professionals with this task and offers a ranking of the predictive models. These existing models supporting ODH assessments are deterministic and provide a single-point estimate of O_2 levels (Stefana et al., 2019a). Consequently, to evaluate their uncertainty, we used Microsoft Excel® spreadsheets and Palisade add-in @RISK (version 7.5.1). @RISK permits including the uncertainty in the models and generating a range of possible outcomes (Stefana et al., 2019a).

The case study is focused on scenarios where ventilation systems draw air from the working environment with a rate greater than release flow rate. The following nomenclature is adopted:

- $C_{O2}(t)$ is the O_2 concentration by volume in the working environment at the time t (%);
- 21% is the fixed initial O_2 concentration in the working environment;
- Q_{out} is the output forced ventilation airflow rate ($m^3\ s^{-1}$);
- R is the flow rate of the gas released in the working environment ($m^3\ s^{-1}$);
- V is the working environment volume (m^3);
- t is the time from the start of the release (at the beginning of the release t = 0) (s);
- $C_{O2,air}$ is the O_2 concentration in the air (due to ventilation airflow rate) (%);
- $C_{O2}(0)$ is the O_2 concentration by volume in the working environment at the time t = 0 s (%).

6. Results

6.1. Dataset characterisation and meta-features

Five training datasets (in the following indicating as Dataset 1, Dataset 2, Dataset 3, Dataset 4, and Dataset 5) are supposed, which consist of several instances representing observations of measurements and information relevant for ODH. Typical collected data are: volume of the working environment, initial indoor O_2 concentration, inert gas and ventilation air flow rates, and O_2 concentration of the ventilation air flow rates. Such datasets are characterised by means of different meta-features, as summarised in Table 3. A set of 24 meta-features are extracted for each dataset, which regard the probability distributions, and minimum, most likely, maximum, mean, and standard deviation values for the parameters R, Q_{out}, $C_{O2}(0)$, and $C_{O2,air}$.

In particular, the release flow rate is described by means of triangular, uniform, or normal distribution, while the output forced ventilation airflow rate approaches a normal or uniform distribution. Triangular distribution is used to represent uncertainty when only upper and lower bounds and a most likely value are known, uniform one characterises phenomena for which only an upper and lower bound can be estimated, and normal one designates the distribution of means of independent observations from any distribution or any combination of distributions as the number of observations becomes large, or the distribution of the sum of samples from a large number of distributions, which may be of any shape, as the number of input distributions increases (Cullen and Frey, 1999). In addition to uniform and normal distributions underlying data, the O_2 concentration in the air due to ventilation airflow rate or the O_2 concentration by volume in the working environment at the time t = 0 s can be characterised through a lognormal distribution, which assumes only non-negative values and describes random variables resulting from multiplicative processes (Cullen and Frey, 1999).

Since no correlations are imposed on the input variables, meta-features relating to dependencies are not hypothesised. Moreover, in this case study, we do not consider simple and fast uncertainty-based and sensitivity-based meta-features because of the already limited number of iterations chosen by the @RISK (setting Auto-Stop mode) during the execution of each model over each dataset.

6.2. Models and performance measures

Among the candidate models available in the literature (reviewed in Stefana et al. (2015)) to estimate the time trend of indoor O_2 levels in scenarios where ventilation systems draw air from the working environment with a rate greater than release flow rate, we focus the attention on a suite of three models. Such models (indicated as Model 1, Model 2, and Model 3) are summarised by means of Eq. (1), Eq. (2), and Eq. (3).

$$C_{O2}(t) = \frac{21\%}{Q_{out}+R}\left[Q_{out}+R\exp\left(-\frac{Q_{out}+R}{V}t\right)\right] \quad (1)$$

$$C_{O2}(t) = 21\%\left\{1-\frac{R}{Q_{out}}\left[1-\exp\left(-\frac{Q_{out}}{V}t\right)\right]\right\} \quad (2)$$

$$C_{O2}(t) = C_{O2,air}\left(1-\frac{R}{Q_{out}}\right)+\left[C_{O2}(0)-C_{O2,air}\left(1-\frac{R}{Q_{out}}\right)\right]\exp\left(-\frac{Q_{out}}{V}t\right) \quad (3)$$

We assumed that the working environment volume is not affected by uncertainty issues, and thus it is a fixed value and equals to 50 m^3. The task in this case study is related to predict the indoor O_2 concentration

Table 3
Meta-features extracted from the five datasets.

Dataset	Pr.distr. for R	Min	M. like	Max	μ	σ	Pr.distr. for Q_{out}	Min	M. like	Max	μ	σ	Pr.distr. for $C_{O2}(0)$	Min	M. like	Max	μ	σ	Pr.distr. for $C_{O2,air}$	Min	M. like	Max	μ	σ
Dataset 1	Triangular	0	0.2	0.4	–	–	Uniform	0.5	–	1	–	–	Lognormal	–	–	23.0%	21.0%	1.0%	Lognormal	–	–	23.0%	21.0%	1.0%
Dataset 2	Uniform	0	–	0.2	–	–	Uniform	0.5	–	1	–	–	Uniform	18.0%	–	21.0%	–	–	Uniform	18.0%	–	21.0%	–	–
Dataset 3	Normal	0	–	0.4	0.2	0.05	Normal	0.5	–	1	0.75	0.05	Lognormal	–	–	23.0%	21.0%	1.0%	Lognormal	–	–	23.0%	21.0%	1.0%
Dataset 4	Normal	0	–	0.4	0.2	0.05	Normal	0.5	–	1	0.75	0.05	Normal	19.5%	–	23.0%	21.0%	1.0%	Normal	19.5%	–	23.0%	21.0%	1.0%
Dataset 5	Triangular	0	0.2	0.4	–	–	Normal	0.5	–	1	0.75	0.05	Lognormal	–	–	23.0%	21.0%	1.0%	Lognormal	–	–	23.0%	21.0%	1.0%

Note: Pr.distr. = Probability distribution; R = Flow rate of the gas released in the working environment ($m^3 s^{-1}$); Min = Minimum; M.like = Most likely; Max = Maximum; μ = Mean; σ = Standard deviation; Q_{out} = Output forced ventilation airflow rate ($m^3 s^{-1}$); $C_{O2}(0)$ = Oxygen concentration by volume in the working environment at initial time 0 (%); $C_{O2,air}$ = Oxygen concentration in the air (due to ventilation airflow rate) (%).

by volume at t = 500 s.

The three predictive models are executed and evaluated on each dataset. Such execution of the models through Latin Hypercube Sampling (LHS) and an automatic number of iterations allows estimating their performance measures, reported in Table 4, Table 5, Table 6, Table 7, and Table 8. Note that LHS is able to produce more accurate results (Albright et al., 2006), and in many cases is preferred as a numerical simulation method (Cullen and Frey, 1999).

We use the following performance metrics to compare the models:

- 90% confidence interval and coefficient of variation for confidence-based performance measures;
- kurtosis as an indication of shape-based performance metrics;
- probabilities that the O_2 concentration by volume is between 10% and 12%, 12% and 15%, and 15% and 18%, and skewness for safety-based performance measures (since ODH is not characterised by a threshold value, but frequently described in terms of adverse negative effects based on O_2 concentrations by volume);
- Spearman rank correlation coefficients for R and Q_{out} (in all models), and of $C_{O2}(0)$ and $C_{O2,air}$ (taking into consideration Model 3) to evaluate sensitivity-based performance metrics.

We focus on the O_2 concentration by volume between 10% and 12%, 12% and 15%, and 15% and 18% because of the criticality of the physiological effects, as underlined in Stefana et al. (2016, 2019b). Indeed, when the O_2 level drops to a concentration lower than 18%, symptoms such as decreased ability to perform tasks, accelerated heartbeat, dizziness, and/or loss of muscle control may manifest. If the O_2 concentration decreases further and achieves values around 10–12%, the symptoms become more severe and workers may experience loss of consciousness, permanent brain damage, possible damage to the heart, and/or very poor muscular coordination.

Finally, we take into account Spearman rank correlation coefficient as a sensitivity-based performance measure because of its property of detecting nonlinear monotonic dependencies (Borgonovo and Plischke, 2016).

6.3. Meta-knowledge and meta-learner

The characterisation of datasets by meta-features and estimation of performance measures of the models permit to obtain meta-knowledge thanks to the analysis of the relationships between them, and later apply a meta-learner for inducing a model associating the meta-features to the different uncertainty and sensitivity metrics.

Confidence-based performance measures highlight that Model 1 outperforms both Model 2 and Model 3: since the first model produces slightly narrower confidence intervals and lower values of the coefficients of variation in comparison to the others, its estimations are more precise (assuming that probability distributions used to represent uncertainty in the input parameters are based on justifiable statistical data). Model 1 also presents a probability that the O_2 concentration by volume is between 10% and 12% equal to 0%, and higher probabilities for concentrations above 12%. From these results, analysts can assess with a reasonable degree of confidence that the O_2 concentration does not drop to a level that may cause serious adverse effects such as loss of consciousness, very poor muscular coordination, or damage to the heart to workers. This model also produces the probabilities of O_2 concentration lower than 10% equal to 0%, thus severe symptoms related to unconsciousness, coma, and death are not likely to occur. Such results could also be influenced by higher means and lower standard deviations of the probability distribution of the O_2 concentration. On the contrary, Model 2 and Model 3 executed on Dataset 1 and Dataset 5 present a probability that the O_2 concentration by volume is between 10% and 12% higher than 5%: analysts should pay attention to the hazardousness of the indoor air pointed out by these models when the release flow rate is described by a triangular distribution.

Table 4
Performance measures for Dataset 1.

Model	90% CI	CV	Kurtosis	Pr(10–12%)	Pr(12–15%)	Pr(15–18%)	Skewness	Corr.coeff. for R
Model 1	±0.267%	0.097	2.8595	0.00%	16.54%	64.71%	0.1685	−0.88
Model 2	±0.317%	0.180	3.2876	5.93%	36.19%	39.74%	−0.5533	−0.89
Model 3	±0.311%	0.176	3.0871	9.11%	30.31%	41.96%	−0.4153	−0.85

Note: CI = Confidence interval; CV = Coefficient of variation; Pr = Probability; Corr.coeff. = Correlation coefficients; R = Flow rate of the gas released in the working environment ($m^3\ s^{-1}$).

Table 5
Performance measures for Dataset 2.

Model	90% CI	CV	Kurtosis	Pr(10–12%)	Pr(12–15%)	Pr(15–18%)	Skewness	Corr.coeff. for R
Model 1	±0.232%	0.076	2.0898	0.00%	0.00%	39.12%	−0.14	−0.95
Model 2	±0.289%	0.097	2.3293	0.00%	5.34%	43.13%	−0.278	−0.95
Model 3	±0.33%	0.120	2.3558	1.31%	21.39%	51.65%	−0.2714	−0.88

Note: CI = Confidence interval; CV = Coefficient of variation; Pr = Probability; Corr.coeff. = Correlation coefficients; R = Flow rate of the gas released in the working environment ($m^3\ s^{-1}$).

Table 6
Performance measures for Dataset 3.

Model	90% CI	CV	Kurtosis	Pr(10–12%)	Pr(12–15%)	Pr(15–18%)	Skewness	Corr.coeff. for R
Model 1	±0.147%	0.053	3.2706	0.00%	3.54%	90.73%	0.322	−0.96
Model 2	±0.257%	0.101	3.1248	2.61%	35.96%	58.10%	−0.1284	−0.97
Model 3	±0.25%	0.098	3.9715	2.03%	36.00%	57.16%	0.0709	−0.81

Note: CI = Confidence interval; CV = Coefficient of variation; Pr = Probability; Corr.coeff. = Correlation coefficients; R = Flow rate of the gas released in the working environment ($m^3\ s^{-1}$).

Table 7
Performance measures for Dataset 4.

Model	90% CI	CV	Kurtosis	Pr(10–12%)	Pr(12–15%)	Pr(15–18%)	Skewness	Corr.coeff. for R
Model 1	±0.154%	0.056	3.883	0.00%	2.44%	92.03%	0.5206	−0.96
Model 2	±0.25%	0.098	2.6042	1.57%	44.34%	52.09%	−0.0802	−0.96
Model 3	±0.241%	0.094	3.3424	1.40%	37.84%	55.28%	0.1313	−0.87

Note: CI = Confidence interval; CV = Coefficient of variation; Pr = Probability; Corr.coeff. = Correlation coefficients; R = Flow rate of the gas released in the working environment ($m^3\ s^{-1}$).

Table 8
Performance measures for Dataset 5.

Model	90% CI	CV	Kurtosis	Pr(10–12%)	Pr(12–15%)	Pr(15–18%)	Skewness	Corr.coeff. for R
Model 1	±0.258%	0.093	2.3777	0.00%	14.31%	63.29%	0.3794	−0.99
Model 2	±0.368%	0.146	2.5515	8.29%	33.71%	46.13%	−0.166	−0.98
Model 3	±0.272%	0.151	2.6271	7.61%	36.03%	41.82%	0.1783	−0.95

Note: CI = Confidence interval; CV = Coefficient of variation; Pr = Probability; Corr.coeff. = Correlation coefficients; R = Flow rate of the gas released in the working environment ($m^3\ s^{-1}$).

Model 2 and Model 3 applied on Dataset 3 and Dataset 4 provide skewness values near to 0: such datasets are characterised by normal and/or lognormal distributions assigned to input parameters. Additionally, when R and Q_{out} are represented by means of a normal distribution, confidence intervals and coefficients of variation assume lower values. Better shape-based performance measures (specifically, lower values of kurtosis) seem to be correlated to uniform distributions assigned to input parameters (Dataset 2). Based on an initial sensitivity analysis considering all the parameters of the models (since the complete outcomes of this sensitivity analysis are not reported in the paper because they do not represent a key result of the application of our framework, please contact the authors for the complete set of results), R is the most influential model input for all datasets for all candidate models. Consequently, in order to reduce the overall uncertainty and thus increase the confidence in the O_2 concentration, analysts should try to decrease the uncertainty involved in this parameter (e.g. constant monitoring of the release flow rate, adoption of gas detection equipment and control devices).

6.4. Ranking

To generate the model rankings, we focus our attention on the 90% confidence interval in terms of confidence-based performance measures, kurtosis for shape-based performance metrics, probability that the O_2 concentration by volume is between 10% and 12% and skewness as safety-based performance measures, and rank correlation coefficients for R for sensitivity-based performance metrics. Table 9, Table 10, Table 11, Table 12, and Table 13 summarise the single-performance measure and final rankings for each dataset. Note that final score is calculated for each model and for each dataset according to Eq. (4).

$$\text{Final score} = \sqrt{\sum_{j}\left((m+1)\ \text{-}\ \text{Ranking}_j\right)^2} \tag{4}$$

where j is the number of single-performance measure rankings (in our case study: j = 5), m is the number of the models (in our case study: m = 3), and Ranking is the single ranking produced for each performance

Table 9
Single-performance measure rankings and final ranking for Dataset 1.

Model	Ranking for CI	Ranking for kurtosis	Ranking for Pr(10–12%)	Ranking for skewness	Ranking for corr.coeff. for R	Final score	Final ranking
Model 1	1	1	1	1	2	6.32	1
Model 2	3	3	2	3	3	2.83	3
Model 3	2	2	3	2	1	4.69	2

Note: CI = Confidence interval; Pr = Probability; Corr.coeff. = Correlation coefficients; R = Flow rate of the gas released in the working environment ($m^3\ s^{-1}$).

Table 10
Single-performance measure rankings and final ranking for Dataset 2.

Model	Ranking for CI	Ranking for kurtosis	Ranking for Pr(10–12%)	Ranking for skewness	Ranking for corr.coeff. for R	Final score	Final ranking
Model 1	1	1	1	1	2	6.32	1
Model 2	2	2	1	3	3	4.36	2
Model 3	3	3	3	2	1	4.00	3

Note: CI = Confidence interval; Pr = Probability; Corr.coeff. = Correlation coefficients; R = Flow rate of the gas released in the working environment ($m^3\ s^{-1}$).

Table 11
Single-performance measure rankings and final ranking for Dataset 3.

Model	Ranking for CI	Ranking for kurtosis	Ranking for Pr(10–12%)	Ranking for skewness	Ranking for corr.coeff. for R	Final score	Final ranking
Model 1	1	2	1	3	2	5.20	1
Model 2	3	1	3	2	3	4.00	3
Model 3	2	3	2	1	1	5.20	1

Note: CI = Confidence interval; Pr = Probability; Corr.coeff. = Correlation coefficients; R = Flow rate of the gas released in the working environment ($m^3\ s^{-1}$).

Table 12
Single-performance measure rankings and final ranking for Dataset 4.

Model	Ranking for CI	Ranking for kurtosis	Ranking for Pr(10–12%)	Ranking for skewness	Ranking for corr.coeff. for R	Final score	Final ranking
Model 1	1	3	1	3	3	4.58	3
Model 2	3	1	3	1	2	4.90	2
Model 3	2	2	2	2	1	5.00	1

Note: CI = Confidence interval; Pr = Probability; Corr.coeff. = Correlation coefficients; R = Flow rate of the gas released in the working environment ($m^3\ s^{-1}$).

Table 13
Single-performance measure rankings and final ranking for Dataset 5.

Model	Ranking for CI	Ranking for kurtosis	Ranking for Pr(10–12%)	Ranking for skewness	Ranking for corr.coeff. for R	Final score	Final ranking
Model 1	1	1	1	3	3	5.39	1
Model 2	3	2	3	1	2	4.36	3
Model 3	2	3	2	2	1	4.69	2

Note: CI = Confidence interval; Pr = Probability; Corr.coeff. = Correlation coefficients; R = Flow rate of the gas released in the working environment ($m^3\ s^{-1}$).

measure considered (in our case study: confidence interval, kurtosis, probability that the O_2 concentration by volume is between 10% and 12%, skewness, and correlation coefficients for R).

The single-performance measure rankings, which are obtained by aggregating the information of the several performance measures, underline that Model 1 outperforms the others on the three datasets characterised by a triangular or uniform distribution assigned to the flow rate of the gas released in the working environment. This model is also assigned rank 1 in the global average ranking (reported in Table 14) for the three candidate models based on all the five datasets considered in this case study. For details about the average ranking method, refer to Abdulrahman et al. (2015), and Brazdil and Soares (2000).

The consideration of a higher range of O_2 concentration by volume as a performance measure leads to a different ranking of the models. If analysts assume that the probability of the O_2 concentration by volume is between 15% and 18% (instead of a probability between 10% and 12%), Model 1 outperforms the others on only two datasets (Dataset 1 and Dataset 2). On the contrary, Model 3 achieves better performance measures for Dataset 3 and Dataset 5. These two datasets are characterised by different probability distributions for the flow rate of the gas released in the working environment, but the same probability distributions for the other three parameters: normal distribution for the output forced ventilation airflow rate, and lognormal distributions for both the O_2 concentration in the air and in the working environment at the time t = 0 s. Model 3 is also assigned rank 1 in the global average ranking for the three candidate models based on all the five datasets. Rank 3 assigns to Model 2 also in this case. An overview of the results obtained in this case is presented in Table 15 and Table 16, where the single-performance measure and final rankings for each model and each dataset, and the average ranking for the models on all dataset are reported, respectively.

Table 14
Average ranking for the models on all datasets.

Model	Average rank	Ranking
Model 1	1.4	1
Model 2	2.6	3
Model 3	1.8	2

7. Discussion

The task investigated through the case study has regarded the prediction of the ranking of the models able to estimate the indoor O_2 levels and thus assess ODH in working environments. Because of its simplicity, not all the aspects of the ProMetaUS framework were covered and

Table 15
Single-performance measure rankings and final ranking for all the models and datasets (considering the probability of the O_2 concentration between 15% and 18%).

Model	Dataset	Ranking for CI	Ranking for kurtosis	Ranking for Pr(15–18%)	Ranking for skewness	Ranking for corr.coeff. for R	Final score	Final ranking
1	1	1	1	3	1	2	5.66	1
2	1	3	3	1	3	3	3.61	3
3	1	2	2	2	2	1	5.00	2
1	2	1	1	1	1	2	6.32	1
2	2	2	2	2	3	3	3.74	3
3	2	3	3	3	2	1	4.00	2
1	3	1	2	3	3	2	4.36	2
2	3	3	1	2	2	3	4.36	2
3	3	2	3	1	1	1	5.66	1
1	4	1	3	3	3	3	3.61	3
2	4	3	1	1	1	2	5.66	1
3	4	2	2	2	2	1	5.00	2
1	5	1	1	3	3	3	4.58	3
2	5	3	2	2	1	2	4.69	2
3	5	2	3	1	2	1	5.20	1

Note: CI = Confidence interval; Pr = Probability; Corr.coeff. = Correlation coefficients; R = Flow rate of the gas released in the working environment ($m^3\ s^{-1}$).

Table 16
Average ranking for the models on all datasets (considering the probability of the O_2 concentration between 15% and 18%).

Model	Average rank	Ranking
Model 1	2.0	2
Model 2	2.2	3
Model 3	1.6	1

applied. However, such case study gives an introductory description of the pivotal components to be defined and steps to be followed by analysts in the proactive ranking of models for the assessment of one hazard based on desired uncertainty conditions by means of meta-learning notions. The same components and steps are valid for consequences, likelihoods, risks, and mitigation measures that are studied in the safety field and can be assessed by models. The developed framework follows the classical ASP architecture, which is largely implemented in computer science field, but never before presented in the safety domain for selecting models for estimating and assessing DRM core elements that considers uncertainty. It represents an analysis and decision support tool in real applications, whose proper functioning does not require user experience and competencies. The attainment of the models ranking by means of the ProMetaUS framework is possible by means of widely spread tools, e.g. Microsoft Excel® and software systems for risk analysis, simulation, uncertainty quantification, and sensitivity analysis.

The ProMetaUS framework is based on an automatic and continuous process that feeds itself from the application of a set of models on different datasets and problems. This is guaranteed by the process of acquiring and exploiting meta-knowledge, which is tightly linked to meta-learning (Vilalta et al., 2004). Such meta-knowledge is built by means of the combination of the meta-features with the performance measures across several datasets solved by different models. In the case study, the meta-knowledge is created through the mapping of the probability distributions and statistical parameters of the main inputs of the models to confidence levels, summary statistics, and sensitivity analysis outcomes obtained by executing each model on each dataset. The obtained results underline that models achieve better confidence-based performance measures when inert gas release flow rates and forced ventilation airflow rates are described by normal distributions; whereas, when uniform distributions are assigned to input parameters, lower values of kurtosis are estimated. For instance, the information about the value of kurtosis helps choose among the models because it permits identifying the candidate ones that calculate low probabilities of achieving extremely low or extremely high O_2 concentration. All these correlations between meta-features and performance measures are collected to generate meta-knowledge and to feed a meta-learner. Then, the meta-learner is applied for suggesting a meta-model able to produce the ranking and predict the performance of the models for new problems.

Additionally, an initial sensitivity analysis considering all the parameters of the models (not reported in this paper for the sake of brevity) highlights that all the candidate models agree that the release gas flow rate is the most influential input: for reducing the overall uncertainty in the O_2 concentration estimations, efforts should be mainly devoted to decrease the uncertainty involved in this parameter. A constant release monitoring, and the adoption of gas detection equipment and control devices are highly suggested for reducing this uncertainty, mainly in working environments where ODH can occur with a high probability (e.g. confined spaces). Note that the O_2 concentrations and the related ranges employed in the case study can be modified and customised based on the assessment purpose. Indeed, analysts can rely on some recommended minimum O_2 levels available in the literature (also including a margin of safety) for understanding the typical human body reactions due to O_2 deficient atmospheres and the consequent need of adoption of adequate measures for mitigating the asphyxiation risk (Stefana et al., 2021).

Variations of the O_2 concentration range permit analysing the behaviour of the candidate models and its impact on the ranks assigned to the models. Such possibility to define different groups of single performance measures depending on the types of tasks and purposes of the safety assessments emphasises the versatility of our framework. The generation of several single rankings based on each performance measure can allow the professional conducting the assessment to choose which criteria including and affecting the selection of the models. The selection is obtained by means of three different types of ranking, such as single-performance measure rankings for each dataset, a final ranking for each dataset, and a global ranking for all the candidate models on all datasets. The rankings are regularly updated whenever new-meta-features are gained, and meta-knowledge is continuously acquired as tests proceed on new datasets. This is in accordance with the definition proposed by Lemke et al. (2015): “A metalearning system must include a learning subsystem, which adapts with experience. Experience is gained by exploiting metaknowledge extracted (a) in a previous learning episode on a single dataset, and/or (b) from different domains or problems”.

The ProMetaUS framework permits obtaining not only a recommendation of the most adequate model(s) among the considered candidate models for a certain task, but also highlighting the confidence level in model results for guiding further uncertainty reduction efforts. In order to do that, the complete set of meta-features involves characteristics that describe datasets, define the problem under investigation, indicate the specific source of each dataset, and summarise outcomes of simple and fast uncertainty and sensitivity analyses performed for each model on the datasets. These meta-features appear in compliance with the principal guidelines available in the literature (Brazdil et al., 2003,

2009; Castiello et al., 2005; Kalousis and Hilario, 2001; Khan et al., 2020; Smith-Miles, 2008a): they should not be too computationally complex, be efficiently and uniformly computable for wide range of problems in a particular domain, and not be too large compared to the amount of available meta-data. However, such a set of meta-features could increase the required resources and computational time for obtaining the models ranking, especially if the models are more complex than ones analysed in the case study. To limit a possible time increment, further experiments by using different combinations of meta-features (e.g. employment of all defined meta-features, application of the meta-features that belong to only few groups, or usage of the most relevant ones chosen by a features selection algorithm) could be carried out. These experiments could thus help identify a reasonable number of meta-features (for the considered problem) that avoids a time-consuming selection process, and balances the trade-off between desired accuracy and computational efficiency. Future research could also investigate the inclusion of some meta-features related to the comparisons between the measured and predicted quantity of interest (if measurement outputs of the quantity of interest are available) in order to look for other relevant relationships between them and performance measures.

For bringing about a more predictive model of the learning process, particular attention should be given to the collection of an adequate number of datasets, also in different domains and problems. In the safety field, this can represent a challenging aspect due to the requested monitoring of many and various conditions and aspects, e.g. working environment settings, hazards, risks, equipment characteristics and failure analysis, present and produced substances.

The development of the ProMetaUS framework is based on a set of general assumptions that could reduce its field of application. For instance, our focus is on epistemic uncertainty because it can be reduced by improved system understanding (Thompson and Warmink, 2017). However, the understanding of the behaviour of complex systems is not trivial and is limited: in such situations a continuous updating of the model parameters and input data, and the analysis of the performance of the system over time are needed. This represents an interesting topic that requires future works. Moreover, we deal with the uncertainty by means of the probabilistic approach and probability distributions. This is possible for those uncertainty sources that are quantifiable through quantitative approaches; however, there are other sources of uncertainty that cannot be described in quantitative terms (e.g. social influences on system performance) and that could be included in an advanced future framework thanks to ad hoc adjustments. Regarding the defined performance measures, we assume that threshold values, if available, are not characterised by uncertainty degrees: although this hypothesis seems reasonable in several scenarios in the safety field (e.g. TLVs published by American Conference of Governmental Industrial Hygienists), a proper caution should be dedicated to this aspect by analysts. A certain level of prudence should also be used when some model inputs are neglected in the analysis because all the models agree that such inputs are not relevant contributing factors to the overall uncertainty. This consideration remains valid as long as the set of models does not change: when a new model is incorporated in the problem space, additional sensitivity analyses should be conducted in order to understand if these model inputs can continue to be overlooked or should be introduced in the assessment.

Finally, this paper offers only one of the possible modes of connection between meta-learning concepts and the safety domain. The application of meta-learning notions seems to be particularly promising for making relevant progress in the risk reduction and prevention improvement. Various future directions could be imagined and should deserve further research efforts. For instance, the study of possible relations between system conditions and the consequent gathered expertise can help to efficiently and effectively assess new risks and define adequate safety measures. Furthermore, interesting stimuli could be derived from the topic of learning from accidents largely addressed in the safety field (Patriarca et al., 2019) for understanding how the complexity of the system contributes to the development of accident scenarios (Seligmann et al., 2019). In such a context, the meta-learning peculiarities applied to plant and procedures could serve as a valuable complement to the intrinsic human process of "learning about learning" in order to try to capture the complexity of socio-technical systems, explore the interactions among different cyber-socio-technical components (Patriarca et al., 2021), and thus enhance the effectiveness of hazard identification.

8. Conclusions

This paper proposes a novel proactive framework, named ProMetaUS after the Greek mythological figure Prometheus, based on a meta-learning system to select and rank models for the assessment of DRM core elements (hazard, consequence, likelihood, risk, or mitigation measures) depending on desired uncertainty conditions. Such framework gives safety managers, practitioners, and researchers a dynamic tool for incorporating further models, evidence, and information when available, without requiring expertise and competencies in all the existing models and data mining techniques.

A case study about ODH assessments in a working environment was presented to offer a preliminary application of the framework in the safety field. Three models predicting the time trend of the indoor O_2 concentration by volume were evaluated on five training datasets. We extracted 24 meta-features (e.g. probability distribution, mean, and standard deviation values for the flow rate of the gas release in the working environment and O_2 concentration in the air) and estimated 8 performance measures (e.g. confidence interval, kurtosis, Spearman rank correlation coefficients for model inputs) for each model and for each dataset. Single-performance measure and final rankings for each dataset and a global average ranking for the candidate models based on all the five datasets were recommended. Additional training datasets and models should be required for generalising the obtained results and improving the correlations between meta-features and performance measures in the occupational context of ODH assessments.

Further case studies about the application of the ProMetaUS framework considering other DRM core elements should be investigated. Interesting future applications could be focused on models for estimating the risk of fire due to O_2 enrichment or occupational exposures to chemicals. A deep analysis about the relevance of the designed meta-features and proposal of other meaningful characteristics to be extracted from datasets is an additional aspect that deserves future investigation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Abdo, H., Flaus, J.-M., Masse, F., 2017. Uncertainty quantification in risk assessment – representation, propagation and treatment approaches: application to atmospheric dispersion modeling. J. Loss Prev. Process Ind. 49, 551–571.

Abdulrahman, S.M., Brazdil, P., 2014. Measures for combining accuracy and time for meta-learning. In: Meta-Learning and Algorithm Selection Workshop at ECAI, pp. 49–50.

Abdulrahman, S.M., Brazdil, P., van Rijn, J.N., Vanschoren J., 2015. Algorithm Selection via Meta-learning and Sample-based Active Testing. In: MetaSel@ PKDD/ECML, pp. 55–66.

Abdulrahman, S.M., Brazdil, P., van Rijn, J.N., Vanschoren, J., 2018. Speeding up algorithm selection using average ranking and active testing by introducing runtime. Mach. Learn. 107, 79–108.

Albright, S.C., Winston, W., Zappe, C., 2006. Data Analysis & Decision Making with Microsoft Excel, third ed. Thomson South-Western, Mason, OH.

Arunraj, N.S., Mandal, S., Maiti, J., 2013. Modeling uncertainty in risk assessment: an integrated approach with fuzzy set theory and Monte Carlo simulation. Accid. Anal. Prev. 55, 242–255.

Bensusan, H., Giraud-Carrier, C., 2000. Discovering task neighbourhoods through landmark learning performances. In: Zighed, D.A., Komorowski, J., Żytkow, J. (Eds.), Principles of Data Mining and Knowledge Discovery. PKDD 2000. Lecture Notes in Computer Science, vol. 1910. Springer, Berlin, Heidelberg, pp. 325–330.
Bhatt, N., Thakkar, A., Bhatt, N., Prajapati, P., 2020. Algorithm selection via meta-learning and active meta-learning. In: Somani, A., Shekhawat, R., Mundra, A., Srivastava, S., Verma, V. (Eds.), Smart Systems and IoT: Innovations in Computing. Smart Innovation, Systems and Technologies, vol. 141. Springer, Singapore, pp. 169–178.
Bhatt, N., Thakkar, A., Ganatra, A., 2012. A survey and current research challenges in meta learning approaches based on dataset characteristics. Int. J. Soft Comput. Eng. 2 (1), 239–247.
Borgonovo, E., Plischke, E., 2016. Sensitivity analysis: a review of recent advances. Eur. J. Oper. Res. 248, 869–887.
Brazdil, P., Giraud-Carrier, C., Issue, Special, 2018. Metalearning and Algorithm Selection: progress, state of the art and introduction to the 2018. Mach. Learn. 107, 1–14.
Brazdil, P., Giraud-Carrier, C., Soares, C., Vilalta, R., 2009. Metalearning. Applications to Data Mining. Springer-Verlag, Berlin, Heidelberg.
Brazdil, P., Vilalta, R., Giraud-Carrier, C., Soares, C., 2017. Metalearning. In: Sammut, C., Webb, G.I. (Eds.), Encyclopedia of Machine Learning and Data Mining, second ed. Springer Science+Business Media, New York, pp. 818–823.
Brazdil, P.B., Soares, C., 2000. A comparison of ranking methods for classification algorithm selection. In: López de Mántaras, R., Plaza, E. (Eds.), Machine Learning: ECML 2000. ECML 2000. Lecture Notes in Computer Science (Lecture Notes in Artificial Intelligence), vol. 1810. Springer, Berlin, Heidelberg, pp. 63–75.
Brazdil, P.B., Soares, C., Da Costa, J.P., 2003. Ranking learning algorithms: using IBL and meta-learning on accuracy and time results. Mach. Learn. 50, 251–277.
Brocal, F., Paltrinieri, N., González-Gaya, C., Sebastián, M.A., Reniers, G., 2021. Approach to the selection of strategies for emerging risk management considering uncertainty as the main decision variable in industrial contexts. Saf. Sci. 134, 105041.
Bucelli, M., Utne, I.B., Rossi, P.S., Paltrinieri, N., 2020. A system engineering approach to subsea spill risk management. Saf. Sci. 123, 104560.
Castiello C., Castellano G., Fanelli A.M., 2005. Meta-data: Characterization of Input Features for Meta-learning. In: Torra V., Narukawa Y., Miyamoto S. (Eds.), Modeling Decisions for Artificial Intelligence. MDAI 2005. Lecture Notes in Computer Science, vol. 3558. Springer, Berlin, Heidelberg, pp. 457–468.
Chang, J.C., Hanna, S.R., 2004. Air quality model performance evaluation. Meteorol. Atmos. Phys. 87, 167–196.
Chen, Y.-C., Ma, H.-W., 2007. Combining the cost of reducing uncertainty with the selection of risk assessment models for remediation decision of site contamination. J. Hazard. Mater. 141, 17–26.
Cohen-Shapira N., Rokach L., Shapira B., Katz G., Vainshtein R., 2019. AutoGRD: model recommendation through graphical dataset representation. In: CIKM '19: Proceedings of the 28th ACM International Conference on Information and Knowledge Management. ACM, pp. 821–830.
Cui, C., Hu, M., Weir, J.D., Wu, T., 2016a. A recommendation system for meta-modeling: a meta-learning based approach. Expert Syst. Appl. 46, 33–44.
Cui, C., Wu, T., Hu, M., Weir, J.D., Li, X., 2016b. Short-term building energy model recommendation system: a meta-learning approach. Appl. Energy 172, 251–263.
Cullen, A.C., Frey, H.C., 1999. Probabilistic Techniques in Exposure Assessment. A Handbook for Dealing with Variability and Uncertainty in Models and Inputs. Plenum Press, New York.
Cunha, T., Soares, C., de Carvalho, A.C.P.L.F., 2018. Metalearning and Recommender Systems: a literature review and empirical study on the algorithm selection problem for Collaborative Filtering. Inf. Sci. 423, 128–144.
das Dôres S.N., Alves L., Ruiz D.D., Barros R.C., 2016. A meta-learning framework for algorithm recommendation in software fault prediction. In: SAC '16: Proceedings of the 31st Annual ACM Symposium on Applied Computing. ACM, pp. 1486–1491.
de Rocquigny, E., 2009. Quantifying uncertainty in an industrial approach: an emerging consensus in an old epistemological debate. S.A.P.I.EN.S 2 (1), 1–18.
de Rocquigny, E., Devictor, N., Tarantola, S. (Eds.), 2008. Uncertainty in Industrial Practice. A guide to Quantitative Uncertainty Management. John Wiley & Sons Ltd, West Sussex, England.
de Souto, M.C.P., Prudêncio, R.B.C., Soares, R.G.F., de Araujo, D.S.A., Costa, I.G., Ludermir, T.B., Schliep, A., 2008. In: Ranking and Selecting Clustering Algorithms Using a Meta-Learning Approach. IEEE, pp. 3729–3735.
dos Santos P.M., Ludermir T.B., Prudêncio R.B.C., 2004. Selection of Time Series Forecasting Models based on Performance Information. In: Fourth International Conference on Hybrid Intelligent Systems (HIS'04). IEEE, pp. 366–371.
Dyrmishi, S., Elshawi, R., Sakr, S., 2019. A decision support framework for AutoML systems: a meta-learning approach. In: 2019 International Conference on Data Mining Workshops (ICDMW). IEEE, pp. 97–106.
EIGA (European Industrial Gases Association), 2018. Hazards of oxygen-deficient atmospheres. Doc 44/18. EIGA, Brussels.
Ferrari, D.G., de Castro, L.N., 2015. Clustering algorithm selection by meta-learning systems: a new distance-based problem characterization and ranking combination methods. Inf. Sci. 301, 181–194.
Filchenkov, A., Pendryak, A., 2015. Datasets Meta-Feature Description for Recommending Feature Selection Algorithm. In: 2015 Artificial Intelligence and Natural Language and Information Extraction, Social Media and Web Search FRUCT Conference (AINL-ISMW FRUCT). IEEE, pp. 11–18.
Giraud-Carrier, C., Vilalta, R., Brazdil, P., 2004. Introduction to the special issue on meta-learning. Mach. Learn. 54, 187–193.
IAEA (International Atomic Energy Agency), 1989. Evaluating the Reliability of Predictions Made using Environmental Transfer Models. Safety Series No. 100. IAEA, Vienna, Austria.
IPCS (International Programme on Chemical Safety), 2008. Uncertainty and Data Quality in Exposure Assessment. Harmonization Project Document No. 6. World Health Organization, Geneva, Switzerland.
Ivings, M.J., Gant, S.E., Jagger, S.F., Lea, C.J., Stewart, J.R., Webber, D.M., 2016. Evaluating Vapor Dispersion Models for Safety Analysis of LNG Facilities, 2nd ed. NFPA. Health & Safety Laboratory, Buxton, Derbyshire, UK.
Kalousis, A., Hilario, M., 2001. Model selection via meta-learning: a comparative study. Int. J. Artif. Intell. Tools 10 (4), 525–554.
Kanda, J., de Carvalho, A., Hruschka, E., Soares, C., Brazdil, P., 2016. Meta-learning to select the best meta-heuristic for the Traveling Salesman Problem: a comparison of meta-features. Neurocomputing 205, 393–406.
Khan, I., Zhang, X., Rehman, M., Ali, R., 2020. A literature survey and empirical study of meta-learning for classifier selection. IEEE Access 8, 10262–10281.
Kozielski M., 2016. A meta-learning approach to methane concentration value prediction. In: Kozielski S., Mrozek D., Kasprowski P., Małysiak-Mrozek B., Kostrzewa D. (Eds.), Beyond Databases, Architectures and Structures. Advanced Technologies for Data Mining and Knowledge Discovery. BDAS 2015, BDAS 2016. Communications in Computer and Information Science, vol. 613. Springer, Cham., pp. 716–726.
Kozielski M., Łaskarzewski Z., 2019. Matching a model to a user – application of meta-learning to LPG consumption prediction. In: Xhafa F., Barolli L., Greguš M. (Eds.), Advances in Intelligent Networking and Collaborative Systems. INCoS 2018. Lecture Notes on Data Engineering and Communications Technologies, vol. 23. Springer, Cham., pp. 495–503.
Kück M., Crone S.F., Freitag M., 2016. Meta-learning with neural networks and landmarking for forecasting model selection. An empirical evaluation of different feature sets applied to industry data. In: 2016 International Joint Conference on Neural Networks (IJCNN). IEEE, pp. 1499–1506.
Lee, S., Landucci, G., Reniers, G., Paltrinieri, N., 2019. Validation of dynamic risk analysis supporting integrated operations across systems. Sustainability 11 (23), 6745. https://doi.org/10.3390/su11236745.
Lemke, C., Budka, M., Gabrys, B., 2015. Metalearning: a survey of trends and technologies. Artif. Intell. Rev. 44, 117–130.
Ler D., Teng H., He Y., Gidijala R., 2018. Algorithm Selection for Classification Problems via Cluster-based Meta-features. In: 2018 IEEE International Conference on Big Data (Big Data). IEEE, pp. 4952–4960.
Loucks D.P., 2002. Quantifying and Communicating Model Uncertainty for Decisionmaking in the Everglades. Risk-Based Decisionmaking in Water Resources X, pp. 40–58.
Loucks, D.P., Van Beek, E., Stedinger, J.R., Dijkman, J.P.M., Villars, M.T., 2005. Model sensitivity and uncertainty analysis. Water Resourc. Syst. Plan. Manage. 255–290.
Makmal, A., Melnikov, A.A., Dunjko, V., Briegel, H.J., 2016. Meta-learning within projective simulation. IEEE Access 4, 2110–2122.
McManus, N., Haddad, A.N., 2015. Oxygen levels during welding: assessment in an aluminum shipbuilding environment. Professional Safety 60 (7), 26–32.
Morgan M.G., Henrion M., 1990. Uncertainty. A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis. Cambridge University Press, Cambridge, United Kingdom.
Muñoz M.A., Kirley M., Halgamuge S.K., 2013. The Algorithm Selection Problem on the Continuous Optimization Domain. In: Moewes C., Nürnberger A. (Eds.), Computational Intelligence in Intelligent Data Analysis. Studies in Computational Intelligence, vol. 445. Springer, Berlin, Heidelberg, pp. 75–89.
Nilsen, T., Aven, T., 2003. Models and model uncertainty in the context of risk analysis. Reliab. Eng. Syst. Saf. 79 (3), 309–317.
Paltrinieri, N., Bonvicini, S., Spadoni, G., Cozzani, V., 2012. Cost-benefit analysis of passive fire protections in road LPG transportation. Risk Anal. 32 (2), 200–219. https://doi.org/10.1111/j.1539-6924.2011.01654.x.
Paltrinieri, N., Comfort, L., Reniers, G., 2019. Learning about risk: machine learning for risk assessment. Saf. Sci. 118, 475–486. https://doi.org/10.1016/j.ssci.2019.06.001.
Paltrinieri, N., Khan, F. (Eds.), 2016. Dynamic Risk Analysis in the Chemical and Petroleum Industry. Evolution and Interaction with Parallel Disciplines in the Perspective of Industrial Application. Butterworth-Heinemann, Oxford, United Kingdom.
Paltrinieri, N., Patriarca, R., Stefana, E., Brocal, F., Reniers, G., 2020. Meta-learning for safety management. Chem. Eng. Trans. 83 https://doi.org/10.3303/CET2082029.
Pfahringer B., Bensusan H., Giraud-Carrier C., 2000. Meta-Learning by Landmarking Various Learning Algorithms. In: Proceedings of the 17th International Conference on Machine Learning (ICML'00), pp. 743–750.
Patriarca, R., Di Gravio, G., Cioponea, R., Licu, A., 2019. Safety intelligence: Incremental proactive risk management for holistic aviation safety performance. Saf. Sci. 118, 551–567. https://doi.org/10.1016/j.ssci.2019.05.040.
Patriarca, R., Falegnami, A., Costantino, F., Di Gravio, G., De Nicola, A., Villani, M.L., 2021. WAx: An integrated conceptual framework for the analysis of cyber-socio-technical systems. Saf. Sci. 136, 105142 https://doi.org/10.1016/j.ssci.2020.105142.
Pimentel, B.A., de Carvalho, A.C.P.L.F., 2019a. A new data characterization for selecting clustering algorithms using meta-learning. Inf. Sci. 477, 203–219.
Pimentel B.A., de Carvalho A.C.P.L.F., 2019b. Unsupervised meta-learning for clustering algorithm recommendation. In: 2019 International Joint Conference on Neural Networks (IJCNN). IEEE, pp. 1–8.
Pinto F., Soares C., Mendes-Moreira J., 2014. A Framework to Decompose and Develop Metafeatures. In: Vanschoren J., Brazdil P., Soares C., Kotthoff L. (Eds.), Meta-Learning and Algorithm Selection Workshop at ECAI 2014. MetaSel 2014, pp. 32–36.

Pinto F., Soares C., Mendes-Moreira J., 2016. Towards Automatic Generation of Metafeatures. In: Bailey J., Khan L., Washio T., Dobbie G., Huang J., Wang R. (Eds.), Advances in Knowledge Discovery and Data Mining. PAKDD 2016. Lecture Notes in Computer Science, vol. 9651. Springer, Cham., pp. 215–226.
Pise, N., Kulkarni, P., 2016. Algorithm selection for classification problems. In: 2016 SAI Computing Conference (SAI). IEEE, pp. 203–211.
Prudêncio R.B.C., de Souto M.C.P., Ludermir T.B., 2011a. Selecting Machine Learning Algorithms Using the Ranking Meta-Learning Approach. In: Jankowski, N., Duch, W., Grąbczewski, K. (Eds.), Meta-Learning in Computational Intelligence. Studies in Computational Intelligence, vol. 358. Springer, Berlin, Heidelberg, pp. 225–243.
Prudêncio, R.B.C., Ludermir, T.B., 2004. Meta-learning approaches to selecting time series models. Neurocomputing 61, 121–137.
Prudêncio, R.B.C., Ludermir, T.B., 2012. Combining Uncertainty Sampling methods for supporting the generation of meta-examples. Inf. Sci. 196, 1–14.
Prudêncio R.B.C., Soares C., Ludermir T.B., 2011b. Combining Meta-learning and Active Selection of Datasetoids for Algorithm Selection. In: Corchado E., Kurzyński M., Woźniak M. (Eds.), Hybrid Artificial Intelligent Systems. HAIS 2011. Lecture Notes in Computer Science, vol. 6678. Springer, Berlin, Heidelberg, pp. 164–171.
Prudêncio, R.B.C., Soares, C., Ludermir, T.B., 2011c. Uncertainty sampling methods for selecting datasets in active meta-learning. In: The 2011 International Joint Conference on Neural Networks. IEEE, pp. 1082–1089.
Prudêncio, R.B.C., Soares, C., Ludermir, T.B., 2011d. Uncertainty Sampling-Based Active Selection of Datasetoids for Meta-learning. In: Honkela, T., Duch, W., Girolami, M., Kaski, S. (Eds.), Artificial Neural Networks and Machine Learning – ICANN 2011. ICANN 2011. Lecture Notes in Computer Science, vol. 6792. Springer, Berlin, Heidelberg, pp. 454–461.
Rao, K.S., 2005. Uncertainty analysis in atmospheric dispersion modeling. Pure Appl. Geophys. 162, 1893–1917.
Reif, M., 2012. A comprehensive dataset for evaluating approaches of various meta-learning tasks. In: Proceedings of the 1st International Conference on Pattern Recognition and Methods (ICPRAM), pp. 273–276.
Reif, M., Shafait, F., Dengel, A., 2012. Dataset generation for meta-learning. In: Wölfl, S. (Eds.), Poster and Demo Track of the 35th German Conference on Artificial Intelligence (KI-2012), pp. 69–73.
Ren, Y., Chi, C., Jintao, Z., 2020. A survey of personalized recommendation algorithm selection based on meta-learning. In: Xu, Z., Choo, K.-K.R., Dehghantanha, A., Parizi, R., Hammoudeh, M. (Eds.), Cyber Security Intelligence and Analytics. CSIA 2019. Advances in Intelligent Systems and Computing, vol. 928. Springer, Cham., pp. 1383–1388.
Rice, J.R., 1976. The algorithm selection problem. Adv. Comput. 15, 65–118.
Romero, C., Olmo, J.L., Ventura, S., 2013. A meta-learning approach for recommending a subset of white-box classification algorithms for Moodle datasets. In: D'Mello, S., Calvo, R., Olney, A. (Eds.), Proceedings of International Conference on Educational Data Mining (EDM), pp. 268–271.
Rossi, A.L.D., Carvalho, A.C.P.L.F., Soares, C., 2012. Meta-learning for periodic algorithm selection in time-changing data. In: 2012 Brazilian Symposium on Neural Networks. IEEE, pp. 7–12.
Rossi, A.L.D., de Leon Ferreira, A.C.P., Soares, C., de Souza, B.F., 2014. MetaStream: a meta-learning based method for periodic algorithm selection in time-changing data. Neurocomputing 127, 52–64.
Rossi, A.L.D., de Souza, B.F., Soares, C., de Carvalho, A.C.P.L.F., 2017. A guidance of data stream characterization for meta-learning. Intell. Data Anal. 21, 1015–1035.
Saltelli, A., Tarantola, S., Campolongo, F., Ratto, M., 2004. Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models. John Wiley & Sons Ltd, Chichester, England.
Santos, E. Jr., Kilpatrick, A., Nguyen, H., Gu, Q., Grooms, A., Poulin, C., 2012. Flexible algorithm selection framework for large scale metalearning. In: 2012 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology. IEEE, pp. 496–503.
Seligmann, B.J., Zhao, J., Marmara, S.G., Corbett, T.C., Small, M., Hassall, M., Boadle, J. T., 2019. Comparing capability of scenario hazard identification methods by the PIC (Plant-People-Procedure Interaction Contribution) network metric. Saf. Sci. 112, 116–129.
Shahoud, S., Khalloof, H., Duepmeier, C., Hagenmeyer, V., 2020. Descriptive statistics time-based meta features (DSTMF) constructing a better set of meta features for model selection in energy time series forecasting. In: APPIS 2020: Proceedings of the 3rd International Conference on Applications of Intelligent Systems, pp. 1–6.
Smith-Miles, K.A., 2008a. Cross-disciplinary perspectives on meta-learning for algorithm selection. ACM Comput. Surveys 41 (1), 6:1–6:25.
Smith-Miles, K.A., 2008b. Towards insightful algorithm selection for optimisation using meta-learning concepts. In: 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence). IEEE, pp. 4118–4124.
Soares C., Brazdil P., 2002. A comparative study of some issues concerning algorithm recommendation using ranking methods. In: Garijo, F.J., Riquelme, J.C., Toro, M. (Eds.), Advances in Artificial Intelligence – IBERAMIA 2002. IBERAMIA 2002. Lecture Notes in Computer Science, vol. 2527. Springer, Berlin, Heidelberg, pp. 80–89.
Soares, C., Brazdil, P.B., 2000. Zoomed ranking: selection of classification algorithms based on relevant performance information. In: Zighed, D.A., Komorowski, J., Żytkow, J. (Eds.), Principles of Data Mining and Knowledge Discovery. PKDD 2000. Lecture Notes in Computer Science, vol. 1910. Springer, Berlin, Heidelberg, pp. 126–135.
Soares, R.G.F., Ludermir, T.B., De Carvalho, F.A.T., 2009. An analysis of meta-learning techniques for ranking clustering algorithms applied to artificial data. In: Alippi, C., Polycarpou, M., Panayiotou, C., Ellinas, G. (Eds.), Artificial Neural Networks – ICANN 2009. ICANN 2009. Lecture Notes in Computer Science, vol. 5768. Springer, Berlin, Heidelberg, pp. 131–140.
Sousa, A.F.M., Prudêncio, R.B.C., Ludermir, T.B., Soares, C., 2016. Active learning and data manipulation techniques for generating training examples in meta-learning. Neurocomputing 194, 45–55.
Stefana, E., Marciano, F., Alberti, M., 2016. A predictive model for estimating the indoor oxygen level and assessing Oxygen Deficiency Hazard (ODH). J. Loss Prev. Process Ind. 39, 152–172. https://doi.org/10.1016/j.jlp.2015.11.022.
Stefana, E., Marciano, F., Cocca, P., 2019a. Uncertainty and sensitivity analyses of models for assessing oxygen deficiency hazard: preliminary results. In: Beer, M., Zio, E. (Eds.), Proceedings of the 29th European Safety and Reliability Conference. European Safety and Reliability Association, Research Publishing, Singapore, pp. 2761–2767.
Stefana, E., Marciano, F., Cocca, P., Alberti, M., 2015. Predictive models to assess Oxygen Deficiency Hazard (ODH): a systematic review. Saf. Sci. 75, 1–14. https://doi.org/10.1016/j.ssci.2015.01.008.
Stefana, E., Marciano, F., Cocca, P., Alberti, M., 2017. A near field-far field model for assessing oxygen deficiency hazard. Process Saf. Environ. Prot. 105, 201–216. https://doi.org/10.1016/j.psep.2016.11.006.
Stefana, E., Marciano, F., Cocca, P., Rossi, D., Tomasoni, G., 2019b. Oxygen deficiency hazard in confined spaces in the steel industry: assessment through predictive models. Int. J. Occupat. Safety Ergon. https://doi.org/10.1080/10803548.2019.1669954.
Stefana, E., Marciano, F., Drolet, D., Armstrong, T.W., 2021. A traditional Near Field-Far Field approach-based model and a spreadsheet workbook to manage Oxygen Deficiency Hazard. Process Saf. Environ. Prot. 149, 537–556. https://doi.org/10.1016/j.psep.2020.11.014.
Stefana, E., Paltrinieri, N., 2020. Meta-learning potential to assess uncertainties in dynamic risk management. In: Baraldi, P., Di Maio, F., Zio, E. (Eds.), Proceedings of the 30th European Safety and Reliability Conference and the 15th Probabilistic Safety Assessment and Management Conference. Research Publishing, Singapore.
Thompson, M., Warmink, J.J., 2017. Natural hazard modeling and uncertainty analysis. In: Riley, K., Webley, P., Thompson, M. (Eds.), Natural Hazard Uncertainty Assessment: Modeling and Decision Support. American Geophysical Union, Washington, D.C., USA, pp. 11–20.
Tripathy, M., Panda, A., 2017. A study of algorithm selection in data mining using meta-learning. J. Eng. Sci. Technol. Rev. 10 (2), 51–64.
Vanschoren, J., 2018. Meta-Learning: A Survey. arXiv preprint arXiv:1810.03548.
Vanschoren, J., 2019. Meta-Learning. In: Hutter F., Kotthoff L., Vanschoren J. (Eds.), Automated Machine Learning. Methods, Systems, Challenges. Springer, Cham, Switzerland, pp. 35–61.
Verma, A.K., Srividya, A., Karanki, D.R., 2010. Uncertainty management in reliability/safety assessment. In: Verma, A.K., Ajit, S., Karanki, D.R., Reliability and Safety Engineering. Springer, London, pp. 435–522.
Vilalta, R., Drissi, Y., 2002. A perspective view and survey of meta-learning. Artif. Intell. Rev. 18, 77–95.
Vilalta, R., Giraud-Carrier, C., Brazdil, P., 2009. Meta-learning – concepts and techniques. In: Maimon, O., Rokach, L. (Eds.), Data Mining and Knowledge Discovery Handbook. Springer, Boston, MA, pp. 717–731.
Vilalta, R., Giraud-Carrier, C., Brazdil, P., Soares, C., 2004. Using meta-learning to support data mining. Int. J. Comput. Sci. Appl. 1 (1), 31–45.
Wolpert, D.H., 1996. The lack of a priori distinctions between learning algorithms. Neural Comput. 8 (7), 1341–1390.
Wolpert, D.H., Macready, W.G., 1995. No Free Lunch Theorems for Search. Technical Report SFI-TR-95-02-010, Santa Fe Institute, Santa Fe.
Wolpert, D.H., Macready, W.G., 1997. No free lunch theorems for optimization. IEEE Trans. Evol. Comput. 1 (1), 67–82.
Yegnan, A., Williamson, D.G., Graettinger, A.J., 2002. Uncertainty analysis in air dispersion modeling. Environ. Modell. Softw. 17, 639–649.
Zhu, X., Yang, X., Ying, C., Wang, G., 2018. A new classification algorithm recommendation method based on link prediction. Knowl.-Based Syst. 159, 171–185.
Zio, E., Aven, T., 2013. Model output uncertainty in risk assessment. Int. J. Performab. Eng. 29 (5), 475–486.
Zio, E., Pedroni, N., 2012. Uncertainty Characterization in Risk Analysis for Decision-making Practice. FonCSI (Fondation pour une culture de sécurité industrielle), Toulouse, France.
Zorrilla, M., García-Saiz, D., 2014. Meta-learning: can it be suitable to automatise the KDD process for the educational domain? In: Kryszkiewicz, M., Cornelis, C., Ciucci, D., Medina-Moreno, J., Motoda, H., Raś, Z.W. (Eds.), Rough Sets and Intelligent Systems Paradigms. Lecture Notes in Computer Science, vol. 8537. Springer, Cham., pp. 285–292.
Zorrilla, M., García-Saiz, D., 2015. Meta-learning based framework for helping non-expert miners to choice a suitable classification algorithm: an application for the educational field. In: Núñez, M., Nguyen, N., Camacho, D., Trawiński, B. (Eds.), Computational Collective Intelligence. Lecture Notes in Computer Science, vol. 9330. Springer, Cham., pp. 431–440.