IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received May 15, 2021, accepted June 17, 2021, date of publication June 21, 2021, date of current version June 29, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3090936

Novel Meta-Features for Automated Machine Learning Model Selection in Anomaly Detection

MILOŠ KOTLAR[1], MARIJA PUNT[1], ZAHARIJE RADIVOJEVIĆ[1], MILOŠ CVETANOVIĆ[1], AND VELJKO MILUTINOVIĆ[2]

[1]School of Electrical Engineering, University of Belgrade, 11000 Belgrade, Serbia
[2]Department of Computer Science, Indiana University Bloomington, Bloomington, IN 47405, USA

Corresponding author: Miloš Kotlar (km175003p@student.etf.bg.ac.rs)

This work was supported in part by the Ministry of Education, Science, and Technological Development of the Republic of Serbia under Grant 44009, and in part by the Science Fund of the Republic of Serbia [AVANTES].

ABSTRACT A growing number of research papers shed light on automated machine learning (AutoML) frameworks, which are becoming a promising solution for building complex machine learning models without human expertise and assistance. The key challenge in enabling AutoML frameworks to build an efficient model for anomaly detection tasks is to determine the best underlying model for a given task and optimization metric. The meta-learning approaches based on a set of meta-features that describes data properties can enable efficient model selection in AutoML frameworks. The existing meta-learning approaches based on statistical and information-theoretic meta-features require large amounts of data and computational resources to extract data properties. This paper proposes a novel set of meta-features for model selection in anomaly detection tasks based on domain-specific properties of data which overcomes the shortcomings of existing meta-features by introducing simple but effective meta-features that can be efficiently extracted or estimated by using a low amount of data. Experiments with 63 datasets from different repositories with varying schemas show that the proposed set of meta-features achieves an accuracy of 87% for model selection, while the achieved accuracy for simple meta-features is 74%, for statistical meta-features 68%, for information theory meta-feature 70%, and for a comprehensive set of meta-features by pyMFE 73%. This demonstrates that the proposed set can be adopted by AutoML frameworks across a diverse range of domains.

INDEX TERMS Anomaly detection, AutoML, data properties, distance functions, meta-features, meta-learning, transfer learning.

I. INTRODUCTION

Anomaly detection is an important machine learning problem studied within diverse research areas [1]–[4]. It has an enormous applicability that includes almost any domain, especially for the purpose of quality monitoring [5], [6].

In practical use, anomalies are diverse: in each domain they have different properties, which makes anomaly detection a challenging task that requires an effective model. For example, manufacturing systems are equipped with an increasing number of sensors to monitor the process where the failure of a device or a sensor often brings about latency or downtime of the system. One of the requirements may be to reduce failure of a sensor device in such a system by predicting anomalies in sensor data and thus decreasing downtime in the system [7], [8]. In transportation networks, it may be required to detect traffic jams in overcrowded places, which represents unusual behavior. Papers [9], [10] addressed approaches for detecting traffic jams by introducing innovative techniques for analyzing spatial and temporal data of transportation networks and detecting unusual behaviors. The most prominent cases of anomaly detection use are fraud detection in finance systems and cyber security intrusion detection [11], [12], which have a huge impact on the confidential aspects of data. In the field of healthcare, monitoring systems equipped with different types of diagnostic tools and sensors rely on anomaly detection models suitable for predictive modeling that record the condition of a patient and predict contextual abnormal behavior early in temporal data [13]. In data mining techniques, removing anomalous data instances such as errors

The associate editor coordinating the review of this manuscript and approving it for publication was Ioannis Schizas.

significantly improves the performance of a model for a given optimization metric [14].

With such use cases, the proliferation of data and devices [15], [16] makes data quality an essential key performance indicator of a particular use case where anomaly detection model performance, or machine learning models in general, strongly depends on human expertise, data properties, and hyperparameters [17]. AutoML frameworks have since recently been used in scenarios which allow non-expert users to make use of models[1] without requiring prior knowledge. In order to enable AutoML frameworks to build an efficient model for anomaly detection, the key challenge is to determine the best underlying model for a given task and optimization metric. For example, linear regression may achieve significant results for anomaly detection tasks in temporal data due to linear data correlation but could fail for high-dimensional data with linear independence [18]. AutoML frameworks contain several components including data preparation, feature engineering, model selection, and application [19], [20], as shown in Fig. 1. Model selection utilizes the meta-learning approach by extracting meta-features that could be predictive for algorithm performance using prior performance knowledge of data with similar properties [21].

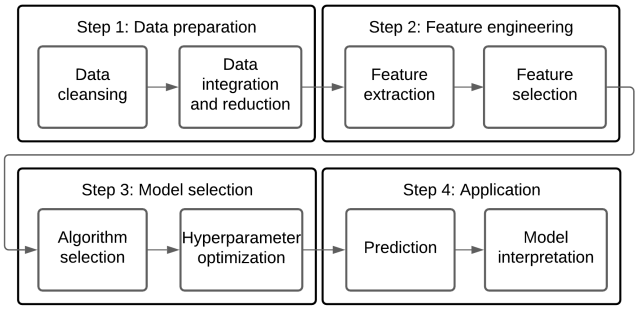

FIGURE 1. The main components of AutoML frameworks are data preparation, feature engineering, model selection, and application. This paper proposes a set of meta-features used for algorithm selection in the model selection component only.

The main contribution of this paper is a novel set of domain-specific meta-features for model selection in anomaly detection tasks. In order to validate the proposed set of domain-specific meta-features, the following has been done:

- Existing solutions were analysed and compared through different complexity aspects.
- A public repository has been created containing an open source implementation enabling full reproducibility of the results (open-source implementation of anomaly detection algorithms, distance functions, and meta-feature extractor).
- A dataset repository has been created as a union of 2 different dataset repositories that contains 63 industry-based with labeled anomalies available for benchmarking.
- Performance and robustness experiments of the proposed meta-features has been designed and conducted.

The existing solutions are based on statistical and information theory methods which implies an enormous amount of mathematical computations in order to generate the appropriate vector of values relevant for the model selection. The proposed solution creates the vector of values for the model selection without much mathematical computation, and only with some logical computation, which reduces tremendously the overall computational costs. With such overall computational costs, a data provider or a domain expert can describe anomalies by estimating the proposed meta-features in scenarios where labeled data is not available.

The presentation and research methodology of this paper is based on [35]. Section II presents overview of the existing meta-learning approaches and compares them by type of computation, whether they are used in anomaly detection so far, level of information required, and amount of data needed for computation. Section III analyzes problems and introduces research questions related to the possibility to define a set of meta-features which would effectively represent the properties of anomalies in data. Section IV defines a novel set of domain-specific meta-features for anomaly detection and gives their details. Section V presents an architecture design for the model selection system used in experiments and discusses different groups of algorithms evaluated in this paper. Section VI describes experiments and obtained results which answer the defined research questions by evaluating the performance of the proposed set of meta-features. Section VII concludes the paper, providing a summary of advantages and drawbacks of the proposed set of meta-features. It also demonstrates why this research is important, who will benefit from it, and outlines new paths for further research.

II. RELATED WORK

Meta-learning is a novel approach for model selection used in AutoML frameworks and decision support systems [22]. The meta-learning approach based on extracting meta-features from data often employs statistical and information theory methods. Based on the amount of data needed, computational resources required, and methods used for computation, meta-features can be divided into the following groups: simple, statistical, information theory, domain-specific, and model-based. A special type of meta-learning approach which is not based on meta-features but rather on simple algorithms for estimating performance of more complex algorithms is called landmarking [23]. Table 1 summarizes different types of meta-learning based on the level of information needed for meta-feature extraction.

Paper [23] provides a comprehensive overview of meta-features and performs an in-depth analysis with a tool for their extraction. Package from this paper (pyMFE) presents an industry adopted meta-features extractor based on simple, statistical, and information theory methods. It brings

[1] In this paper, the words model and algorithm are used synonymously and in the same manner.

TABLE 1. Overview of meta-learning approaches by type, whether they are used in anomaly detection domain so far, level of information required, and amount of data needed for computation. It is important to note that domain-specific meta-features are not used in anomaly detection domain, prior to this paper. The level of information needed for computation is divided into the following categories from the smallest to the highest one, whereby each category implicitly includes the previous categories: (I) only domain knowledge is required (doesn't exist in anomaly detection domain prior to this paper), (II) size of data is required, (III) column types are required, (IV) data distribution is required, (V) part of data is required, and (VI) complete data is required.

Meta-features	Extracted from	Used in anomaly detection domain	Level of information required	Amount of data	Computational complexity	Use cases
Simple	data	yes	III - column types	low	low	[22], [23]
Statistical	data	yes	IV - data distribution	significant	significant	[23]–[28]
Information theory	data	yes	V - part of data	significant	significant	[29], [30], [30]
Domain-specific	data	no	II - size of data	low	low	[25], [31]
Model-based	model	no	VI - complete data	significant	significant	[23], [32], [33]
Landmarking	model	yes	V - part of data	significant	low	[34]

cutting edge meta-features, a topic that was proposed in recent literature. This package is later used in the experiments for comparing the proposed solution against it.

Simple meta-features are based on mathematical computation and contain basic data properties that are directly extracted from data with low computational costs using methods with low complexity. They are also referred to as general meta-features and include basic dataset properties such as the number of attributes, number of columns, number of categorical or numerical attributes, etc. [22].

Statistical meta-features are based on mathematical computation contain statistical properties of data that are extracted from data with significant computational costs using statistical methods with significant complexity. Statistical methods often require hyperparameters and they are based on numerical attributes only. They constitute the largest and the most diversified group of meta-features which are extracted separately for attributes and include data distribution properties, average, standard deviation, correlation and covariance, min, max, mean, sparsity, and similar [23]. The number of existing solutions utilize statistical meta-features due to attribute diversity that can be applied across a range of different domains [24]–[28].

Information theory meta-features are based on mathematical computation and represent data complexity and contain the level of information embedded in data. They are directly extracted from data with significant computational costs using methods with significant complexity but without hyperparameters, requiring significant amounts of data. They are based on discrete attributes and include properties such as entropy which captures the amount of information and complexity of data, mutual information which mostly determines the relation of attributes and target class used for classification problems [25]. They are used for presenting different behavior patterns [29], for performing high quality recommendations, and for representing inner correlations between different classes [30].

Domain-specific meta-features are based on logical computation and contain domain-based knowledge directly extracted from data with low computational costs using methods with low complexity. If data is not available, they can be estimated by a data provider or a domain expert without computational costs, requiring only domain-based knowledge. So far, domain specific meta-features are not used in the anomaly detection domain. They are only used in text classification, where domain-based knowledge present vocabulary length, words overlap, number of text categories, corpus hardness, domain broadness, and similar [25], [31].

Model-based meta-features contain properties that describe the model; they are extracted by applying predictive learning algorithms, namely, decision trees and clustering algorithms. They are only applicable to supervised problems where all measures are deterministic and require hyperparameters with high computational costs. They are extracted from model properties which require significant computational costs and complexity. For example, using a decision tree, one can extract properties such as the number of leaves, number of nodes, depth and width of the tree, etc. [23]. Papers [32], [33] analyze the meta-learning approach for determining the number of clusters in data by proposing quality metrics meta-features to describe the structure of data.

Landmarking is a special case of meta-learning that describes data using the performance of simple and fast learners. The result of meta-learning is not extracting meta-features from data but rather predicting which algorithm will provide the best results for a given dataset by running simple learners with significant amounts of data and high computational costs. Simple learners are often based on classification and clustering algorithms, such as eliteNN, where the algorithm is based on the 1NN model with the most informative attributes of a dataset [34].

Along with the well-known meta-learning approaches there are novel approaches based on techniques like morphing [36] which transforms data and observes changes in behavior of learning algorithms. Meta-learning is also used for other data interoperability tasks such as feature selection [37], [38]. The meta-features presented in this paper are extracted from data and are thus compared against the same group of meta-features, which means that experiments do not evaluate meta-features based on models (e.g. model-based meta-features or landmarking).

III. ANOMALY DETECTION USING META-FEATURES

In AutoML frameworks with strict latency and runtime demands, model selection is the key challenge. The goal is to determine which algorithm will provide the best results

for a given task and optimization metric. Many papers address this challenge by proposing different approaches for model selection [39], including brute-force and meta-learning approaches. The brute-force approach requires large amounts of computational resources to obtain the result which does not fit into AutoML framework requirements and is not suitable for anomaly detection scenarios which rely on streaming data pipelines. Meta-learning approaches are based on less compute-intensive methods compared to the brute-force approach. However, existing meta-features for model selection in anomaly detection still have several unresolved issues. Those will be addressed in the following sections, where conditions of interest to be fulfilled are defined and the research questions are formalized.

A. PROBLEM STATEMENT

Anomaly detection ensures data quality and constitutes an important task in scenarios that are often related to processing and analyzing heterogeneous data streams and detecting different patterns in a near real-time window [40]. Such scenarios have strict demands and thus eliminate solutions that rely on complex operations and require significant computational resources. Model selection based on statistical and information theory meta-features are compute-intensive and require significant amounts of data and computational resources to extract meta-features. On the opposite side, model selection based on simple meta-features can efficiently extract data properties with low amounts of data, but they often do not provide enough information to achieve satisfactory results [41]. Besides performance-related issues, existing solutions predominantly focus on frameworks for classification or text processing and do not provide solutions for anomaly detection tasks which can be represented as a special case of classification with highly imbalanced classes.

B. PROBLEM ANALYSIS

Model selection depends on algorithm selection, hyperparameter optimization and data preprocessing, like reduction and integration, which are often heuristically found by the systematic grid, a random search, or by applying machine learning [42], and this paper addresses algorithm selection only. Using logical domain-specific meta-features would make it possible to efficiently provide results for a given task and budget by minimizing computational resources and amount of data required for model selection in AutoML frameworks, which would speed up the learning process and improve overall performance. Such a set of meta-features should meet the following requirements in order to match demands of AutoML frameworks:

- Schema agnosticism: Meta-features should be able to efficiently describe properties for data from different domains, containing different attribute types, having different types of anomalies, and different anomaly spaces.
- Scalability: Meta-features should be able to utilize simple methods for their extraction without complex and compute-intensive operations.
- Relation: Meta-features should be able to efficiently describe the relations between data and achieve significant performance for different optimization metrics.
- Simplicity: Meta-features should be efficiently estimated by a data provider or a domain expert, which is crucial when anomalies in data are not labeled.

The existing solutions based on simple, statistical, and information theory meta-features do not meet one or more above-mentioned defined requirements. Simple meta-features meet all complexity-based requirements except the ability to be efficiently estimated by a data provider or a domain expert. Statistical and information theory meta-features implement significant correlation in data but require significant amounts of computational resources and thus do not meet the scalability requirement. The importance of above-mentioned requirements will over the time grow together with data volume caused by the rapid development of IoT and WSN. In such environments, the need for significant data quality is immense [43].

C. RESEARCH QUESTIONS

Research questions addressed in this paper are the following:

1) Is it possible to define domain-specific set of meta-features which would effectively represent the properties of anomalies in data?
2) Whether domain-specific set of meta-features achieve the same or even better results compared to existing solutions for different data types?
3) Whether domain-specific set of meta-features achieve the same or even better results compared to existing solutions for different anomaly types?
4) Whether domain-specific set of meta-features achieve the same or even better results compared to existing solutions for different data domains?
5) Whether a particular type of distance function achieves significant results for domain-specific set of meta-features?
6) Whether domain-specific set of meta-features reduce overall computational complexity thus a domain expert can estimate the properties of anomalies in data?

IV. DEFINITION OF PROPOSED META-FEATURES

The demand for a systematic approach for model selection used by AutoML frameworks creates an opportunity for the meta-learning approach based on anomaly detection domain-based knowledge. The existing meta-features are facing issues related to schema agnosticism, scalability, relation, and simplicity. To overcome these limitations, this paper proposes a set of domain-specific meta-features for anomaly detection which would meet the defined requirements. The proposed meta-features extractor combines the versatility of domain-specific meta-features with simple meta-features.

Let d be a dataset with n instances, where each instance $x = [v1, v2, v3, \ldots]$ is a vector with m attributes and optional target attribute which indicates whether this instance is an anomaly. Meta-feature c presents the result of a function

defined as $f(d) = c$, which when applied to dataset d, returns a vector of values which represent the properties of dataset d. These values are predictive for the performance of anomaly detection algorithms when they are applied to the dataset d. Depending on meta-features, they do not require complete datasets for applying meta-features extractor functions. The proposed meta-features that describe the properties of anomalies in data are: anomaly space, anomaly type, anomaly ratio, type of data, and data domain.

A. ANOMALY SPACE

In a dataset, depending on the number of attributes, the anomaly detection task can be referred to as univariate and multivariate. The univariate anomaly detection task creates a model based on data with anomalies for each individual attribute, while the multivariate anomaly detection task creates a single model for all attributes in the dataset. Datasets with a single attribute are considered as data with univariate anomaly space. Otherwise, datasets with more than one attribute are considered as data with multivariate anomaly space. In dataset d, anomaly space meta-feature is extracted by applying (1).

$$space(d) = \begin{cases} uni, & \text{if } attrNum(d) = 1 \\ multi, & \text{otherwise} \end{cases} \tag{1}$$

Univariate anomaly detection is based on data distribution of a single attribute space where anomalies often represent extreme values or errors in temporal data. Such anomalies can be effectively detected by using probabilistic and statistical methods [44]. Multivariate anomaly detection is based on the data distribution of a n-th attribute space. Most of the existing uses from industry are based on multivariate anomaly detection tasks, where anomalies are described as unusual behavior of at least two attributes in a dataset.

B. ANOMALY TYPE

Depending on the environment where anomalies occur, they are referred to as global anomalies, local anomalies, and micro-cluster anomalies. Based on the anomaly score, the flagging is as follows: instances with an extreme score compared to other anomalies are flagged as global anomalies, instances with an extreme score compared only to the neighbor instances are flagged as local anomalies, while instances with a score larger than normal and having neighbor instances with a similar score are flagged as micro-clusters. The anomaly type does not have to be unique for a dataset, which means that more than one anomaly type may occur in the dataset. In dataset d, anomaly type is extracted by applying (2).

$$type(i, d) = \begin{cases} global, & \text{if } score(i, d) \geq \lambda \\ local, & \text{if } score(i, \varepsilon(i, d)) \geq \lambda \text{ and} \\ & score(i, d) < \lambda \\ cluster, & \text{if } score(i, d) \geq \lambda \text{ and} \\ & score(\varepsilon(i, d), d) > \lambda \end{cases} \tag{2}$$

Anomaly type meta-feature requires a small amount of labeled data to precisely determine anomaly types for a dataset. Anomaly detection tasks are usually performed with a part of manually labeled data where baselines (λ) are determined automatically [4]. However, if labeled data is not available, an anomaly type meta-feature can be estimated by the data provider or domain expert so a dataset may be characterized by one or more type of data properties. For example, global anomalies are different from the dense areas with respect to their attributes; they represent extreme values for all instances in a dataset, which are detected by using max and min functions. Local anomalies are considered instances with a higher score compared to close-by neighborhood only. They represent extreme values only for neighbor instances in the dataset and can be detected using max and min functions over a subset of neighbor instances. Micro-cluster anomalies have scores larger than normal in- stances with close-by neighbors which also have scores larger than normal instances. They represent clusters of instances in the dataset, often indicate novelty in data, and represent an undiscovered group.

C. ANOMALY RATIO

Anomaly ratio is a simple meta-feature: it indicates the number of anomalies per total number of instances in a dataset. If a dataset with labeled instances is available, this meta-feature can be extracted with low computational costs. However, if the dataset with labeled anomalies is not available prior to the meta-learning task, then it can be estimated by a data provider or a domain expert. The anomaly ratio meta-feature in dataset d is extracted by applying (3).

$$ratio(d) = \frac{anomNum(d)}{instNum(d)} \tag{3}$$

D. TYPE OF DATA

The type of data is a meta-feature that enables the relation between similar data types. Datasets can be classified into the following categories by type: nominal, temporal, spatial, high-dimensional, and network-based. If a dataset contains attributes with a particular data type, and these attributes are included in anomaly space, it can be considered that the dataset belongs to the particular group. Data types in dataset d are extracted by applying (4).

$$data\_type(d) = \begin{cases} nominal, & \text{if } hasText(d) \\ temporal, & \text{if } hasTime(d) \\ spatial, & \text{if } hasCoord(d) \\ high\text{-}dim, & \text{if } attrNum(d) \geq \delta \\ network, & \text{if } isGraph(d) \end{cases} \tag{4}$$

This meta-feature can be estimated by a data provider or a domain expert. The dataset may be characterized by one or more type of data properties. In order to give an insight into different data types used in anomaly detection tasks, there are a number of papers that summarize the main differences between the types and present anomaly detection approaches for them [45]–[54].

E. DATA DOMAIN

Data domain is a meta-feature which divides data into several categories and thus enables the relation between data within the same domain. It provides the relation between datasets within the same domain and thus enables knowledge transfer between them. This meta-feature is determined by a data provider or a domain expert. The dataset may be characterized by one or more data domain properties. It can also be extracted by finding similar datasets using a distance function with extracted meta-features. According to available data repositories, datasets are divided into the following categories: manufacturing, transportation, finance, healthcare, text, software, and social.

V. METHODOLOGY

The model selection approach that relies on meta-learning is a special case of transfer learning where the model is determined by evaluation of the previously achieved performance of similar data. Such approach has been inspired by supervised learning, where training data is evaluated and described using a set of meta-features. Later, test data is described using the same set of meta-features where performance is estimated by evaluations of similar training data based on meta-features. The goal of such model selection approach is to predict the best algorithm for a given dataset and optimize metrics in anomaly detection tasks. In this paper, the model selection system contains the following components: metadata semantic storage used for data properties and algorithm performance evaluations; meta-features extractors for extracting the proposed meta-features; anomaly detection algorithms and distance functions based on tensorflow implementations; and datasets with labeled anomalies from different repositories.

A. TRAINING AND TESTING PHASES

The twofold training process is shown in Fig. 2. First, meta-feature extractor calculates meta-features for training datasets and stores them into metadata semantic storage. Then the system evaluates datasets using different anomaly detection algorithms for accuracy, precision, recall, and f1 optimization metrics. Those are then stored into metadata semantic storage.

After the training process, metadata semantic storage contains data properties and algorithm evaluations. During the testing phase, the meta-feature extractor calculates meta-features for test datasets and for each dataset the system scores training datasets by meta-features using a distance function. For the given optimization metric, the system predicts an algorithm by taking the best algorithm from similar datasets, as shown in Fig. 3.

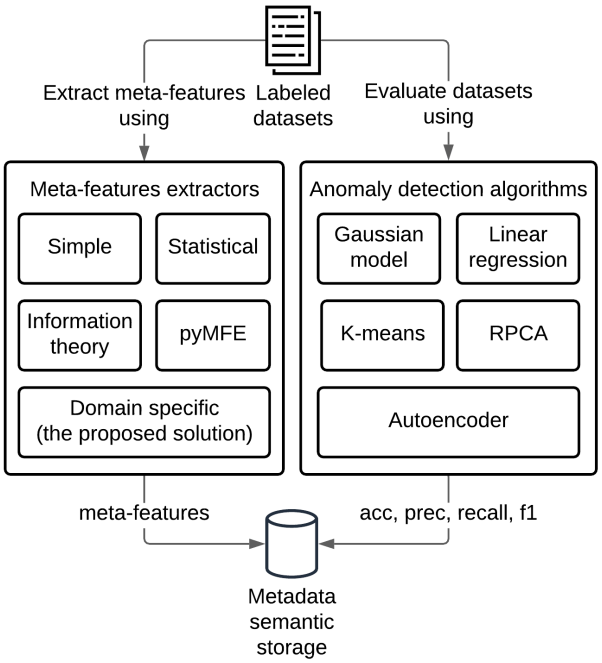

FIGURE 2. An example of model selection system architecture design [55] used for the training phase in experiments. Different meta-features extractors calculate data properties, and the system stores them into metadata semantic storage using the training set of data. The system later evaluates dataset performance using different anomaly detection algorithms and stores them into metadata semantic storage.

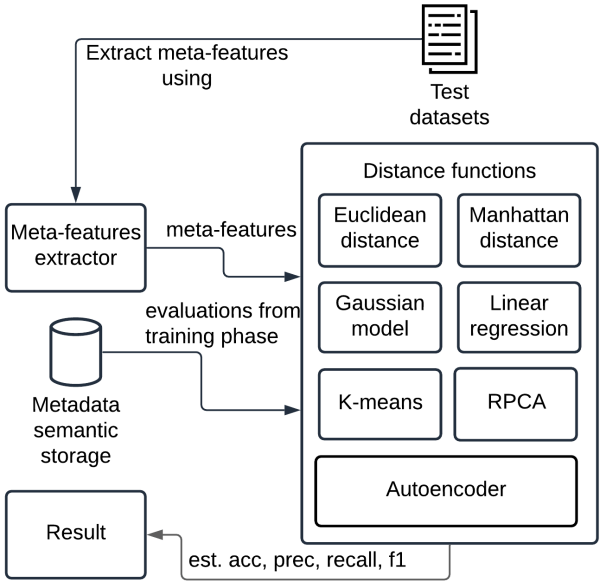

FIGURE 3. An example of model selection system architecture design [55] used for the testing phase in experiments. Meta-feature extractors calculate meta-features for test datasets. The system then scores training datasets by meta-features using different distance functions. For a given optimization metric, the system predicts an algorithm by taking the best algorithm of a dataset that has a minimal distance to the test dataset.

B. ANOMALY DETECTION ALGORITHMS

Anomaly detection algorithms create models of normal behavior patterns in data and compute anomaly scores for data instances. These models can be generative, regression-based, or proximity-based where all of them generate different properties about behavior patterns. An incorrect choice of the data model may lead to inability to achieve significant performance. Anomaly detection tasks are usually

unsupervised or semi-supervised, where a small number of anomalies is labeled. In such cases, model selection for anomaly detection presents a more challenging problem compared to the supervised problems like classification where labeled data is available. The most important anomaly detection algorithms are the following [56]: multivariate gaussian distribution, linear regression and tensor decomposition, k-means,[2] and autoencoders. From each group of algorithms with the same model properties one is selected and implemented [57] using tensorflow framework [58] for the purpose of this paper.

C. DISTANCE FUNCTIONS

When the model selection system evaluates datasets using anomaly detection algorithms and extracts domain-specific data properties, distance functions are used for measuring similarity between data using their properties. A number of distance functions are predominantly used for measuring geometry distance and do not exploit the relation between data properties by applying simple arithmetic operations [59]. To overcome these limitations, this paper uses an approach to distance functions that allows the same set of algorithms to evaluate datasets and measure distances between them. These algorithms are compared against well-known distance functions [60], [61]. Multivariate gaussian distribution calculates the distance between datasets by calculating the probability of meta-features for data distribution in situations where datasets with similar probabilities are close to each other. In linear regression, distances between datasets are calculated by measuring distances to the regression line. Again, datasets with similar distances are close to each other. Robust PCA and autoencoders calculate distances between datasets by transforming meta-features into lower-dimensional subspace and measure reconstruction errors, where datasets with similar reconstruction errors are close to each other. K-means divides data into clusters and calculates distances between datasets by measuring distances from data to cluster centers, where datasets with similar distances are close to each other.

VI. EXPERIMENTS

The experiments are designed so as to support the defined requirements and enable the adoption of meta-features in AutoML frameworks. The experiments are two-fold: training and testing phases evaluate accuracy of the proposed solution, and meta-features shifting evaluates the robustness of the solution. In the training phase, datasets are evaluated using different anomaly detection algorithms where data properties and these evaluations are stored in metadata semantic storage. In the testing phase, datasets are *1*-folded where in each iteration neighbor datasets are determined using distance functions. The achieved results are later compared against existing solutions that rely on meta-features extracted from data only. Meta-features shifting results in estimation error functions for the meta-features which have a higher tendency to be estimated with an error.

In the experiments, requirements related to schema agnosticism and relation are ensured by including data with varying schemas, from different domains and with different data types. The scalability requirement is ensured by proposing domain-specific meta-features that can be extracted by using simple operations with low computational costs. Simplicity requirement is ensured by possibility for meta-features to be estimated by a data provider or a domain expert in cases when labeled data is not available.

A. DATASETS FOR BENCHMARKING

In general, anomaly detection tasks can be considered as binary classification tasks. There are a number of dataset repositories with binary classes available, such as the UCI machine learning repository [62]. These datasets can be preprocessed for anomaly detection tasks by sub-sampling a small number of instances randomly. Unfortunately, such an approach does not reflect real-life scenarios with anomalies and they cannot be reproduced. Even if it is possible to recreate a dataset, it is not guaranteed that anomalies do not fit into normal behavior patterns; therefore this does not constitute a valid approach for anomaly detection benchmarking.

TABLE 2. Dataset repository consists of 63 datasets with labeled anomalies, varying schemas, and different anomaly types. One dataset may represent more than one class. Total column presents total number of datasets for a particular classification, while average column presents average number of anomalies in datasets for the particular classification. The created dataset repository presents a comprehensive collection for algorithm benchmarking where datasets are preprocessed using techniques such as data cleansing, data reduction, and integration.

		Anomaly type					
		Local		Global		Micro-cluster	
		total	avg	total	avg	total	avg
Data type	High-dim.	5	12×10^{-4}	5	47×10^{-4}	2	25×10^{-4}
	Temporal	41	64×10^{-6}	14	61×10^{-6}	1	18×10^{-4}
	Nominal	0	0	1	36×10^{-4}	1	36×10^{-4}
	Spatial	1	30×10^{-4}	0	0	0	0
Data domain	Manufac.	4	15×10^{-4}	0	0	0	0
	Transp.	5	67×10^{-6}	3	15×10^{-4}	0	0
	Finance	6	14×10^{-4}	0	0	0	0
	Healthc.	0	0	3	33×10^{-4}	2	36×10^{-4}
	Text	2	88×10^{-5}	2	86×10^{-4}	0	0
	Software	30	97×10^{-6}	2	17×10^{-5}	2	16×10^{-4}
	Social	0	0	10	22×10^{-6}	0	0

Dataset repository contains 63 datasets with labeled anomalies with varying schemas and different anomaly types, as summarized in Table 2. Such a diverse selection of datasets enables the validation of requirements that each meta-feature extractor must meet. The dataset repository is a comprehensive collection for algorithm benchmarking in which datasets are preprocessed using techniques like data cleansing, data reduction, and integration. Datasets from the repository meet important requirements regarding the domain background such as significant deviation from the data norm, as well as regarding the semantic background, such that evaluations simulate real-life industry cases. Datasets are collected from

[2]In order to achieve stable performance of K-means clustering, the results are obtained by multiple evaluations.

repositories in [63], [64]. The experiments are based on an industry-standard datasets [63], [64]. Datasets includes all existing data types, according to [56]. The study covers 7 different application domains (manufacturing, transportation, finance, healthcare, text, software, and social), which, according to [4] cover a great majority of general applications relevant for this study. The level of details of this study has been chosen to be equally granular or more granular, compared to the widely used and highly referenced approach of [23]. Based on the type of anomaly, it contains global and local anomalies with a small number of micro-clusters. Based on anomaly space, it contains significant amounts of univariate anomalies that are related to temporal data.

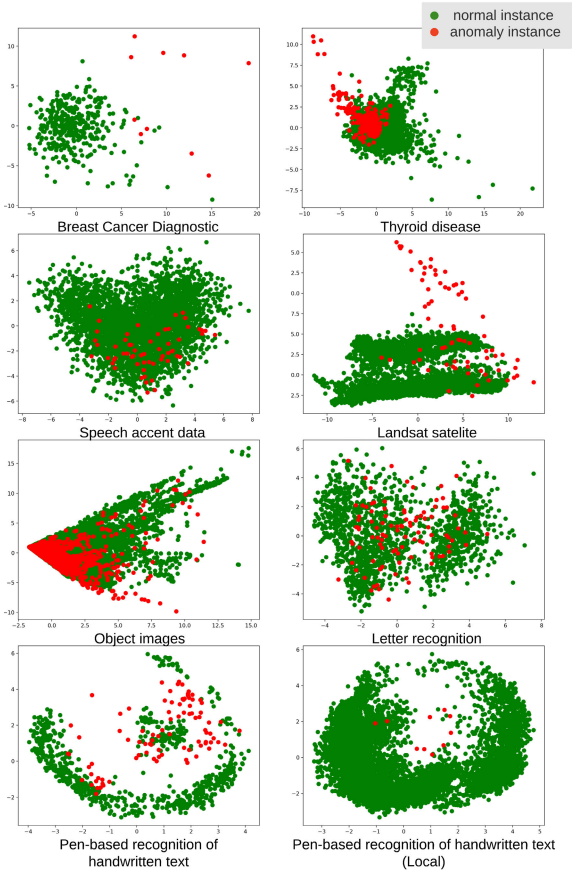

FIGURE 4. High-dimensional datasets from the dataset repository transformed in 2D sub-space using PCA linear transformation. Green points represent normal instances and red points represent anomalies. Such data distribution makes anomaly detection tasks more challenging and confirms that the proposed repository is a competitive benchmarking repository for anomaly detection.

In order to give a better insight into data from the repository, Fig. 4 presents high-dimensional datasets transformed in 2D sub-space. It is shown that the dataset repository contains different data distributions, which makes anomaly detection tasks more challenging and confirms that the proposed repository is a competitive benchmarking repository for anomaly detection.

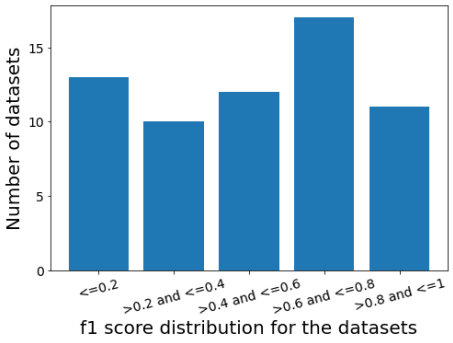

FIGURE 5. Uniform-based performance distribution of anomaly detection algorithms for datasets from the repository using $f1$ evaluation metric. Such dataset repository provides a fair performance indicator for anomaly detection tasks.

In anomaly detection, model overfitting and underfitting can occur with optimization of particular evaluation metrics due to the high disbalance between normal and anomalous instances. For example, optimization of accuracy evaluation metric creates a model where only a few anomalies are classified correctly. In the same manner, optimization of recall optimization metric creates a model where a number of normal instances are misclassified as anomalies. In order to overcome this drawback, experiments are designed to evaluate only $f1$ evaluation metric, which provides a fair performance indicator for anomaly detection tasks. Fig. 5 shows uniform-based performance distribution of anomaly detection algorithms for datasets from the repository using $f1$ evaluation metric.

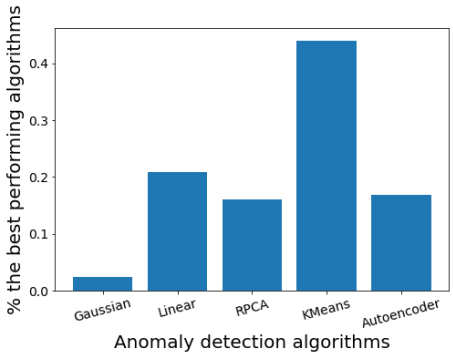

FIGURE 6. Anomaly detection algorithm comparison for datasets from the repository by the number of datasets where a particular algorithm achieves the best performance for $f1$ optimization metric.

B. EVALUATION METRICS AND BASELINES

Based on data properties and algorithm hyperparameters, some algorithms may perform better than others. Comparison of anomaly detection algorithms for datasets from the repository is shown in Fig. 6 by the number of datasets where a particular algorithm achieves the best performance for $f1$ optimization metric. It is shown that density-based algorithms achieve significant results for a number of datasets while

TABLE 3. An overview of the compared solutions with their attribute counts and descriptions.

Name	Attributes count	Description
Simple	11	Simple meta-features contain basic data properties that are directly extracted from data with low computational costs using methods with low complexity.
Statistical	26	Statistical meta-features contain statistical properties of data that are extracted from data with significant computational costs using statistical methods with significant complexity.
Information theory	8	They are directly extracted from the data with significant computational costs using methods with significant complexity but without hyperparameters, requiring significant amounts of data.
Domain-specific (**proposed solution**)	5	The proposed solution which contains domain-based knowledge where meta-features are directly extracted from data with low computational costs using methods with low complexity.
pyMFE	93	Provides a comprehensive robust set of meta-features requiring significant amounts of data and computational costs. It also includes simple, statistical, and information theory groups of meta-features.

probabilistic algorithms achieve significant results for only few datasets.

Experiments compare the proposed solution against the following meta-features: simple meta-features, statistical meta-features, and information-theory meta-features [23]. A brief overview of compared solutions is provided in Table 3. The algorithms used for anomaly detection tasks in the experiments are multivariate gaussian distribution, linear regression, tensor decomposition, k-means, and autoencoders. Moreover, these algorithms are used as distance functions and compared against euclidean and manhattan distance functions. Distance functions are an essential component of model selection based on the meta-learning approach. By choosing a less effective distance function, a model can achieve poor performance even with meta-features that meet the defined requirements. Distance functions have a hyperparameter k which determines k-nearest neighbors in meta-feature space used for algorithm selection. If k is a large number, algorithms that achieve significant results for a number of datasets have a higher impact on algorithm selection. In this paper, such experiments create models with average performance. However, if k is a small number, data properties and meta-features have a higher impact on algorithm selection. In this paper, such experiments create models which provide either effective or ineffective results. Depending on the evaluation metric, k can be determined, where in this paper the evaluation metric model with $k = 1$ achieves significant performance for $f1$.

C. RESULTS AND DISCUSSION

Results of the experiments that evaluate accuracy and robustness are presented in Table 4 and Figure 7 respectively. Accuracy results present a comparison of the proposed meta-features against the existing meta-features through different aspects. The results contain accuracy of an approach, the number of datasets for which an algorithm is correctly predicted, and distance function that provides the best accuracy. The robustness results present the estimation error functions for the anomaly ratio and anomaly type meta-features which have a higher tendency to be estimated with an error. Answers to the research questions that validate the proposed solution are presented below.

1) **Question:** Is it possible to define a set of meta-features which would effectively represent the properties of anomalies in data?
   Answer: It is shown that the proposed solution achieves an accuracy of 87% and consistently meets the baselines and in particular cases outperforms these baselines in the experiments with 63 industry-based datasets. The results in Table 4 present accuracy limits achieved by using the proposed solution and the best existing solutions [23]. It also reduces tremendously the overall computational complexity by using only some logical computation.
2) **Question:** Whether the proposed domain-specific meta-features achieve the same or even better results compared to existing solutions for temporal and high-dimensional data?
   Answer: According to the Table 4, the proposed solution achieves the same or better results compared against the existing solutions for 55 temporal industry-based datasets and 9 high-dimensional industry-based datasets.
3) **Question:** Whether the proposed domain-specific meta-feature achieves better results compared to existing solutions for different anomaly types?
   Answer: According to the Table 4, the proposed solution achieves the same or better results compared to existing solutions for all anomaly types, except for local anomalies where only the information theory meta-features achieve 11% better results.
4) **Question:** Whether the proposed domain-specific meta-features achieve better results compared against existing solutions for different data domains?
   Answer: According to the Table 4, the proposed solution achieves the same or better results compared against existing solutions for all data domains except for transportation, where only the comprehensive set of meta-features [23] achieves 25% better results.
5) **Question:** Whether a particular type of distance function achieves significant results for such meta-features?
   Answer: Distance functions based on reconstruction errors, such as RPCA and autoencoders, obtain significant performance for the proposed meta-features

TABLE 4. Results achieved using the proposed solution and the existing solutions [23] for different data types, anomaly types, and data domains. Total datasets are the total number of datasets for a particular type while dataset score is the number of datasets with the best proposed algorithm using denoted distance function. Distance functions are denoted as: G - Gaussian, K - KMeans, R - RPCA, L - Linear regression, E - Euclidean, M - Manhattan, A - Autoencoders. Datasets may belong to several categories for a particular type, which means that the sum of datasets in the total datasets row exceeds the total number of evaluated datasets. The results are obtained using the open source code available at: https://github.com/kotlarmilos/meta-features-anomaly-detection.

		Data type		Anomaly type			Data domain						
	Total	High-dimen	Temp-oral	Local	Global	Micro-cluster	Manufac-turing	Trans-portation	Fin-ance	Health-care	Text	Software	Social
Total datasets	63	9	55	45	18	3	3	8	6	2	4	30	10
Simple													
Accuracy	74%	56%	75%	76%	83%	67%	67 %	88%	100%	100%	50%	83%	90%
Datasets score	47	5	41	30	15	2	2	7	6	2	1	25	9
Distance function	L	A	K	E	R	R	A	R	K	E	L	K	R
Statistical													
Accuracy	68%	56%	91%	71%	83%	67%	67%	63%	100%	100%	50%	83 %	90 %
Datasets score	43	5	50	30	15	2	2	5	6	2	1	28	9
Distance function	K	R	R	L	R	R	L	E	R	E	R	R	K
Information theory													
Accuracy	70%	56%	71%	78%	67%	67%	67 %	75%	83%	100%	50%	77%	90%
Datasets score	44	5	39	30	12	2	2	6	5	2	1	23	9
Distance function	R	M	A	R	E	R	G	E	A	E	L	L	G
Domain-specific (proposed solution)													
Accuracy	87%	56%	91%	67%	83%	67%	67%	63%	100%	100%	50%	93%	90%
Datasets score	55	5	50	30	15	2	2	5	6	2	2	28	9
Distance function	R	R	R/A	R	R/A	R/A	R	R	R	L	L	R	R
PyMFE													
Accuracy	73%	56%	91%	67%	83%	33%	33%	88%	100%	100%	50%	67%	80%
Datasets score	46	5	50	30	15	1	1	7	6	2	2	20	8
Distance function	K	L	R	L	R	R	R	L	R	L	L	K	L

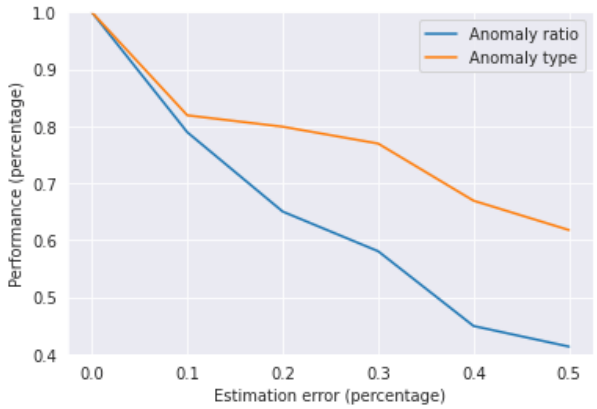

FIGURE 7. Estimation error functions (robustness) of the proposed anomaly ratio and anomaly type meta-features that could be performed by data provider or domain expert. The rest of the proposed meta-features don't have a higher tendency to be estimated with an error.

compared to other distance functions used, as shown in Table 4.

6) **Question:** Whether such meta-features reduce overall computational complexity thus a domain expert can estimate the properties of anomalies in data?
Answer: According to the Fig. 7, the proposed meta-features reduces overall computational complexity and enable robustness, thus can be estimated by a data provider or a domain expert, which makes them more generalizable and competitive for industry use. For example, if a domain expert can estimate the meta-features with 10% error, the performance of predicting the best model for a given dataset and optimization metric decreases by 20%.

D. THREATS TO VALIDITY

It is important to note that this paper does not shed light on AutoML frameworks in general and its' components, such as hyperparameters optimization, data preparation, feature engineering, and application.

Additional concerns are noise in data and validity of the experiments. In order to enhance the validity of extensive experiments, it is important to consider noise in the data. Noise in the data may provide invalid results and disprove the answered research questions. In order to eliminate such scenarios, experiments are designed to minimize this possibility. First, datasets are collected from different repositories, which reduces the possibility for noise in the data created by the same data source. Second, the proposed set of meta-features is compared against different types of existing meta-features, which reduces the possibility for noise in baseline results. Also, the proposed solution could be validated further using additional industry-based datasets with labeled anomalies, as well as to compare the proposed solution with other model-based meta-learning approaches.

VII. CONCLUSION

This paper proposes a novel set of domain-specific meta-features for model selection in anomaly detection tasks, where two sets of conclusions can be derived. One is related

to meta-learning in general and its advantages through several aspects, while the other one is related to utilization of the proposed meta-features in model selection systems and AutoML frameworks in general. In the meta-learning domain, this paper introduces a novel approach for extracting anomaly-related properties from data by proposing meta-features, which gives better results compared to existing solutions. In addition, the proposed set of meta-features meets the following requirements: schema agnosticism, scalability, relation, and simplicity. In the AutoML domain, the proposed solution focuses on improving the essential component of model selection for anomaly detection, which effectively solves existing performance-related issues.

The main research questions answered in this paper apply to the possibility to define a set of meta-features which would effectively represent the properties of anomalies from data with low computational costs. Additional research questions answered in this paper apply to the possibility to determine the relation between dataset properties and algorithms that obtain the best performance for a given optimization metric, which depends on the proposed meta-features. Experiments show that the two similar datasets could have the same best algorithms for anomaly detection tasks by measuring distances between the extracted data properties.

AutoML frameworks are becoming an important tool for bringing science to industry. This is where this research is an important step towards automated anomaly detection frameworks. The comparative analysis of methods used for anomaly detection, the created dataset repository, the proposed distance functions, and meta-features pave the way for further research in this domain. The results obtained in this paper represent baselines for benchmarking and further research in this domain.

REFERENCES

[1] M. Carletti, C. Masiero, A. Beghi, and G. A. Susto, "Explainable machine learning in industry 4.0: Evaluating feature importance in anomaly detection to enable root cause analysis," in *Proc. IEEE Int. Conf. Syst., Man Cybern. (SMC)*, Oct. 2019, pp. 21–26.

[2] M. Braun, P. Converse, and F. Oswald, "The accuracy of dominance analysis as a metric to assess relative importance: The joint impact of sampling error variance and measurement unreliability," *J. Appl. Psychol.*, vol. 104, no. 4, p. 593, 2019.

[3] R. Chalapathy and S. Chawla, "Deep learning for anomaly detection: A survey," 2019, *arXiv:1901.03407*. [Online]. Available: http://arxiv.org/abs/1901.03407

[4] M. Goldstein and S. Uchida, "A comparative evaluation of unsupervised anomaly detection algorithms for multivariate data," *PLoS ONE*, vol. 11, no. 4, Apr. 2016, Art. no. e0152173.

[5] S. Wolfert, L. Ge, C. Verdouw, and M.-J. Bogaardt, "Big data in smart farming—A review," *Agricult. Syst.*, vol. 153, pp. 69–80, May 2017.

[6] Y. Wang, L. Kung, and T. A. Byrd, "Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations," *Technological Forecasting Social Change*, vol. 126, pp. 3–13, Jan. 2018.

[7] J. Liu, J. Guo, P. Orlik, M. Shibata, D. Nakahara, S. Mii, and M. Takáč, "Anomaly detection in manufacturing systems using structured neural networks," in *Proc. 13th World Congr. Intell. Control Autom. (WCICA)*, Jul. 2018, pp. 175–180.

[8] B. Lindemann, F. Fesenmayr, N. Jazdi, and M. Weyrich, "Anomaly detection in discrete manufacturing using self-learning approaches," *Procedia CIRP*, vol. 79, pp. 313–318, Jan. 2019.

[9] Q. Lu, F. Chen, and K. Hancock, "On path anomaly detection in a large transportation network," *Comput., Environ. Urban Syst.*, vol. 33, no. 6, pp. 448–462, Nov. 2009.

[10] M. H. Hassan, A. Tizghadam, and A. Leon-Garcia, "Spatio-temporal anomaly detection in intelligent transportation systems," *Procedia Comput. Sci.*, vol. 151, pp. 852–857, Jan. 2019.

[11] J. O. Awoyemi, A. O. Adetunmbi, and S. A. Oluwadare, "Credit card fraud detection using machine learning techniques: A comparative analysis," in *Proc. Int. Conf. Comput. Netw. Informat. (ICCNI)*, Oct. 2017, pp. 1–9.

[12] N. Margaliot, "Systems and methods for derivative fraud detection challenges in mobile device transactions," U.S. Patent App. 15 184 818, Dec. 29, 2016.

[13] J. Knights, Z. Heidary, and J. M. Cochran, "Detection of behavioral anomalies in medication adherence patterns among patients with serious mental illness engaged with a digital medicine system," *JMIR Mental Health*, vol. 7, no. 9, Sep. 2020, Art. no. e21378.

[14] X. Wu and X. Zhu, "Mining with noise knowledge: Error-aware data mining," *IEEE Trans. Syst., Man, Cybern. A, Syst. Humans*, vol. 38, no. 4, pp. 917–932, Jul. 2008.

[15] A. Oussous, F. Z. Benjelloun, A. A. Lahcen, and S. Belfkih, "Big data technologies: A survey," *J. King Saud Univ. Comput. Inf. Sci.*, vol. 30, no. 4, pp. 431–448, Oct. 2018.

[16] W. A. Günther, M. H. Rezazade Mehrizi, M. Huysman, and F. Feldberg, "Debating big data: A literature review on realizing value from big data," *J. Strategic Inf. Syst.*, vol. 26, no. 3, pp. 191–209, Sep. 2017.

[17] K. Wagstaff, "Machine learning that matters," 2012, *arXiv:1206.4656*. [Online]. Available: http://arxiv.org/abs/1206.4656

[18] S. W. Yahaya, A. Lotfi, and M. Mahmud, "A consensus novelty detection ensemble approach for anomaly detection in activities of daily living," *Appl. Soft Comput.*, vol. 83, Oct. 2019, Art. no. 105613.

[19] X. He, K. Zhao, and X. Chu, "AutoML: A survey of the state-of-the-art," 2019, *arXiv:1908.00709*. [Online]. Available: http://arxiv.org/abs/1908.00709

[20] X. He, K. Zhao, and X. Chu, "AutoML: A survey of the state-of-the-art," *Knowl.-Based Syst.*, vol. 212, Jan. 2021, Art. no. 106622.

[21] C. Castiello, G. Castellano, and A. M. Fanelli, "Meta-data: Characterization of input features for meta-learning," in *Proc. Int. Conf. Modeling Decisions Artif. Intell.* Heidelberg, Germany: Springer, 2005, pp. 457–468.

[22] A. Filchenkov and A. Pendryak, "Datasets meta-feature description for recommending feature selection algorithm," in *Proc. Artif. Intell. Natural Lang. Inf. Extraction, Social Media Web Search FRUCT Conf. (AINL-ISMW FRUCT)*, Nov. 2015, pp. 11–18.

[23] A. Rivolli, L. P. F. Garcia, C. Soares, J. Vanschoren, and A. C. P. L. F. de Carvalho, "Characterizing classification datasets: A study of meta-features for meta-learning," 2018, *arXiv:1808.10406*. [Online]. Available: http://arxiv.org/abs/1808.10406

[24] H. S. Jomaa, L. Schmidt-Thieme, and J. Grabocka, "Dataset2Vec: Learning dataset meta-features," 2019, *arXiv:1905.11063*. [Online]. Available: http://arxiv.org/abs/1905.11063

[25] G. J. Aguiar, E. J. Santana, S. M. Mastelini, R. G. Mantovani, and S. Barbon, Jr., "Towards meta-learning for multi-target regression problems," in *Proc. 8th Brazilian Conf. Intell. Syst. (BRACIS)*, Oct. 2019, pp. 377–382.

[26] M. Hu, Z. Ji, K. Yan, Y. Guo, X. Feng, J. Gong, X. Zhao, and L. Dong, "Detecting anomalies in time series data via a meta-feature based approach," *IEEE Access*, vol. 6, pp. 27760–27776, 2018.

[27] M. Canizo, I. Triguero, A. Conde, and E. Onieva, "Multi-head CNN–RNN for multi-time series anomaly detection: An industrial case study," *Neurocomputing*, vol. 363, pp. 246–260, Oct. 2019.

[28] A. Cohen and N. Nissim, "Trusted detection of ransomware in a private cloud using machine learning methods leveraging meta-features from volatile memory," *Expert Syst. Appl.*, vol. 102, pp. 158–178, Jul. 2018.

[29] R. S. Oyamada, L. Shimomura, S. B. Junior, and D. Kaster, "Towards proximity graph auto-configuration: An approach based on meta-learning," in *Proc. Eur. Conf. Adv. Databases Inf. Syst.* Heidelberg, Germany: Springer, 2020, pp. 93–107.

[30] Y. Zhang, R. Zhu, Z. Chen, J. Gao, and D. Xia, "Evaluating and selecting features via information theoretic lower bounds of feature inner correlations for high-dimensional data," *Eur. J. Oper. Res.*, vol. 290, no. 1, pp. 235–247, Apr. 2021.

[31] J. Madrid and H. J. Escalante, "Meta-learning of text classification tasks," in *Proc. Iberoamerican Congr. Pattern Recognit.* Heidelberg, Germany: Springer, 2019, pp. 107–119.
[32] J. A. Saez and E. Corchado, "A meta-learning recommendation system for characterizing unsupervised problems: On using quality indices to describe data conformations," *IEEE Access*, vol. 7, pp. 63247–63263, 2019.
[33] B. A. Pimentel and A. C. P. L. F. de Carvalho, "Unsupervised meta-learning for clustering algorithm recommendation," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2019, pp. 1–8.
[34] A. E. Gutierrez-Rodríguez, S. E. Conant-Pablos, J. C. Ortiz-Bayliss, and H. Terashima-Marín, "Selecting meta-heuristics for solving vehicle routing problems with time windows via meta-learning," *Expert Syst. Appl.*, vol. 118, pp. 470–481, Mar. 2019.
[35] V. Milutinovic, "The best method for presentation of research results," *IEEE TCCA Newslett.*, vol. 214, no. 2, pp. 1–6, Sep. 1996.
[36] A. Correia, C. Soares, and A. Jorge, "Dataset morphing to analyze the performance of collaborative filtering," in *Proc. Int. Conf. Discovery Sci.* Heidelberg, Germany: Springer, 2019, pp. 29–39.
[37] I. Tanfilev, A. Filchenkov, and I. Smetannikov, "Feature selection algorithm ensembling based on meta-learning," in *Proc. 10th Int. Congr. Image Signal Process., Biomed. Eng. Informat. (CISP-BMEI)*, Oct. 2017, pp. 1–6.
[38] Y. Zhang, F. Feng, C. Wang, X. He, M. Wang, Y. Li, and Y. Zhang, "How to retrain recommender system?: A sequential meta-learning method," in *Proc. 43rd Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, Jul. 2020, pp. 1479–1488.
[39] J. Ding, V. Tarokh, and Y. Yang, "Model selection techniques: An overview," *IEEE Signal Process. Mag.*, vol. 35, no. 6, pp. 16–34, Nov. 2018.
[40] A. Akbar, A. Khan, F. Carrez, and K. Moessner, "Predictive analytics for complex IoT data streams," *IEEE Internet Things J.*, vol. 4, no. 5, pp. 1571–1582, Oct. 2017.
[41] B. Bilalli, A. Abelló, and T. Aluja-Banet, "On the predictive power of meta-features in OpenML," *Int. J. Appl. Math. Comput. Sci.*, vol. 27, no. 4, pp. 697–712, Dec. 2017.
[42] A. Ngom, I. Stojmenovic, and V. Milutinovic, "STRIP—A strip-based neural-network growth algorithm for learning multiple-valued functions," *IEEE Trans. Neural Netw.*, vol. 12, no. 2, pp. 212–227, Mar. 2001.
[43] I. Yaqoob, I. Hashem, A. Gani, S. Mokhtar, E. Ahmed, N. Anuar, and A. Vasilakos, "Big data: From beginning to future," *Int. J. Inf. Manage.*, vol. 36, no. 6, pp. 1231–1247, Dec. 2016.
[44] M. Braei and S. Wagner, "Anomaly detection in univariate time-series: A survey on the state-of-the-art," 2020, *arXiv:2004.00433*. [Online]. Available: http://arxiv.org/abs/2004.00433
[45] A. Taha and A. S. Hadi, "Anomaly detection methods for categorical data: A review," *ACM Comput. Surveys*, vol. 52, no. 2, pp. 1–35, May 2019.
[46] V. Vercruyssen, W. Meert, G. Verbruggen, K. Maes, R. Baumer, and J. Davis, "Semi-supervised anomaly detection with an application to water analytics," in *Proc. IEEE Int. Conf. Data Mining (ICDM)*, Nov. 2018, pp. 527–536.
[47] W.-K. Wong, A. Moore, G. Cooper, and M. Wagner, "Rule-based anomaly pattern detection for detecting disease outbreaks," in *Proc. AAAI/IAAI*, 2002, pp. 217–223.
[48] M. Landauer, M. Wurzenberger, F. Skopik, G. Settanni, and P. Filzmoser, "Dynamic log file analysis: An unsupervised cluster evolution approach for anomaly detection," *Comput. Secur.*, vol. 79, pp. 94–116, Nov. 2018.
[49] Y. Djenouri, A. Belhadi, J. C.-W. Lin, and A. Cano, "Adapted K-nearest neighbors for detecting anomalies on spatio–temporal traffic flow," *IEEE Access*, vol. 7, pp. 10015–10027, 2019.
[50] H. H. Bosman, G. Iacca, A. Tejada, H. J. Wörtche, and A. Liotta, "Spatial anomaly detection in sensor networks using neighborhood information," *Inf. Fusion*, vol. 33, pp. 41–56, Jan. 2017.
[51] H. Song, Z. Jiang, A. Men, and B. Yang, "A hybrid semi-supervised anomaly detection model for high-dimensional data," *Comput. Intell. Neurosci.*, vol. 2017, pp. 1–9, Nov. 2017.
[52] G. O. Campos, A. Zimek, J. Sander, R. J. G. B. Campello, B. Micenková, E. Schubert, I. Assent, and M. E. Houle, "On the evaluation of unsupervised outlier detection: Measures, datasets, and an empirical study," *Data Mining Knowl. Discovery*, vol. 30, no. 4, pp. 891–927, Jul. 2016.
[53] M. Salehi and L. Rashidi, "A survey on anomaly detection in evolving data: [With application to forest fire risk prediction]," *ACM SIGKDD Explor. Newslett.*, vol. 20, no. 1, pp. 13–23, May 2018.
[54] M. Yoon, B. Hooi, K. Shin, and C. Faloutsos, "Fast and accurate anomaly detection in dynamic graphs with a two-pronged approach," in *Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Jul. 2019, pp. 647–657.
[55] A. R. S. Parmezan, H. D. Lee, and F. C. Wu, "Metalearning for choosing feature selection algorithms in data mining: Proposal of a new framework," *Expert Syst. Appl.*, vol. 75, pp. 1–24, Jun. 2017.
[56] C. C. Aggarwal, "Outlier analysis," in *Data Mining*. Heidelberg, Germany: Springer, 2015, pp. 237–263.
[57] M. Kotlar. (2021). *Novel Meta-Features for Automated Machine Learning Model Selection in Anomaly Detection*. [Online]. Available: https://github.com/kotlarmilos/meta-features-anomaly-detection
[58] M. Abadi, "TensorFlow: Learning functions at scale," in *Proc. 21st ACM SIGPLAN Int. Conf. Funct. Program.*, Sep. 2016, p. 1.
[59] A. Mucherino and D. S. Gonçalves, "An approach to dynamical distance geometry," in *Proc. Int. Conf. Geometric Sci. Inf.* Heidelberg, Germany: Springer, 2017, pp. 821–829.
[60] X. Gao and G. Li, "A KNN model based on manhattan distance to identify the SNARE proteins," *IEEE Access*, vol. 8, pp. 112922–112931, 2020.
[61] Y. Huang, W. Jin, B. Li, P. Ge, and Y. Wu, "Automatic modulation recognition of radar signals based on manhattan distance-based features," *IEEE Access*, vol. 7, pp. 41193–41204, 2019.
[62] D. Dua and C. Graff, "UCI machine learning repository," School Inf. Comput. Sci., Univ. California, Irvine, CA, USA, Tech. Rep., 2017.
[63] M. Goldstein, "Unsupervised anomaly detection benchmark," Harvard Univ., Cambridge, MA, USA, Tech. Rep., 2015.
[64] S. Ahmad, A. Lavin, S. Purdy, and Z. Agha, "Unsupervised real-time anomaly detection for streaming data," *Neurocomputing*, vol. 262, pp. 134–147, Nov. 2017.

MILOŠ KOTLAR received the B.Sc. and M.Sc. degrees in electrical and computer engineering from the School of Electrical Engineering, University of Belgrade, Serbia, in 2016 and 2017, respectively, where he is currently pursuing the Ph.D. degree with the School of Electrical Engineering. His general research interests include implementation of energy efficient tensor implementations using the dataflow paradigm (FPGA and ASIC accelerators) and meta learning approaches for anomaly detection tasks.

MARIJA PUNT received the B.Sc., M.Sc., and Ph.D. degrees in electrical and computer engineering from the School of Electrical Engineering, University of Belgrade, Serbia, in 2004, 2009, and 2015, respectively. She is currently an Assistant Professor with the University of Belgrade. She teaches several courses on computer architecture and organization, web design, and human–computer interaction. Her research interests include computer architecture, digital systems simulation, consumer electronics, data analysis, and human–computer interaction.

ZAHARIJE RADIVOJEVIĆ received the B.Sc., M.Sc., and Ph.D. degrees in electrical and computer engineering from the School of Electrical Engineering, University of Belgrade, Serbia, in 2002, 2006, and 2012, respectively. He is currently an Associate Professor with the University of Belgrade. He teaches several courses on computer architecture and organization, e-business infrastructure, and mobile device programming. His research interests include computer architecture and organization, concurrent and distributed programming, data analysis, simulations, and reverse engineering.

MILOŠ CVETANOVIĆ received the B.Sc., M.Sc., and Ph.D. degrees in electrical and computer engineering from the School of Electrical Engineering, University of Belgrade, Serbia, in 2003, 2006, and 2012, respectively. He is currently an Associate Professor with the University of Belgrade. He teaches several courses on databases and database software tools, information systems, and e-business infrastructure. His research interests include the area of database and information systems, artificial intelligence, big data, and reverse engineering.

VELJKO MILUTINOVIĆ received the Ph.D. degree from the University of Belgrade, Serbia. He spent about a decade on a various faculty positions in the USA (mostly at Purdue University and more recently at Indiana University Bloomington). He was a co-designer of the DARPAs pioneering GaAs RISC microprocessor on 200 MHz (about a decade before the first commercial effort on that same speed) and also a co-designer of the related GaAs Systolic Array (with 4096 GaAs microprocessors). Later, for almost three decades, he taught and conducted research at the University of Belgrade, for the departments of EE, MATH, BA, and PHYS/CHEM. His research interests include data mining algorithms and dataflow computing, with the emphasis on mapping of data analytics algorithms onto fast energy efficient architectures. Most of his research was done in cooperation with industry (Intel, Fairchild, Honeywell, Maxeler, HP, IBM, NCR, and RCA).