A literature review on one-class classification and its potential applications in big data

Naeem Seliya[1*], Azadeh Abdollah Zadeh[2] and Taghi M. Khoshgoftaar[2]

*Correspondence:
seliyana@uwec.edu;
nseliya@gmail.com
[1] University of Wisconsin-Eau
Claire, Eau Claire, WI 54702,
USA
Full list of author information
is available at the end of the
article

Abstract

In severely imbalanced datasets, using traditional binary or multi-class classification typically leads to bias towards the class(es) with the much larger number of instances. Under such conditions, modeling and detecting instances of the minority class is very difficult. One-class classification (OCC) is an approach to detect abnormal data points compared to the instances of the known class and can serve to address issues related to severely imbalanced datasets, which are especially very common in big data. We present a detailed survey of OCC-related literature works published over the last decade, approximately. We group the different works into three categories: outlier detection, novelty detection, and deep learning and OCC. We closely examine and evaluate selected works on OCC such that a good cross section of approaches, methods, and application domains is represented in the survey. Commonly used techniques in OCC for outlier detection and for novelty detection, respectively, are discussed. We observed one area that has been largely omitted in OCC-related literature is its application context for big data and its inherently associated problems, such as severe class imbalance, class rarity, noisy data, feature selection, and data reduction. We feel the survey will be appreciated by researchers working in these areas of big data.

Keywords: One-class classification, Big data, Outlier detection, Novelty detection, Deep learning, Class imbalance

Introduction

The commonly known five Vs of big data are volume, variety, value, veracity, and velocity. The enormous volume of big data poses unique challenges, e.g., in a binary classification problem, the number of instances in the positive class (the class of interest) is miniscule compared to the number of instances in the negative class. This brings up issues such as how to handle the very high-class imbalance in big data, the presence of class rarity of the positive class instances in big data [1–4], and modeling bias toward the negative class (the class of less interest). Variety indicates that big data can have data obtained from multiple sources. Value is often considered the most important aspect of big data, and that is because mining such large data corpus should yield results that of practical business value to the end user. Veracity in big data often refers to the truthfulness of the data points in the big data set—for example, how to handle missing data points? How to cleanse the data set? And how accurate are the data points? Velocity indicates the rate

at which data is incoming and how it could potentially change the characteristics of the volume of big data. Is it better to have limited data in real time than lots of data a low speed?

While we do not attempt to put focus on every aspect of big data in this paper, we focus on how one-class classification (OCC) can aid specific issues attributed to big data. Some of these include severe class imbalance, presence of class rarity, data cleansing for improved data quality, feature selection, and data volume reduction. Toward that end, it is important to clearly understand the area of one-class classification in the field of data mining and machine learning. In this paper, we put emphasis on exploring various works done in one-class classification. In addition, we comment on whether sufficient work has been done on OCC with big data to provide researchers with techniques to address big data problems mentioned above. We felt a current survey of OCC methodologies would provide insight into their use for addressing some of the specific problems encountered with big data.

In a binary classification problem with instances from the positive and negative classes, a traditional machine learning algorithm aims to discriminate between the two classes and build a prediction model that can accurately classify unlabeled (previously unseen) instances of the two classes. However, in the case of class imbalance, the number of instances in the negative class is disproportionately high compared to the same in the positive class (the class of interest). Under such circumstances, a typical classifier will show bias toward the class with the larger number of instances, i.e., the negative class. When the class imbalance is severe, accurately classifying the positive class is very challenging and sometimes impractical with traditional binary classifiers. For example, detecting illegal bank transactions is fraught with severe class imbalance as the number of positive instances (illegal transactions) is much smaller than the number of negative instances (legal transactions). In such a scenario, if data on the positive instances is available while data on the negative instances is either not available or is unlabeled, how does one perform classification-based prediction modeling? To address such a problem, approaches based on the one-class classification (OCC) concept can be used.

One-class classification is a specific type of multi- or binary-classification where the classification problem is addressed by examining and analyzing instances of only one class, which is usually the class of interest. In an OCC problem scenario, labeled instances of the positive class(es) are either not available or are not in adequate numbers to train a traditional machine learner. Revisiting the problem of classifying legal/illegal bank transactions, OCC can be used to classify a previously unseen transaction as legal or illegal. We discuss OCC further in the next section. In this study, we present a survey of approaches, methodologies, and algorithms for OCC in the literature approximately over the last 10–11 years, i.e., 2010–2021 (May). The goal of the survey is to provide a good cross section of different methods and approaches to OCC and its applications investigated during the last 10–11 years and is not meant to be an exhaustive survey of all related works.

Outlier detection and novelty detection are observed as the primary areas of use for one-class classification in our surveyed works. In addition, we also categorize surveyed works based on the use of deep learning in the context of one-class classification. Outlier detection and novelty detection have a subtle difference in their concepts and

applications. In novelty detection, the anomalies are detected in the test dataset while the training dataset does not contain any anomalous data points. In outlier detection, the training dataset may contain both normal and anomalous data points and the task is to determine the boundary(ies) between the two. The boundary is subsequently applied on the test dataset, which again may contain both normal and anomalous data points.

The remainder of the paper is structured as follows. "One-class classification" section provides further detail on OCC and its primary types. "Summary of surveyed works" section provides detailed summaries of the surveyed works on OCC in the context of outlier detection, novelty detection, and application of deep learning in OCC. This section also discusses previous survey papers on OCC, and how this paper is different from those papers. "Discussion" section provides a discussion of the surveyed works and the overall OCC problem. "Conclusion" section provides a conclusion of this paper, including some suggestions for future work.

One-class classification

In several real-world datasets, labeled examples are available for only one class. Since the number of unlabeled samples can be large, it increases the learning time for standard classification approaches, primarily due to the large size of the dataset. At this point, one of the solutions to solve the problem of classification is applying one-class classification to classify the unseen transaction as legal (normal) or illegal (abnormal). Since one-class classification is performed by only instances of one class, it involves more complex solutions to attain accurate results. One-class classification (OCC) is a specific type of multi- or binary-classification task done by only instances of one class. The other class samples are either not available or not adequate in numbers for training a more traditional (non-OCC) classifier. In some cases, the number of collected samples is unsatisfactory.

To clarify the concept of OCC, we consider some examples. Consider specific problems such as granting credit cards to customers. In this example, organizations that give credit cards need to evaluate new customers' applications or the behavior of existing customers to accept or reject them. Since most customers pay their loan and very few people default, we do not have an acceptable portion of defaulters, and the datasets are extremely imbalanced. As another example, the normal data regarding device status are abundant in the health monitoring of turbines or offshore platforms. However, abnormal statuses occur infrequently, and experts are interested in detecting those rare conditions. Other similar examples can be cited to explain the use and importance of OCC.

The class of well-sampled adequate instances of the training set is assumed as the target class while the outlier class instances are very sparse or unavailable. Unavailability of the outlier class may result in measurement difficulty, or the high cost of collecting the examples. In several one-class classification algorithms, finding the decision boundary over the training set is a goal. The primary characteristic of OCC is that it can distinguish one class object from other objects by only one-class learning. It means that even when there is no sample of another class, OCC is applicable. Furthermore, since one of the goals of OCC is identifying hidden outliers against target-class samples, producing robust decision boundaries is a fundamental part of OCC. The objective of a one-class classifier may be obtained as different types, such as assigning a class label, considering an area around one class, or belonging (and not-belonging) of an object to a class. One

Seliya *et al. J Big Data*   (2021) 8:122

Page 4 of 31

of the popular reasons for utilizing OCC is its potency in detecting abnormal objects or outliers or suspicious patterns. Using just target class objects for the training empowers OCC to be a practical choice of outlier detection and novelty detection.

The lack of instances from one-class has the potential to disrupt the class discrimination process. Having only one well-trained class makes the decision boundary differentiation among its examples hard. Moreover, single class instances impose problems on the feature selection [5, 6] as we must deal with only one class as compared to the traditional binary or multi-class problem. Consequently, finding the best subset of features to have a proper separation between classes is burdensome. As outlier instances are not available, the training set only contains target instances, making the data boundary non-convex [7]. As a result, additional number of instances are needed to train the model in comparison with the more conventional or traditional multi/binary classification problem. In a typical one-class classification, the decision to accept a data point as an inlier or outlier is based on two parameters: a parameter to calculate the distance of a sample to the target class and a threshold limit defined by a user to compare the distance and accept or reject the object as inlier [8]. Khan et al. [9] categorize OCC techniques based on the model of the classifier, the data types being analyzed, and temporal relations of the features. The model of the classifier is divided into three types, such as density-based, boundary-based, and reconstruction-based.

Density-based one-class classification methods perform based on estimating the training data density, which is compared to a threshold, a model parameter. These types of methods are applicable in well-sampled data with a high number of training samples. The Gaussian Method, Mixture of Gaussians, and Parzen Density are categorized as density-based methods. In boundary-based methods, a closed border and boundary around the inliers are built, which makes the optimization of the boundary a modeling challenge. Any sample outside the border of the boundary is considered as an outlier. One-class support vector machine (OCSVM) is one of the kernel-based methods based on support vector machines (SVMs). OCSVM is built by developing a hyperplane that maximizes the distance from the origin and separates the outliers from inliers [10]. Another kernel-based one-class classification method, Support Vector Data Description (SVDD), builds a hypersphere with a minimum radius, which comprises target samples, and any sample outside of the hypersphere is considered as outlier [11]. Boundary-based methods require fewer data samples comparing to density-based methods for similar performances. In reconstruction-based methods, domain-specific historical data (prior knowledge) is needed as an assumption in generating the models. Outlier samples would typically not comply with the historical data assumption embedded in the models, and thus, any sample with a high reconstruction error is considered an outlier. In this method, an input pattern is represented as output, and the reconstruction error is minimized. The K-means clustering-based one-class classifier [12], Principal Component Analysis (PCA) based one-class classifier [13], Learning Vector Quantization (LVQ) based one-class classifier [14], and Auto-Encoder [15] or Multi-layer Perceptron (MLP) [16] methods are reconstruction-based models.

An ensemble-based one-class classifier is the combination of multiple one-class classifiers to collectively benefit from each of them. Desir et al. [17] proposed One-Class Random Forest (OCRF), which boosts some weak classifiers and where an artificial outlier

generation process is integrated to change the one-class classification to a binary learner. One-Class Clustering-based Ensemble (OCClustE) builds clusters from feature space [18]. This method considerably reduces processing time. One-Class Linear Programming (OCLP) is an efficient method to detect dissimilarity representations [19]. The OCLP method has the advantage of reducing the number of objects for testing. Graph-based semi-supervised one-class classification with OCSVM is used to detect abnormal lung sounds with few labeled normal samples [20]. The authors built a spectral graph to show relationships between samples. A comprehensive comparison of extreme learning (ELM) based one-class classification, which includes two types of boundary-based methods and reconstruction-based methods, is presented by [21]. Weighted one-class support vector machine with incremental learning and forgetting is presented by Krawczyk and Wozniak [22]. In incremental learning, the data is regularly used to increase the model knowledge, which changes the prior decision boundaries. This method is useful for data streaming modeling and analysis.

Summary of surveyed works

A select group of works on one-class classification is summarized in this section. The select group is obtained from OCC related works during the last decade (2010–2021). While not meant to be an exhaustive survey of all OCC related works, we have attempted to present a good cross section (to the best of our knowledge) of one-class classification works published during the last ten years. Based on the focus and approach of the surveyed work, we group them into three categories: outlier detection and OCC, novelty detection and OCC, and deep learning and OCC.

Outlier detection and OCC

Bartkowiak [7] presents a case study on detecting abnormal patterns (or masqueraders) in computer system calls. The data set represents 50 users with each having a sequence of 15,000 system calls. The collection of system calls was abstracted into two sets, i.e., 50 blocks (Part A) and 100 blocks (Part B), each containing 100 calls. Part A contained no masqueraders, while in Part B some of the blocks were replaced by blocks from 20 users posing as masqueraders. The OCC problem here is to detect these masquerader blocks. A detailed analysis is provided for the blocks of one user that has about 20 abnormal blocks. In masquerader detection, decision boundaries are built by using OCC to model the density of data. Constructions are based on classic Gaussian distribution, robust Gaussian distribution, and SVM. The author shows that to monitor an unusual event in the context of the case study, applying OCC methods is practical. It is also shown that reconstruction methods may be useful since the user investigated about half of the implanted blocks (masqueraders) needed to be detected. In addition to the case study, the paper discusses the benefits of statistical method and machine learning methods for network anomaly detection. The study could have more reliable appeal for masquerader detection if actual alien (unauthorized) users were involved in the data set and were detected. Moreover, a case study with a larger number of users and system calls would lend to improved generalizability of the work.

Leng et al. [23] present a one-class classifier based on extreme learning machine (ELM), in which the hidden layer of the neural network does not need tuning, and the

output weights are computed analytically leading to relatively faster learning time. They compare their proposed method with the autoencoder neural networks, with a reconstruction method for building a one-class classifier. An outlier detection analysis in the context of seven UCI data sets and three artificially generated data sets is conducted. While both random feature mappings and kernels can be used for the proposed classifier, the latter yields better results than the former. The primary comparative study between their ELM-based model and the autoencoder neural network suggested the former has an analytical solution that obtains better generalization performance and that too at relatively faster network learning times. A downside of this study is that the data sets investigated in the study are relatively small in size, leaving the research gap of how the proposed method would scale up to much larger data sets especially since neural networks are notorious for relatively slow learning. How would the authors' ELM-based approach work effectively with big data?

Gautam et al. [21] present six OCC approaches, grouped into two categories: three reconstruction-based OCC methods and three boundary-based OCC methods. The proposed OCC methods are based on ELM and online sequential extreme learning machine (OSELM). The authors discuss both online and offline approaches for OCC. Among the four offline methods, two approaches perform random feature mapping while the other two perform kernel feature mapping. The case study data sets consisted of two artificially created data sets and eight benchmark data sets from different domains for evaluating the performance of the OCC models. The authors state that the proposed classifiers perform better than ten traditional OCC and two ELM-based classifiers. ELM, in the context of OCC, is used by other studies as well, e.g., Dai et al. [24] and Leng et al. [23]. While the authors use some benchmark datasets, their analysis and conclusion are also based on artificially generated datasets.

Dreiseitl et al. [25] examine outlier detection with one-class support vector machines in the context of detecting abnormal cases of melanoma prognosis. The one-class classification aims to model the distribution of melanoma patients who have not obtained metastases status, which in this context is the normal class (case) for patients with melanoma. The case study data was obtained from the Department of Dermatology of the Medical University of Vienna. The post-cleansing data set consisted of 270 serologic blood tests, including those from 37 patients with metastatic disease and 233 patients without metastatic disease. The one-class SVM approach was compared with regular two-class SVM and Artificial Neural Networks (ANN) algorithms. These were investigated using the WEKA data mining tool [26]. Their empirical work suggests that one-class SVMs are a good alternative to standard classification algorithms in the case where there are only few cases available from the class of interest, i.e., patients with metastatic disease in this context. The one-class SVM models performed better than the two-class models when the latter used less than half the number of cases in the minority class. A potential problem with this study is the very small data set size of the case study, and whether their approach is scalable to larger data sets, such as big data.

Mourão-Miranda et al. [27] present a method that classifies patient brain activity by one-class SVM (OCSVM). The method analyzes Functional Magnetic Resonance Imaging (fMRI) feedback to sad facial expressions in patients with depression. They examine the fMRI of those patients, compare them to healthy (not depressed) patients, and

conclude that fMRI responses of depressed patients are classified as outliers. The data set consisted of 19 depressed patients and 19 not-depressed patients. The OCSVM classification reveals that there is a strong interconnection between healthy patient boundaries and the Depression Hamilton Rating Scale. Moreover, two subcategories among patients are discovered by OCSVM. These subcategories are categorized based on the patient's reactions to treatment. To classify the individual as depressed and healthy, the proposed algorithm used two types of brain data such as voxel (the voxel size is the spatial 3D resolution of an image) of whole brain and regions of the brain, and it extracts around 500 for the whole brain and 348 features for the region. The consideration of brain region-based images and using OCSVM to treat patients makes this study a notable work in the application of OCC in the healthcare. The very small data set size of the case study makes it difficult to make broad generalization conclusions, especially in the context of big data.

Bartkowiak and Zimroz [28] investigate the planetary gearbox (placed in a bucket wheel excavator) vibration signal and detected outlier data. They collected the two datasets from segmented vibration signal spectrums as a "Good" dataset and a "Bad" dataset. In the case that the gearbox is in bad condition, it makes many harmonic signals with high Signal to Noise Ratio (SNR), and when in good condition, harmonics and SNR become relatively lower. The number of samples in the good dataset is 951, with 15 attributes. They applied the Neuroscale technique (a visualization method) to reduce the attributes to two features, and thus, the data can be plotted on an x–y plane. To estimate the distribution of the data, the authors used three methods, including Parzen Windows, Support Vector Data Description (SVDD), and a mixture of Gaussians. Since these methods are boundary methods, the decision boundary of one-class is built for the good data, and the bad data are used to test the models. Results show that on the test dataset, the models identify 98% as bad, i.e., outliers. This work is a good example of finding outliers as faults in a mechanical system as such information is useful in system diagnostics.

Desir et al. [29] present an empirical study to investigate the behavior of their previously proposed One Class Random Forests (OCRF) [17], which is based on the random forest learner and a novel outlier generation procedure. The latter reduces both the number of artificial outliers to create as well as the size of the feature space in which the outliers are generated. In [29] the authors present, in the context of several UCI datasets, a comparative case study of OCRF with a number of reference one class classification algorithms, namely Gaussian density models, Parzen estimators, Gaussian mixture models and one class support vector machines. Their work shows that the OCRF approach with outlier generation performs similar to or better than the above-mentioned reference algorithms. Moreover, their proposed solution demonstrates stable performance in the presence of higher dimension feature spaces, where some other OCC algorithms may not fare well. While not explored in [29], we feel their approach can potentially be investigated for big data where a large number of features is often a problematic issue.

Krawczyk et al. [18] present a multiple classifier system based on weighted one-class support vector machines (OCSVM) and in the context of clustering of the data points in the target class. A multiple classifier system builds an ensemble of classifiers, which in this case is a classifier built upon clusters derived from the pool of instances of the target class. The authors propose "an elastic and efficient framework for this task, which

requires only the selection of several components, namely, the clustering algorithm, individual classifier model, and fusion method [18]." Empirical case studies with several benchmark datasets (including 19 from the UCI library) demonstrate that the proposed method outperforms several OCC methods, including OCSVM for single class and multi-class problems. The authors do not compare with SVDD, an effective OCC approach based on our observation of the various studies explored in the survey. Moreover, all of the case study datasets were relatively small in size, putting the question of model scalability up front.

Lang et al. [20] present a novel approach using a graph-based semi-supervised OCSVM. The application domain is the detection of abnormal lung sounds which is important in the diagnosis of pulmonary diseases and patient monitoring in telemedicine. The proposed approach can describe normal lung sounds and detect the abnormal ones by using a small number of labeled normal instances and a large number of unlabeled instances. "A spectral graph is constructed to indicate the relationship of all the samples, which enriches the information provided by only a small number of labeled normal samples. Then, a graph-based semi-supervised OCSVM model is built, and its solution is provided. Employing the information in the spectral graph, the proposed method can enhance the effect of recognition and generalization which are crucial for the effective detection of abnormal lung sounds." [20]. The performance of the proposed method improves as the number of unlabeled abnormal instances increases.

Krawczyk and Woźniak [22] address the problem of coping with data streams especially in the presence of concept-drift. The authors discuss that OCC is a promising research direction for data stream analysis and can be used for binary classification with only instances from one class, outlier detection, and novelty detection. The authors present a novel weighted OCSVM, which can deal with gradual concept drift. The proposed OCC can adapt its decision boundary to new, incoming data as it also employs a forgetting scheme that boosts the ability of the classifier to follow the changes in the model. Moreover, different strategies for incremental learning and forgetting are proposed, which are evaluated in the context of several case studies. The primary conclusion was the effective usability of the proposed OCC for the problem of data stream classification with the presence of concept drift. It would be interesting to observe the efficacy of the proposed solution in the context of concept drift in big data. A comparison with other popular OCC methods would provide stronger validation for the proposed method.

Das et al. [30] study OCC in the context of applying sensor networks in smart homes to monitor activities of persons with dementia. Monitoring such events is invariably associated with detection errors, which in the context of [30] implies that the person (with dementia) does not correctly complete an activity. The problem of activity completion and errors is formulated as one class classification for outlier detection. Case studies were based on monitoring completion or lack thereof of common household activities, such as vacuuming, dusting, watering plants, answering phones, etc. A problem in completing an activity completely is considered an outlier. Different types of sensors for motion detection and pressure detection vibration are used for data collection. The proposed classification model, Detecting Activity Errors in Real-Time (DERT), is trained with no-error data (i.e., one class) consisting of 580 data points. DERT, based upon OCSVM, is shown to outperform a simple baseline outlier detection approach.

The validation of the proposed approach needs support via a comparative study with other OCC techniques, including SVDD.

Deng et al. [31] focus on the problem of outlier detection in IoT sensor data. They developed the One-Class Support Tucker Machine (OCSTuM), which is an unsupervised outlier detection approach involving the Tucker decomposition technique. Tucker decomposition represents tensors by the production of a core tensor and factor matrices. The case study data is inflicted with the high dimensionality problem, requiring feature subset selection as part of the solution. The authors propose a method (called GA-OCSTuM) applying a genetic algorithm that improves feature selection and outlier detection in OCSTuM. Their work involved several datasets including the Montes Sensor dataset, TAO Project Sensor dataset, Daily and Sports Activities dataset (DSAD), Gas Sensor Arrays in Open Sampling Settings dataset (GSAOSD), and University of South Florida Gait Dataset (USFGD). The OCC training data set is clean without any outliers, but the testing data is mixed with 5% outlier samples. The proposed algorithms were compared with baseline methods, for example OCSVM. The empirical results show that the GA-OCSTuM approach outperforms the baseline methods (including SVDD, R-SVDD, OCSVM, and OCSTuM) for all datasets. The datasets considered in the study is less of a big data problem than a high dimensionality problem in the context of OCC outlier detection. Moreover, genetic algorithms (GAs) are known to have slow computational performance, and the study does not shed light onto GAs impact on the time performance of the proposed solution, GA-OCSTuM.

Gautam et al. [32] developed a Deep Kernel-based One-class Classifier (DKRLVOC) model to reduce object variance and improve feature learning with the aid of a couple of autoencoders. The proposed method is examined over 18 datasets and two real-world datasets, which consist of an Alzheimer's detection by the fMRI dataset and breast cancer detection by pathological images dataset. The proposed minimum variance embedded deep kernel-based one class classification approach consists of three layers: minimum variance embedded kernel-based autoencoder, kernel-based autoencoder, and kernel-based OCC. The approach is compared with three kernel-based extreme learning machine approaches, including OCKELM, VOCKELM [33], and ML-OCKELM. Additional details on these models are presented in [32]. Empirical results show that for the smaller biomedical datasets, the proposed approach performed best with respect to the F1 score. For the mid-size biomedical datasets, the proposed approach is higher than that of ML-OCKELM and OCKELM, but lower than that of VOCKELM. The authors compare the different models in the context of small and medium biomedical datasets, which puts some doubt on how their recommended approaches would perform on much larger datasets, such as big data.

Kauffmann et al. [34] developed a method, One-Class Deep Taylor Decomposition (OCDTD), to explain outliers in one class support vector machines. After the outlier detection process, it is beneficial to provide an interpretive explanation, which indicates that those inputs are responsible for generating the outliers. The explanation maximized the advantage of the structure created by a neural network. In their approach, an OCSVM is fed to a "neuralized" process to reveal the structure of the explanation of outliers. Subsequently, the structure is fed to the Deep Taylor Decomposition and the prediction is propagated backward to the show the inputs that are effective in generating the

outliers. The features that are the most impactful in generating outliers are represented as a heat map. To maximize the advantage of using a neural network, the Layer-wise Relevance Propagation technique is applied, where a collection of propagation rules is applied to propagate the prediction backward [35]. Given backward propagation is used in the neural network environment, a computational time performance study would provide an improved insight into the experimental results and analysis of the study.

Aguilera et al. [36] propose two variants of the k-Strongest-Strengths (kSS) algorithm [37] in the context of OCC. The two algorithms are named OCC-kSS and Global Strength Classifier (gSC) and are evaluated using depression and anorexia benchmark datasets. In addition, the authors introduce mass in the context of the kSS methods as a measure to determine relevance of texts for depression and anorexia in social media data. The algorithms are evaluated with four datasets, named Dep2017, Dep2018, Anx2018, and Anx2019, which are datasets from the 2017–2109 editions of the eRisk shared tasks. The authors conclude that in general the gSC algorithm yielded better results than the OCC-kSS algorithm. The work lacks comparison with other existing OCC approaches, especially several of the ones discussed in this paper.

Wang et al. [38] propose a combined approach to anomaly detection in the context of network intrusion detection systems (NIDS) using a modified version of the KDD intrusion detection dataset, abbreviated as NSL-KDD. The approach combines Sub-Space Clustering (SSC) and OCSVM for anomaly detection for NIDS, and compare the same with the K-means, DBSCAN, and SSC-EA methods [39]. Based on true positive rate, false positive rate, and the ROC curve (for two thresholds), the authors demonstrate their approach yields better performance than the other three methods. The computation time of the proposed method was reported to be higher than that of K-means and DBSCAN. The KDD dataset and its variations are a bit outdated in network security and intrusion detection. There are more current datasets in the area for researchers to explore, which however, is not done in their study.

In the context of autonomous structural health monitoring of bridges, Favarelli and Giorgetti [40] present a machine learning approach toward the automatic detection of anomalies in a bridge structure from vibrational data. They propose two anomaly detection methods, named One-Class Classifier Neural Networks, OCCNN and OCCNN2. The case study data is based on a database of accelerometric data collected for a bridge structure (Z-24) [40]. The OCCNN detects the normal class boundaries of the feature space under normal operating conditions using a two-step approach: coarse boundary estimate and fine boundary estimate. The OCCNN2 is based on combining the two-step approach of the OCCNN method with an autoassociative neural network (ANN) [40]. The two approaches are compared with some existing anomaly detection methods, including: PCA, Kernel PCA, Gaussian mixture model (GMM), and ANN. The OCCNN method demonstrates better accuracy and F1 scores compared to the other approaches; however, the OCCNN2 method demonstrates the best performance with respect to responsiveness, accuracy, and F1 score.

Mahfouz et al. [41] present an OCSVM based model for network intrusion detection, where the model trains on normal network traffic samples forming regions in the n-dimensional feature space where the normal data has a high probability density. Subsequently, data samples that do not occur within or represent those (normal) regions

are tagged as anomalous (i.e., intrusions). While their definition of network instruction anomaly detection is not novel, the paper's primary contribution lies in the network intrusion dataset created and used in its case study. The authors implement the modern honey network (MHN), a centralized server to manage and collect data from honeypots [41]. They create a data set tool using Excel that aggregates data from the separate network monitors (of the different honeypots) into a single data set. With a 70:30 training and testing data split, the accuracy of the proposed model is slightly under 98%. The authors do not compare their approach with the several existing methods for network intrusion anomaly detection.

In a preliminary study, Zaidi and Lee [42] discussed that existing methods for bug triage in software development cannot assign a newly added developer to the bug report. "Bug triage is a software engineering problem in which a developer is assigned to a bug report." [42]. The authors cite existing methods that use social network analysis, topic modeling, mining repositories, machine learning, and deep learning for the task where a developer is assigned to a bug report. However, such methods cannot assign a newly added developer to the bug report. Bug report data from the Eclipse [43] and Mozilla [44] software projects are used in their empirical study. An OCSVM model is built using the positive samples, which can thus detect negative samples. The authors state their empirical results are acceptable, and additional research is warranted for the challenging problem of assigning a newly added developer to the bug report.

Table 1 summarizes the key information of the surveyed work on OCC and outlier detection.

Novelty detection and OCC

As stated previously, outlier detection and novelty detection have a subtle difference in their concepts and applications. In novelty detection, the anomalies are detected in the test dataset while the training dataset does not contain any anomalous data points. In outlier detection, the training dataset may contain both normal and anomalous data points and task is to determine the boundary(ies) between the two, and then apply the boundary on the test dataset which again may contain both normal and anomalous data points.

Clifton et al. [45] utilize an improved OCSVM approach for the novelty detection in the context of identification of patient deterioration based on vital-sign health data, such as respiration rate, blood oxygen saturation, heart rate, etc. The novelty detection model is trained by normal data and then examined to classify test data as normal or abnormal. The training data is collected by monitoring 19 patients, yielding a dataset of 1500 instances. Two models, Gaussian mixture model (GMM) and the OCSVM, are tested with the proposed method and the OCSVM outperforms the GMM model. The case study data is collected from the Step-Down Unit (SDU) which is less acute than data from the Intensive Care Unit. The very small size of the dataset puts some doubt in the generalization of the results and conclusions obtained.

Kemmler et al. [46] proposed a novelty detection framework with a Gaussian Process regression and approximate Gaussian classification in the context of one class classification. Their approach is compared with novelty detection methods of SVDD and Parzen Density Estimation methods. Experiments are performed using datasets from multiple

Table 1 Surveyed works on outlier detection and OCC

#	Paper	Method(s)	Domain
1	Anomaly, novelty, one-class classification: a comprehensive introduction [7]	Gaussian, OCSVM	Masquerader detection
2	One-class classification with extreme learning machine [23]	OCC based on ELM	UCI datasets
3	Outlier detection with one-class SVMs: an application to melanoma prognosis [25]	OCSVM, ANN	Healthcare
4	Patient classification as an outlier detection problem: an application of the one-class support vector machine [27]	OCSVM	Healthcare
5	Outlier analysis and one class classification approach for planetary gearbox diagnosis [28]	Parzen, SVDD, Gaussian	Vibration signal Outlier detection
6	One class random forests [29]	OCRF	UCI datasets
7	Clustering-based ensembles for one-class classification [18]	OCSVM	Miscellaneous datasets
8	Graph-based semi-supervised one-class support vector machine for detecting abnormal lung sounds [20]	Graph-based Semi-Supervised OCSVM	Healthcare
9	On the construction of extreme learning machine for online and offline one-class classification—an expanded toolbox [21]	ELM, OSELM	Miscellaneous datasets
10	One-class classifiers with incremental learning and forgetting for data streams with concept drift [22]	Weighted OCSVM	Streaming data classification
11	One-class classification-based real-time activity error detection in smart homes [30]	OCSVM	Smart home sensors Healthcare
12	An intelligent outlier detection method with one class support tucker machine and genetic algorithm toward big sensor data in Internet of Things [31]	OCSTuM, GA-OCSTuM	Miscellaneous datasets Internet of things
13	Minimum variance-embedded deep kernel regularized least squares method for one-class classification and its applications to biomedical data [32]	Deep Kernel OCC	Miscellaneous datasets Healthcare
14	Towards explaining anomalies: a deep Taylor decomposition of one-class models [34]	OCDTD OCSVM	Miscellaneous datasets
15	Depression and anorexia detection in social media as a one-class classification problem [36]	OCC-kSS, gSC	Healthcare
16	A Hybrid Unsupervised Clustering-Based Anomaly Detection Method [38]	SSC, OCSVM	Network intrusion detection
17	Machine Learning for Automatic Processing of Modal Analysis in Damage Detection of Bridges [40]	OCCNN, OCCNN2	Bridge structural vibration analysis
18	Network intrusion detection model using one-class support vector machine [41]	OCSVM	Novel NIDS Dataset
19	One-class classification based bug triage System to assign a newly added developer [42]	OCSVM	Software bug triage

domains and using different image kernel functions. The case study results demonstrate that the proposed approach either performs similar to, or outperforms, both of the other methods. The application of their approach, especially Gaussian process regression-based OCC scores, would be an interesting study in understanding the class rarity problem in big data.

Beghi et al. [47] investigated an OCSVM approach for novelty detection in HVAC systems. Monitoring likely faults preemptively helps save costs and energy. In such systems,

data for anomalies are rare and usually unavailable. The authors study four types of faults, including condenser fouling, refrigerator leak, evaporator water flow reduction, and condenser water flow reduction. The case study data investigated is obtained from the American Society of Heating, Refrigeration and Air Conditioning Engineers (ASHRAE). The authors incorporate Principal Component Analysis (PCA) with the OCSVM model and observed that the AUC performance improves when combining PCA with OCSVM as compared to using OCSVM alone. The authors do not compare with other novelty detection methods in the literature, which limits the generalization and application validity of their work in a broader sense.

Domingues et al. [48] propose an unsupervised modeling approach for novelty detection based on Deep Gaussian Processes (DGP) in autoencoder configuration. The proposed DGP autoencoder is trained by approximating the DGP layers using random feature expansions, and by conducting stochastic variational inference on the subsequent approximate model. The DGP autoencoder can model the complicated data distribution and helps to propose a scoring method of novelty detection. In the context of seven UCI datasets and four datasets obtained from an international airline service provider, the proposed model is compared to the Isolation Forest and Robust Density Estimation methods. The empirical results demonstrate the proposed model outperforms the other two methods. While the authors experiment on multiple datasets, most of them are relatively small in size, providing little insight into their approach's performance on big data.

Sadooghi and Khadem [49] introduce a preprocessing step to OCSVM toward improving its performance. The context of their work is novelty detection in bearing vibration signals of a rotating system. The preprocessing consists of a novel denoising scheme, feature extraction, vectorization, normalization, and dimensionality reduction, each of which is implemented using a detailed systematic approach. The case study is obtained from the Case Western Reserve University bearing data center, the Tarbiat Modares University test rig data, and the PRONOSTIA platform data. For further details on these data resources, the reader is referred to [49]. The proposed systematic approach shows that nonlinear features alone elevate the performance of novelty detection effectively, including improving the classification rate of OCSVM significantly (up to 95% to 100% in some cases). The proposed modification schemes to OCSVM appear to be tightly coupled with the domain of the case study, and the application of those schemes to other domains is not identified which limits the application of the same to other domains.

Yin et al. [50] investigate and propose an active learning-based approach to improve SVDD in the context of novelty detection. SVDD is one of the most widely used approach for novelty detection, and hence making improvements to it is a good research direction adopted by the authors in this paper. However, SVDD may perform poorly when the amount of data is too large, or the data is of poor quality. Describing the data distribution with a small number of labeled samples has its benefits in machine learning, e.g., one can assure the limited data is noise free and of good quality. The proposed active learning-based approach for SVDD can reduce the amount of labeled data needed, generalize the distribution of the data, and reduce the impact of noise by using the local density to guide the selection process. The case study data includes three UCI datasets (Ionosphere, Splice, and Image Segmentation) and the Tennessee Eastern Process benchmark data. The empirical results show that the active learning based SVDD is

significantly better for the UCI datasets. The active learning is based on supplanting the unlabeled data with data labeled by experts ("specialists"), but little to no information is provided on the expert-based data labeling process. Moreover, while the paper's goal is to incorporate active learning with SVDD for improving its performance on larger datasets, a study on varying the dataset size and investigating the performance of the active learning based SVDD is not performed.

Mohammadian et al. [51] study a novelty detection method based on OCSVM to detect anomalous activities of Parkinson and Autism patients. Patient monitoring with wearable and Inertial Measurement Unit (IMU) sensors has gained considerable attention in Parkinson and Autism Spectrum Disorders (ASD) diseases. Early detection of an unusual physical move by a patient is crucial and helpful in terms of their care and treatment. The authors compensate for the usual deficiency of OCSVM's low performance in large data and noisy data by using deep normative modeling. Because of the limitation of labeled data, a normal model is generated to demonstrate the normal movement of the patient, and large (substantial) changes to the normal movement model are considered as abnormalities. The method is tested on the Freezing of Gait (FOG) and Stereotypical Motor Movements (SMMs) datasets and it is shown that the model is an alternative option of novelty detection in relatively larger data, and it has potential for real-time atypical movement detection. The authors state that their approach is limited only to distance-based novelty detection methods, and thus, is not applicable to density-based novelty detection methods.

Sabokrou et al. [52] proposed a Generative Adversarial Network (GAN) [53] for novelty detection in context of different image and video datasets. The authors propose an end-to-end deep network of the OCC problem. The architecture consists of two modules, R and D. The R module refines the input and gradually injects discriminative rules into the learning process to create the positive and novelty instances (inliers and outliers), while the second module (the detector) separates the positive and novelty instances. Their approach is investigated with two image datasets including the MNIST and Caltech-256 dataset. In addition, they also investigate one video dataset, UCSD-Ped2. For the image datasets, the approach is shown to yield a higher F1-score compared to the Local Outlier Factor (LOF) and Discriminate Reconstructions AutoEncoder (DRAE) methods. For the video dataset, pedestrians in the video data are considered as the positive class and anything else are considered as anomalies. The anomaly detection approach is shown as comparable to some novelty detection methods. In a related work (Sabokrou et al. [54]), propose an adversarial training model to detect outliers in an end-to-end deep learning model. They test their approach on image and video datasets and conclude that the proposed model can effectively learn to detect outliers. The efficacy of their approaches remains to be seen in the context of domains other than image/video data, especially big data.

Oosterlinck et al. [55] present a study in novelty detection in which one-class classification is compared with an expert-based two class classification. The authors investigate an approach to detect fraud in a Telecom company's subscription of a new mobile family plan service. Financial loss to organization and corporations can be quite substantial due to fraud, and detection of those transactions is appealing. An efficient fraud detection system is a pivotal prerequisite of every service provider corporation. To deal with this

issue, human behavior tracking is practical to detect anomalies of human activities and fraud detection. The authors investigated the usefulness of combining synthetic negative samples, which are prepared by experts with positive samples. This work confirms that using expert knowledge in building negative samples and converting the one-class classification to binary classification boost the performance of the classifier. The two-class expert generated sample method outperforms the artificially generated and traditional one-class classification methods. Incorporating experts for decision making during the modeling process can lead to human-errors, and its impact on model performance is not studied in the paper.

Xing and Liu [56] proposed a modified AdaBoost algorithm combined with OCSVM to improve the performance of one-class classification. AdaBoost [57] combined with SVM generally improves performances of binary and multi-class classification problems; however, AdaBoost combined with OCSVM does not improve performance of OCC. The authors present a Robust AdaBoost based Ensemble of OCSVM, which applied the Newton–Raphson technique to change the weights of AdaBoost. The case study data consisted of two synthetic datasets, sine-outlier and square-outlier, and twenty datasets from the UCI repository. The proposed approach is shown to outperform various one-class classification methods, including AdaBoost Ensemble of OCSVM, Random Subspace Method based Ensemble of OCSVM, Clustering-based Ensemble of OCSVM, and OCSVM with Gaussian Kernel. The average performance of the proposed approach outperformed most of the other methods. The scalability of the proposed approach needs further investigation, since all the datasets explored are relatively small.

A one-class GAN (OCGAN) model is proposed by Perera et al. [58] for novelty detection, where the solution is based on learning latent representations of in-class samples using a denoising auto-encoder network. The authors contend that novelty detection involves modeling two types of representations, which include ensuring in-class samples are well represented and ensuring out-of-class samples are poorly represented. They state the latter has not been addressed by prior existing work in novelty detection, and that is where their primary contribution is made. Their proposed model considers modeling both types of representation requirements. The case study data consists of four publicly available multi-class object recognition data sets, including MNIST, FMNIST, COIL100, and CIFAR10 [58]. The proposed model performs better than some existing one-class novelty detection methods for the four data sets considered in the paper. The comparative work between the different techniques lacks statistical verification and validation with respect to the models' performances. Moreover, the applicability of the proposed approach to non-image datasets is not discussed by the authors.

In the context of image novelty detection, Zhang et al. [59] propose their "adversarially learned one-class novelty detection with confidence estimation" model. The authors contend that most existing methods for novelty detection, especially those using deep learning technology, are not end-to-end and tend to be overconfident in the novelty detection predictions. The proposed model consists of two modules: representation module and detection module, which are adversarially modeled to collaboratively train and learn the inlier distribution of the data corpus. In addition, the model uses confidence-based estimation to ensure higher efficacy in its predictions. The model is examined with four publicly available image datasets, namely: MNIST, FMINST, COIL100,

Table 2 Surveyed works on novelty detection and OCC

#	Paper	Method(s)	Domain
1	Identification of patient deterioration in vital-sign data using one-class support vector machines [45]	GMM, OCSVM	Healthcare
2	One-class classification with Gaussian processes [46]	Gaussian, SVDD	Generic datasets
3	A one-class SVM based tool for machine learning novelty detection in HVAC chiller systems [47]	PCA, OCSVM	HVAC systems
4	Deep Gaussian Process autoencoders for novelty detection [48]	DGP	UCI datasets
5	Improving one class support vector machine novelty detection scheme using nonlinear features [49]	OCSVM	Vibration signals
6	Active learning-based Support Vector Data Description method for robust novelty detection [50]	Active Learning based SVDD	UCI datasets
7	Novelty detection using deep normative modeling for imu-based abnormal movement monitoring in Parkinson's disease and Autism Spectrum Disorders [51]	OCSVM	Patient monitoring, healthcare
8	Adversarially learned one-class classifier for novelty detection [52, 54]	OCC GAN, LOF, DRAE	Image processing, video processing
9	From one-class to two-class classification by incorporating expert knowledge: Novelty detection in human behaviour [55]	Expert-based Two-Class Classification	Fraud detection
10	Robust AdaBoost based ensemble of one-class support vector machines [56]	AdaBoost, OCSVM	UCI datasets
11	OCGAN: One-class novelty detection using GANs with constrained latent representations [58]	OCGAN	Image processing, object recognition
12	Adversarially learned one-class novelty detection with confidence estimation [59]	Adversarial OCC	Image processing

and CIFAR10, and is compared with several existing methods for novelty detection [59]. The authors conclude that their proposed model outperformed several existing one-class novelty detection methods. Moreover, an ablation study indicated that each module of the proposed model is critical in its functionality. Similar to the previous study, the comparative work between the different techniques in this study lacks statistical verification and validation with respect to the models' performances.

Table 2 summarizes the key information of the surveyed work on OCC and novelty detection.

Deep learning and OCC

Kim et al. [60] present a novel deep learning model that involves the Restricted Boltzmann Machine and a modified SVDD, called Deep SVDD (DSVDD). The latter involves hidden layers where each layer has k SVDD nodes for a typical k-class

problem. All layers in the network include hidden layers and the last layer extracts the decision for the test data samples. The context of the proposed study is reducing the overfitting problem by combining deep learning's representation capability and the generalization performance of SVDD. The case study data includes three UCI datasets, including Wisconsin Breast Cancer, Climate Model Simulation Crashes, and Pima Indians Diabetes datasets. A confidence measure of 0.55 is set to filter the samples as suspicious or confident. Empirical results demonstrate DSVDD yielded better performance than SVM, SVDD, and Deep Belief Networks. As the datasets used in the study are relatively small, the performance goodness of the different methods needs to be investigated in the context of big data, i.e., does the DSVDD model perform well with datasets that are very large and are severely imbalanced.

Erfani et al. [61] address the high-dimensionality problem in context of anomaly detection. The presence of irrelevant features can hide anomalous samples, making their detection difficult. Anomaly detection in the context of high-dimensional data poses some issues, such as exponential search space, data-snooping bias, and irrelevant extracted features. Toward addressing such issues, the proposed method combines a deep learning model with a one-class classifier. The paper presents an unsupervised method of anomaly detection called DBN-1SVM, which benefits from DBN for extracting robust features and OCSVM for training purposes. Since DBNs are efficient in the learning of complex high dimensional datasets and it reduces the dimensionality of the data, it is advantageous to combine it with OCSVM, i.e., the results of the non-linear dimensional reduction of DBNs are fed into OCSVMs for learning. The authors performed a test on the Forest Adult Gas Sensor Array Drift, Opportunity Activity Recognition, Daily and Sport Activity, and Human Activity Recognition using the Smartphones datasets as real-life datasets and the Banana and Smile datasets as synthetic ones. The DBN-1SVM is compared to Plane-based one-class SVM (PSVM), SVDD, and an AutoEncoder (AE) with the choice of linear or RBF kernels. Experiments reveal that DBN-1SVM is more robust compared to other methods and yields better performance. Given the good results obtained by the proposed approach, it would be interesting to investigate the approach's efficacy and efficiency in the context of other domains as well as with very large and complex datasets.

Sun et al. [62] present a deep one-class classification (DOC) learner for detecting anomalies in surveillance videos. Automated detection of abnormal events (anomalies) in video surveillance data is an important problem for intelligent monitoring and alarm systems. The DOC model is the integration of an SVM and Convolutional Neural Network (CNN) [63], where the SVM performs as an optimizer in addition to discriminating between normal and abnormal objects. The proposed model, DOC, is compared to four other methods, including mixtures of dynamic textures [64], sparse reconstruction cost [65], appearance and motion deepnet [66], and the approach for anomaly detection in surveillance videos by Adam et al. [67]. The case study data are the UCSD pedestrian datasets, Ped1 and Ped2. The different approaches were compared at both the frame-level and the pixel-level. Results for the pixel-level experiments show the proposed approach, DOC, outperforms the others. At the frame-level, both DOC and the appearance and motion deepnet approach yield comparable performance. It would be interesting to compare the proposed method with

SVDD and Deep SVDD, as they are commonly very effective with OCC for anomaly detection.

Zhang et al. [68] evaluate a one-class CNN classifier model by comparing it to a typical two-class classifier CNN model. The model represents an integration of OCC and Deep Neural Network, and where the hybrid model is applied to image defect detections. The general architectures of the two models are similar, except that the two-class classifier CNN model has an additional two-neuron output layer fully connected to the last layer of the one-class classifier. The loss function of the one-class classifier is the contrastive loss function while that of the two-class classifier is the softmax loss function. The case study dataset consists of a small and limited number of electronic component images. The defect images are of various types, including incomplete white part, incomplete gray part, deformation, spots, scratches, pits, etc. The manually labeled dataset consists of 1090 images with 600 non-defective images and 490 defective images. The two models are applied to the electronic component dataset, and the one-class classifier outperforms the two-class classifier and is more robust as well. An OCC scheme should be compared with other OCC approaches for gauging an effective and fair comparison of the proposed OCC scheme; however, that is not done in this study.

Gutoski et al. [69] investigate an image anomaly detection classifier using the Deep Embedded Clustering (DEC) method, which is based on a "Deep Autoencoder." Images that do not contain anomalies are considered as normal, and only normal data are used in their one-class classification studies. Model training is started with the Stacked Denoising AutoEncoder (SDAE) with normal instances and its output is entered into the Deep Embedded Clustering optimizer which also determines the cluster center simultaneously. The case study data consists of three datasets: (1) STL-10, which consists of 96×96 pixels color image data divided into 10 classes, each with 800 images of which 500 are labeled instances; (2) MNIST, which consists of 60,000 images as training data and 10,000 images as test data, where each image is in gray scale and 28×28 pixels; (3) NOTMINST, which is a printed character dataset consisting of 10 classes of letters from A to J, and contains 200,000 images as training data and 10,000 images as test data with the images in gray scale and 28×28 pixels. The authors conclude that applying SDAE followed by Deep Embedded Clustering improves accuracy of the one-class classification, i.e., performance of anomaly detection in image classification.

Ruff et al. [70] introduce the Deep Support Vector Data Description (Deep SVDD) method in the context of anomaly detection and one-class classification. As noted earlier the performance of the commonly used OCC-based anomaly detection is degraded with high dimensional datasets, largely due to the high computational costs and data complexity. The Deep SVDD model parameters are optimized using the Adam [71] and Stochastic Gradient Descent (SGD) approaches. SGD leads to the parallel execution of the training data batches and helps with scalability of Deep SVDD. The authors conduct two case studies in this work. The first one involves the MNIST and CIFAR-10 datasets. The second one focuses on the possibility of adversarial attacks, such as boundary attack [72], in the context of the German Traffic Sign Recognition Benchmark (GTSRB) [73] dataset. The Deep SVDD approach is compared to some OCC approaches, such as SVDD, kernel density estimation, Isolation Forest, deep convolutional autoencoder, and anomaly detection based on Generative Adversarial Networks (GANs) [74]. For the

MNIST and GTSRB datasets, the Deep SVDD approach yields the best performance. However, for the CIFAR-10 dataset, the kernel density estimation approach and the SVDD approach outperform Deep SVDD. Because of the mixed results across the three datasets the efficacy of Deep SVDD is not quite clear in the case of image datasets as well as other domain datasets.

Chalapathy et al. [75] present a deep neural network-based approach to one-class classification in the context of anomaly detection. It is a hybrid approach based on the work of Ruff et al. [70]; however, the authors replace SVDD with OCSVM in their overall anomaly detection approach. The hidden layer of the network performs the data representation, while in the subsequent stage OCSVM detects any anomalies that occurred. The case study data included the German Traffic Sign Recognition Benchmark dataset [73], MNIST, CIFAR-10, and a synthetically produced dataset. The proposed strategy is a different than other approaches which use a hybrid method of learning deep features using an autoencoder and then feeding the features to a separate anomaly detection method like OCSVM or OCSVDD. The case study results show that the proposed approach performs similar to state-of-the-art OCC methods. Some of the methods that the proposed approach was compared with include OCSVDD, Isolation Forest, kernel density estimation, deep convolutional autoencoder, etc.

Schlachter et al. [76] investigate Intra-Class Splitting (ICS) in the context of using deep learning networks for OCC. The approach splits data from one class (normal class) into two subsets, "typical normal" and "atypical normal." By splitting the normal class into two subsets, the approach uses a binary loss and defines an auxiliary subnetwork for distance-based constraints. Three different rules are considered for the distance-based constraints for the subsets: small distances with typical normal instances; large distances between typical and atypical normal instances; and large distances among atypical normal instances. The OCC uses an arbitrary deep neural network in which the first layer performs the feature extraction, and the subsequent layer represents the classification subnetwork. A third subnetwork, the distance subnetwork, is used only during the training process to satisfy constraints of the two subsets. The case study data consists of applications to the MNIST, Fashion-MNIST, and CIFAR-10 datasets, for which the proposed approach is compared to other OCC methods, including OCSVM with RBF kernel, Isolation Forest, ImageNet with OCSVM, Naïve neural network without ICS, neural network with ICS but without the subset distance constraints, and Deep SVDD [70]. The comparative results determine that the proposed method performs better than most of the other OCC methods.

Perera and Patel [77] present a deep-learning method for one-class transfer learning where labeled data from an unrelated task is used for feature learning in OCC (for a survey study on transfer learning please refer to Pan and Yang [78]). The authors introduce a joint optimization framework based on two loss functions—compactness loss and descriptiveness loss. The former is for assessing the compactness of the class under consideration in the learned feature space, while the latter assesses the descriptiveness based on the use of an external multi-class dataset. The neural networks backbone architectures are Alexnet and VGG16, which are two successful pre-trained neural networks [77]. The case study involves studying datasets addressing three scenarios respectively: image novelty detection, abnormal image detection, and active authentication. For the

image novelty detection, the Caltech 246 dataset is used with 30,607 images and 256 classes. Again, the performance of both neural networks showed improvement over other OCC methods. The 1001 abnormal objects datasets with six classes are used in the abnormal detection study, in which performance of both neural networks showed improvement over other OCC methods. In the case of the active authentication study, the UMDAA-02 mobile AA dataset [79] is used, which contains multi-modal sensor face observations from 48 users. The proposed approach did not show promising results for this case for both neural networks.

Burlina et al. [80] present an unsupervised approach for diagnosing myopathic disease (myositis is a rare form of disease in myopathies), in which they investigate deep learning and one-class novelty detection for the Myositis3K benchmark dataset. The latter, developed by the authors, is composed of ultrasound images acquired of seven muscle groups imaged bilaterally per subject [80]. The complete dataset consisted of images of size 476×476 from a total of 89 subjects, including 35 normal/control and 54 with myositis (19 with inclusion body myositis, 15 with polymyositis, and 20 with dermatomyositis). This acquisition resulted in a dataset of 3586 images. The general approach is as such: by applying deep feature embedding, a new ultrasound image representation is built; performing PCA to reduce the dimensionality of the images; applying t-Distributed Stochastic Neighbor Embedding (t-SNE) [81]; and finally, the novelty detection scoring algorithm is applied to detect the anomalies. The proposed approach is compared to other one-class novelty detection methods, such as Isolation Forest (IF), Elliptic Envelope (EE), Local Outlier Factor (LOF), OCSVM, and Generative Adversarial Network for anomaly detection (GANomaly). The ultrasound images are partitioned using two ways: image-based partitioning (IP) and patient-based partitioning (PP). The experiments found promising results for the use of deep learning techniques for novelty detection applied to myositis. The best results were obtained by OCSVM, followed by EE, IF, and LOF. Deep learning is applied only to part (i.e., deep feature embedding) of the modeling process instead of the entire process for unsupervised learning for detecting myopathic disease.

Ghafoori and Leckie [82] present the Deep Multi-sphere Support Vector Data Description (DMSVDD) approach, which assumes the training dataset can have multiple data distributions compared to the conventional assumption of a singular data distribution by most classification schemes. Applying the k-means clustering method [83], data is mapped into clusters and projection from input space into hyper-spherical clusters considering the minimum volume of the sphere as a metric. Parametric mapping is done using an autoencoder which helps to reduce the reconstruction error. The case study datasets are MNIST, CIFAR10, and MobiAct, where the latter includes data for 67 persons, capturing 11 normal activities and four abnormal activities. The following versions of the MNIST datasets are considered: MNIST0, where the 0 digit is normal and other digits are abnormal; MNIST01, where the 0 and 1 digits are normal and other digits are abnormal; and MNIST013, where the 0, 1, and 3 digits are normal and other digits are abnormal. The proposed method is compared with SVDD, kernel density estimation (KDE), RBF, IF, OCSVM, Deep SVDD (DSVDD), and Deep Convolutional AutoEncoders (DCAE). Since OCSVM, SVDD, and RBF converge to similar solutions, results of only SVDD are reported. For the MNIST0 dataset, DCAE, DSVDD, and DMSVDD have

similar accuracy. For the MNIST01 dataset there is a slight performance decrease for DCAE and DSVDD, and considerable performance decrease for SVDD, KDE, and IF. The proposed approach, DMSVDD, showed promising results for MNIST01. For the MNIST013 dataset, DMSVDD yielded the best result followed by DCAE; however, a performance decrease was observed for SVDD, KDE, IF, and DSVDD. For the CIFAR-10 dataset, the best performance was observed with SVDD, followed by DSVDD and then DMSVDD, indicating that the proposed method performed poorly compared to a non-deep method. For the MobiAct dataset, DMSVDD performed the best followed by DCAE, while poor performances were observed with SVDD, KDE, and IF. The proposed method is sensitive to the underlying dataset, as observed in the results of the paper. Because of the mixed results across the three datasets the efficacy of DMSVDD is not quite clear.

Liu et al. [84] present a One-Class Presentation Attack Detection (OCPAD) method for Optical Coherence Technology (OCT) images based on fingerprint presentation attack detection. The dataset used in their empirical study consists of 121 (121×400 B-scans) presentation attacks from 101 materials and 233 (233×400 B-scans) bonafides (real fingerprints) from 137 subjects. The proposed model is compared with existing PAD methods, including a feature-based method, a supervised learning-based method, and a one-class GAN approach. The empirical studies demonstrate that the proposed method outperforms those three methods. More specifically, the proposed method obtains a true positive rate (TPR) of 99.43% when the false positive rate (FPR) is 10%, and a TPR of 96.59% when the FPR is 5%. While ROC curves for performances of the three methods are shown in the paper, the AUC values are not reported. Moreover, a statistical verification and validation study is not provided toward the significance of performance differences between the three methods.

Cao et al. [85] argue that existing approaches of OCC approaches for outlier detection, while being robust to the Gaussian noises, are less effective in detecting large number of outliers. To address this problem, they propose a maximum correntropy criterion-based OCC ELM model (MC-OCELM) and the model is further extended to a hierarchical network to improve its capability in characterizing complex data, where the extended model is abbreviated at HC-OCELM. Their empirical case study is based on eight UCI benchmark data sets to illustrate the efficacy of the proposed model, which is then compared with several existing methods, including Parzen, Naïve Parzen, K-means, K-centers, 1-NN, KNN, AutoEncoder, PCA, minimum spanning tree (MST) based OCC, Minimax probability machine (MPM), SCDD, linear programming dissimilarity-data description (LPDD), SVM, and OC-ELM. The two proposed methods outperform the other methods with respect to the F-score metric. The paper also examines the CIFAR-10 data set, for which the proposed methods are compared against Deep SVDD, IF, Deep Convolutional AEs, kernel density estimation, Soft-Boundary Deep SVDD, and Deep Convolutional GAN. The HC-OCELM model provides better results for seven of the ten CIFAR-10 data sets, which Deep SVDD provides better results on two data sets. Overall, the authors conclude their proposed method is better than the other methods in their empirical study.

Fontella-Romero et al. [86] present the distributed singular value decomposition autoencoder (DSVD-AUTO) that facilitates learning in distributed scenarios without

the need to share raw data. Moreover, data privacy preservation is addressed by the proposed method. Sharing data in distributed learning environments and data privacy are two issues facing big data analytics. The case study includes 10 datasets, with sizes ranging from 420 samples to 11,000,000 samples, with instances with missing data being omitted in the modeling. The proposed method is compared with four other methods for OCC anomaly detection, including LOF, OCSVM, AUTO-NN, and APE [86]. While the mean AUC of the OCSVM method yielded better results than those of the proposed method, the authors contend the former requires tuning of multiple hyperparameters. However, additional investigation is required before the efficacy of the proposed method can be stated confidently from a generalization point of view.

Moustafa et al. [87] present a Distributed Anomaly Detection (DAD) system to detect zero-day attacks in edge networks. The system uses Gaussian Mixture-based Corren-tropy, which is an OCC model. Important data features are selected using PCA and are then passed to the proposed system to detect anomalies. The datasets included in their case studies include the NSL-KDD and UNSW-NB15 datasets. Using 350,000 data samples from both datasets, the proposed model is compared with five anomaly detection approaches, including Multivariate Correlation Analysis (MCA), Triangle Area Near-est Neighbors (TANN), Geometric Area Analysis (GAA-ADS), Outlier Dirichlet Mix-ture (ODM), and Attack Detection-based Convolutional Neural Network (AD-CNN). The authors conclude that their approach outperforms these five approaches for anom-aly detection, with respect to detection rate, false positive rate, and processing time. It would be interesting to investigate the performance of the proposed model using the complete datasets (NSL-KDD and UNSW-NB15) instead of using a limited-size sample from both datasets. Such a study would be considered an application of the proposed approach to big data; however, this is not done in the above paper.

Pourreza et al. [88] present a GAN-based deep learning approach for one-class clas-sification. The authors state that training and implement GANs for OCC is often cum-bersome. Toward simplifying the process, they treat the OCC problem as a binary classification task, in which two deep neural networks (generator and discriminator) are trained in a GAN setting on the normal samples. During the early stages of the training process, the generator is likely to fail to produce the normal samples correctly, which is then consequently considered an anomaly sample generator. By doing so, two sets of samples are generated by the deep neural networks—normal and anomalies. Subse-quently, a binary classifier is trained on these generated samples (using both normal and anomaly samples) for anomaly detection. The proposed G2D model consists of three primary modules [88]: (1) irregularity generator network, (2) critic network, and (3) detector network. The case study involves image anomaly detection and video anomaly detection and includes the following data sets [88]: UCSD, MNIST, and Caltech-256. The proposed approach is shown to be competitive with the compared approaches for outlier detection, including R-graph, REAPER, OutlierPursuit, LRR, SSGAN, and ALOCC [88].

We now present a very brief coverage of some other works in OCC involving deep learn-ing in some form or another. We do so because the surveyed works presented in this sec-tion represents a good cross section of works that involved deep learning in OCC. Chong et al. [89] address the problem of hypersphere collapse (or mode collapse) in the context of Deep SVDD. This problem tends to occur if the architecture of the model does not comply

with specific architectural constraints, such as removal of bias terms, thereby constraining the adaptability and performance of the model. Variations of the approach in both problem context and application context are presented by Tan et al. [90] and Golan et al. [91]. A Deep Semi-Supervised Anomaly Detection (Deep SAD) is presented and evaluated by Ruff et al. [92]. The focus of the study is to improve performance of the traditional unsupervised approach to anomaly detection with the aid of some labeled instances. Goyal et al. [93] address the mode collapse problem by proposing and evaluating a deep robust one-class classification approach, which is motivated by the assumption that the interesting class lies on a locally linear low dimensional manifold.

Table 3 summarizes the key information of the surveyed work on Deep Learning and OCC.

Surveys on OCC

In their 2014 survey, Khan and Madden [9] provide "a unified view of the general problem of OCC by presenting a taxonomy of study for OCC problems, which is based on the availability of training data, algorithms used, and the application domains applied." The literature works that are grouped based on availability of training data involve learning only with positive data or learning with unlabeled data, positive data, and some outlier instances. The works grouped based on methodology used involved approaches based on one-class support vector machines (OSVMs) or non-OSVM approaches (e.g., classification ensemble methods). The third group in the taxonomy, application domain, involves whether OCC was applied for text analysis/classification or applied to other application domains. We believe the taxonomy used by the authors is apt; thus, we do not provide a similar categorization of the surveyed works in our study. However, we do feel the survey presented in this study provides a much more updated collection of OCC related works compared to those covered in [9], particularly since several works in OCC have been published since 2014.

Pimentel et al. [94] present a review of novelty detection as an OCC problem. The surveyed works are categorized into five groups, including probabilistic-based, distance-based, reconstruction-based, domain-based, and information theoretic-based approaches. The application domains surveyed in their work include electronic information technology security, healthcare informatics, medical diagnostics and monitoring, industrial monitoring and damage detection, image processing, video surveillance, text mining, and sensor networks [94]. When selecting novelty detection methods, various considerations were made, such as training data availability, application domain, and data characteristics including data dimension, data format, and data continuity. Compared to [94], this survey paper summarizes works in OCC-based outlier detection and deep learning in OCC, in addition to OCC-based novelty detection. Thus, we provide a more expansive survey study on one-class classification compared to the survey work presented in [94].

Discussion

Outlier detection and novelty detection are observed as the primary areas of use for one-class classification in our surveyed works. However, important issues in data mining and machine learning, such as class imbalance problems in big data as well as other associated big data related problems are not addressed in the surveyed works. These and their related topics could be good application areas of OCC, e.g., class rarity detection and

Table 3 Surveyed works on deep learning and OCC

#	Paper	Method(s)	Domain
1	Deep learning with Support Vector Data Description [60]	Deep SVDD	UCI datasets
2	High-dimensional and large-scale anomaly detection using a linear one-class SVM with deep learning [61]	DBN-1SVM	Anomaly detection
3	Abnormal event detection for video surveillance using deep one-class learning [62]	DOC, SVM, CNN	Video surveillance
4	The application of one-class classifier based on CNN in image defect detection [68]	One-Class CNN, CNN	Image processing
5	A clustering-based deep autoencoder for one-class image classification [69]	SDAE, Deep Embedded Clustering	Image processing
6	Deep one-class classification [70]	Deep SVDD	Image processing object recognition
7	Anomaly detection using one-class neural networks [75]	Deep OCC, OCSVM, OCSVDD, IF, KDE	Image processing object recognition
8	Deep one-class classification using intra-class splitting [76]	OCSVM-RBF, IF, ImageNet-OCSVM, NN with/without ICS, Deep SVDD	Image processing
9	Learning deep features for one-class classification [77]	Deep OCC, Alexnet, VGG16,	Image processing
10	Deep embeddings for novelty detection in myopathy [80]	IF, EE, LOF, OCSVM, GANomaly	Healthcare
11	Deep multi-sphere support vector data description [82]	DMSVDD	Image processing human activity
12	One-class fingerprint presentation attack detection using autoencoder network [84]	OCPAD, PAD, OCGAM	Image processing attach detection
13	Maximum Correntropy criterion-based hierarchical one-class classification [85]	OC-ELM, Parzen, K-means, K-centers, 1-NN, KNN, AE, PCA, MS-OCC, MPM, SCDD, LPDD, SVM,	Image processing
14	DSVD-autoencoder: A scalable distributed privacy-preserving method for one-class classification [86]	AUTO-NN, LOF, OCSVM, APE	Miscellaneous datasets Privacy preserving
15	DAD: A distributed anomaly detection system using ensemble one-class statistical learning in edge networks [87]	DAD, MCA, TANN, GAA-ADS, ODM, AD-CNN	Network intrusion detection
16	G2D: generate to detect anomaly [88]	G2D, GAN, DNN, LPR, R-graph, REAPER, Outlier Pursuit, SSGAN, ALOCC	Image/video processing

influence of severe class imbalance in big data. For example, when data is highly imbalanced, detecting classes of interest could potentially be done more efficiently with OCC as the latter focuses on detecting the positive class(es) of interest.

Among the surveyed works, OCC is frequently applied in the analysis of biomedical data. The Parzen Windows approach is used often; however, kernel-based methods such as SVDD and OCSVM perform relatively better. In outlier detection, OCSVM appears to be a prominent choice. It has been applied to detect outliers in various diseases, including tuberculosis, depression, and Alzheimer's. In the context of patient monitoring applications, novelty detection helps in early detection of the disease. Since the

abnormal activities are rare data points, the datasets related to these applications are severely imbalanced. Moreover, labeling data for supervised learning is time-consuming and difficult, making supervised one-class learning not practical in this context.

In the neuroimaging domain, data has high dimensionality in the feature space and with relatively few one class samples available. Under such circumstances, traditional learners tend to perform poorly. Moreover, in cases where the clinical diagnosis data are not clear or where the interest is in identifying subgroups of patients, traditional binary classification tends to perform poorly. High class imbalance is the key problem associated under such circumstances. OCC has been applied to address classification problems in dementia detection and activity completion performance of the elderly. For these domain-specific problems, OCSVM yields promising results. It is observed that applying ensemble method along with OCSVM does not boost performance.

Among the one-class classifiers, SVDD is well suited for outlier detection and loss reduction of outliers. However, SVDD may show overfitting in the case of noisy or uncertain data. In some data types, like sensor data, which may have sampling error, the target hypersphere is not optimized. SVDD performs well in high dimensional feature space without any prior information about the data. In the field of knowledge discovery in databases, SVDD, Parzen Windows, LOF, and LOCI are commonly used outlier detection approaches using OCC. Among them, SVDD and LOF have yielded better results in related literature.

As noted previously, another application of one-class classification is novelty detection, where outliers are detected in the test dataset while the training dataset is does not contain any outliers. In the case of heterogeneous or high dimensional data, novelty detection is a solution to detect anomalies. In general, novelty detection is an unsupervised learning solution when labeled data is difficult to obtain.

Since industrial data distribution is not Gaussian, Gaussian process regression is beneficial for one-class classification in terms of mitigating implementation overhead and improving performance. With the help of deep learning methods, many supervised learning problems have been addressed in the context of OCC. However, in this area considerable research remains to be done regarding choice of the unsupervised learning method. To address the problem of overfitting in deep neural networks used for novelty detection, Deep Gaussian Processes are advantageous and empower the deep model to train overly high complex data or mixed-type features. One of the preferred methods to monitor the reliability and safety of moving parts of industrial machines is novelty detection since plenty of high-quality data is available. The novelty detection methods require preprocessing to improve performance; particularly, nonlinear features because nonlinear vibration signals affect the performance dramatically.

Big data obtained by IoT requires profound consideration in terms of computing time and complexity. Since the accuracy of outlier detection in such systems is related to the number of features extracted, finding perfect subsets of features is critical. With most one-class classification methods, however, vector-based representation is applicable and IoT data is the most suitable for tensor-based methods. For such complex and high dimensional datasets, tensor-based one-class classification methods surpass the vector-based methods. One of the applicable methods to deal with insufficient labeled data in big data, such as industrial data, is one-class active learning which accepts unlabeled

data and requires relatively fewer labeled data points. In HVAC systems, finding system faults is hard due to the variability of data in such systems, and the anomalous condition is mistaken with normal conditions. PCA is observed to be a practical solution for ignoring those mistaken conditions and when combined with OCSVM shows promising performance.

Fraud detection is essential in all financial services and transactions. To detect the fraudulent transaction, one of the best solutions is human behavior tracking and using a one-class classification method. However, such a problem suffers from a few numbers of negative samples; thus, generating a negative sample either artificially or using experts is one of the practical ways to overcome the issue of sample shortage. The performance of the model based on expert knowledge in one-class classification is significantly different especially because of the high diversity of human behavior in data creation. In general, the most popular one-class classification method for novelty detection is OCSVM. However, as noted earlier, ensemble learning with OCSVM does not improve learning. One of the practical and feasible methods to raise the performance is loss function modification such as a combination of two or more loss functions.

SVDD has high generalization performance, and it is robust with outliers in the dataset. Furthermore, SVDD is stable under noisy data conditions. However, the overfitting problem is still relatively prevalent and an important topic of continuing research. Using a deep network method adopted to SVDD algorithm boosts the generalization performance. The deep layers embedded with SVDDs grants representational power to the combined model and it boosts performance. In addition to advantages of the deep learning with one-class classification, deep SVDD has the problem of having some constraints and limitations like mode collapse or hypersphere collapse. While regularization by introducing random noise may help, this problem needs further continuing research.

There is some weakness regarding traditional one-class classification methods such as low performance in high dimensional data due to sensitive hyper-parameters [76]. Furthermore, they need expert labeled data which is time-consuming and cumbersome to obtain. To improve the performance and speed of anomaly detection in complex datasets, using OCSVM-based hybrid model with deep neural networks is promising. Integrating SVM and deep neural networks provides accuracy improvement in anomaly detection. Among linear and RBF kernels, some works confirm that non-linear kernels are more efficient and provide improvements. While OCSVM has been shown to be beneficial for OCC, it suffers from high computational complexity due to its iterative nature during modeling and needs some kind of regularization such as kernel regularized least squares.

In anomaly detection for computer vision, one of the best methods to alleviate the problem of high dimensionality is the application of deep learning methods along with optimization algorithms such as Stochastic Gradient Descent (SGD) with one-class classification. In this way, the optimization algorithm brings scalability and online learning to the hybrid model [70]. Incorporating deep learning with one-class anomaly detection and novelty detection shows promise in the detection of abnormal instances; however, continuing research in a variety of domains is needed to further verify this potential.

There are practically no studies exclusively on big data and one-class classification. A recent study [95] developed an edge-processing unit in the context of railway condition

monitoring systems. The unit consists of two components: a data classification model and a data transmission unit. By using data sampling instead of using the entire big data corpus available, we feel this work is not an example of OCC applied for big data analytics.

The problem of high dimensionality is a big data problem in certain domains; however, in the surveyed works the publicly available image datasets are not big data. While class imbalance remains at the core of the one-classification problem, datasets with severe class imbalance have not been studied in the literature. Nor has the problem of class rarity in big data and one-class classification. While the application of deep learning to OCC problems is actively being investigated, more research needs to be undertaken in this area. For example, the presence of noise in big data and other datasets needs better solutions than those observed in the literature. Moreover, feature engineering in the context of one-class classification with big data is a promising area of research. Innovative approaches, such as transfer learning, need additional attention in the context of one-class classification solutions. Investigating boosting-based methods in the context of OCC may yield interesting results; particularly, since it has shown extensive promise in traditional classification problems.

Conclusion

This study presented a large survey of one-class classification methods and approaches, including domain-specific applications, presented in the literature over the last decade, i.e., 2010–2021 (May 2021). Our survey categorizes the different works into three categories: outlier detection and OCC, novelty detection and OCC, and deep learning and OCC. The paper examined selected works on one-class classification such that a good cross section of approaches, methods, and application domains is represented in the survey. In addition, the paper highlighted potential applications of OCC in the context of various problems faced when dealing with big data.

While not meant to be an exhaustive survey, to our knowledge this is the most recent survey paper on one-class classification works in the literature. One area that has been largely omitted in one-class classification is the application context of big data and its inherently associated problems, such as severe class imbalance, class rarity, noisy data, and feature engineering. We feel this survey paper on one-class classification will be much appreciated by researchers and scientists working in the big data area of data mining and machine learning. Moreover, regarding one-class classification, additional research is needed in areas such as transfer learning [96], deep learning, feature engineering and selection [97], data quality and data collection, active learning [98], semi-supervised learning, and for additional domain applications [99].

Abbreviations
OCC: One-class classification; SVM: Support vector machine; OCSVM: One-class support vector machine; ELM: Extreme learning machine; ANN: Artificial Neural Networks; OSELM: Online sequential extreme learning machine; fMRI: Functional Magnetic Resonance Imaging; SVDD: Support Vector Data Description; PCA: Principal Component Analysis; LVQ: Learning Vector Quantization; OCClustE: One-Class Clustering-based Ensemble; OCLP: One-Class Linear Programming; UCI: University of California Irvine; WEKA: Waikato Environment for Knowledge Analysis; SNR: Signal to Noise Ratio; OCRF: One-Class Random Forests; DERT: Detecting Activity Errors in Real-Time; OCSTuM: One-class Support Tucker Machine; GA: Genetic algorithm; DSAD: Daily and Sports Activities Dataset; GSAOSD: Gas Sensor Arrays in Open Sampling Settings Dataset; USFGD: University of South Florida Gait; DKRLVOC: Deep Kernel-based One-class Classifier; OCKELM: One-Class Kernel Extreme Learning Machine; OCDTD: One-Class Deep Taylor Decomposition; kSS: K-Strongest-Strengths; OCC-kSS: One-Class Classification k-Strongest-Strengths; gSC: Global Strength Classifier; NIDS: Network intrusion detection

systems; SSC: Sub-Space Clustering; OCCNN: One-Class Classifier Neural Networks; GMM: Gaussian mixture model; MHN: Modern honey network; SDU: Step-Down Unit; ASHRAE: American Society of Heating, Refrigeration and Air Conditioning Engineers; AUC: Area under the ROC curve; ROC: Receiver operating characteristic; DGP: Deep Gaussian Process; IMU: Inertial Measurement Unit; ASD: Autism Spectrum Disorders; GAN: Generative Adversarial Network; DRAE: Discriminate Reconstructions AutoEncoder; LOF: Local Outlier Factor; OCGAN: One-Class Generative Adversarial Network; DSVDD: Deep Support Vector Data Description; DBN-1SVM: Deep belief network support vector machine; PSVM: Plane-based one-class support vector machine; DOC: Deep one-class classification; CNN: Convolutional Neural Network; DEC: Deep Embedded Clustering; SDAE: Stacked Denoising AutoEncoder; SGD: Stochastic Gradient Descent; GTSRB: German Traffic Sign Recognition Benchmark; OCSVDD: One-Class Support Vector Data Description; ICS: Intra-Class Splitting; t-SNE: T-Distributed Stochastic Neighbor Embedding; IF: Isolation forest; EE: Elliptic Envelope; DMSVDD: Deep Multi-sphere Support Vector Data Description; KDE: Kernel Density Estimation; RBF: Radial basis function; DCAE: Deep Convolutional AutoEncoders; OCPAD: One-Class Presentation Attack Detection; OCT: Optical Coherence Technology; TPR: True positive rate; FPR: False positive rate; MC-OCEL: Maximum Correntropy One-Class Classification Model; KNN: K-nearest neighbor; MST: Minimum spanning trees; MPM: Minimax probability machine; LPDD: Linear programming dissimilarity-data description; MCA: Multivariate Correlation Analysis; TANN: Triangle Area Nearest Neighbors; GAA-ADS: Geometric Area Analysis Attack Detection System; ODM: Outlier Dirichlet Mixture; AD-CNN: Attack Detection-based Convolutional Neural Network.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]University of Wisconsin-Eau Claire, Eau Claire, WI 54702, USA. [2]Florida Atlantic University, Boca Raton, FL 33431, USA.

References
1. Bauder RA, Khoshgoftaar TM, Hasanin T. An empirical study on class rarity in big data. In: 2018 17th IEEE international conference on machine learning and applications (ICMLA). IEEE; 2018.
2. Herland M, Bauder RA, Khoshgoftaar TM. The effects of class rarity on the evaluation of supervised healthcare fraud detection models. J Big Data. 2019;6(1):1–33.
3. Tawfiq H, Khoshgoftaar TM, Leevy JL, Bauder RA. Investigating class rarity in big data. J Big Data. 2020;7(1):1–7.
4. Bauder RA, Khoshgoftaar TM. A study on rare fraud predictions with big Medicare claims fraud data. Intell Data Anal. 2020;24(1):141–61.
5. Heredia B, Khoshgoftaar TM, Prusa JD, Crawford M. Improving detection of untrustworthy online reviews using ensemble learners combined with feature selection. Soc Netw Anal Min. 2017;7(1):1–18.
6. Hasanin T, Khoshgoftaar TM, Leevy J, Seliya N. Investigating random undersampling and feature selection on bioinformatics big data. In: 2019 IEEE fifth international conference on big data computing service and applications (BigDataService). IEEE; 2019.
7. Bartkowiak AM. Anomaly, novelty, one-class classification: a comprehensive introduction. Int J Comput Inf Syst Ind Manag Appl. 2011;3(1):61–71.
8. Kennedy K, Mac Namee B, Delany SJ. A study of one-class classification and the low-default portfolio problem. In: Irish conference on artificial intelligence and cognitive science; 2009, p. 174–87.
9. Khan SS, Madden MG. One-class classification: taxonomy of study and review of techniques. Knowl Eng Rev. 2014;29(3):345–74.

10. Schölkopf B, Platt JC, Shawe-Taylor J, Smola AJ, Williamson RC. Estimating the support of a high-dimensional distribution. Neural Comput. 2001;17(3):1443–71.
11. Tax DM, Duin RP. Support vector data description. Mach Learn. 2004;54(1):45–66.
12. Jiang MF, Tseng SS, Su CM. Two-phase clustering process for outliers detection. Pattern Recogn Lett. 2001;22(6–7):691–700.
13. Bishop CM. Neural networks for pattern recognition. Oxford: Oxford University Press; 1995.
14. Carpenter GA, Grossberg S, Rosen. Fuzzy ART: fast stable learning and categorization of analog patterns by an adaptive resonance system. Neural Netw. 1991;4(6):759–71.
15. Salekshahrezaee Z, Leevy JL, Khoshgoftaar TM. A reconstruction error-based framework for label noise detection. J Big Data. 2021;8(1):1–16.
16. Japkowicz N, Myers C, Gluck M. A novelty detection approach to classification. In: International joint conference on artificial intelligence (IJCAI); 1995.
17. Desir C, Bernard S, Petitjean C, Heutte L. A new random forest method for one class classification. In: IAPR international workshop on statistical techniques in pattern recognition (SPR), Hiroshima, Japan, vol. 7626 of LNCS; 2012, p. 282–90.
18. Krawczyk B, Woźniak M, Cyganek B. Clustering-based ensembles for one-class classification. Inf Sci. 2014;264:182–95.
19. Pekalska EE, Tax DM, Duin R. One-class LP classifiers for dissimilarity representations. In: Advances in neural information processing systems; 2003.
20. Lang R, Lu R, Zhao C, Qin H, Liu G. Graph-based semi-supervised one-class support vector machine for detecting abnormal lung sounds. Appl Math Comput. 2020;364:124487.
21. Gautam C, Tiwari A, Leng Q. On the construction of extreme learning machine for online and offline one-class classification—an expanded toolbox. Neurocomputing. 2017;261:126–43.
22. Krawczyk B, Woźniak M. One-class classifiers with incremental learning and forgetting for data streams with concept drift. Soft Comput. 2015;19(12):3387–400.
23. Leng Q, Qi H, Miao J, Zhu W, Su G. One-class classification with extreme learning machine. In: Mathematical problems in engineering, vol. 15. London: Hindawi Publishing; 2015.
24. Dai H, Cao J, Wang T, Deng M, Yang Z. Multilayer one-class extreme learning machine. Neural Netw. 2019;115:11–22.
25. Dreiseitl S, Osl M, Scheibböck C, Binder M. Outlier detection with one-class SVMs: an application to melanoma prognosis. In: AMIA annual symposium proceedings; 2010.
26. Witten IH, Frank E, Hall MA, Pal CJ. Data mining: practical machine learning tools and techniques. 4th ed. Burlington: Morgan Kaufmann; 2017.
27. Mourão-Miranda J, Hardoon DR, Hahn T, Marquand AF, Williams SC, Shawe-Taylor J, Brammer M. Patient classification as an outlier detection problem: an application of the one-class support vector machine. Neuroimage. 2011;58(3):793–804.
28. Bartkowiak AM, Zimroz R0 Outliers analysis and one class classification approach for planetary gearbox diagnosis. In: 9th international conference on damage assessment of structures (DAMAS 2011); 2011.
29. Désir C, Bernard S, Petitjean C, Heutte L. One class random forests. Pattern Recogn. 2013;46(12):3490–506.
30. Das B, Cook DJ, Krishnan NC, Schmitter-Edgecombe M. One-class classification-based real-time activity error detection in smart homes. IEEE J Select Top Signal Process. 2016;10(5):914–23.
31. Deng X, Jiang P, Peng X, Mi C. An intelligent outlier detection method with one class support tucker machine and genetic algorithm toward big sensor data in Internet of Things. IEEE Trans Industr Electron. 2018;66(6):4672–83.
32. Gautam C, Mishra PK, Tiwari A, Richhariya B, Pandey HM, Wang S. Minimum variance-embedded deep kernel regularized least squares method for one-class classification and its applications to biomedical data. Neural Netw. 2020;123:191–216.
33. Mygdalis V, Iosifidis A, Tefas A, Pitas I. One class classification applied in facial image analysis. In: 2016 IEEE international conference on image processing (ICIP); 2016.
34. Kauffmann J, Müller KR, Montavon G. Towards explaining anomalies: a deep Taylor decomposition of one-class models. Pattern Recogn. 2020;101:107198.
35. Montavon G, Binder A, Lapuschkin S, Samek W, Müller KR. Layer-wise relevance propagation: an overview. Explainable AI: interpreting, explaining and visualizing deep learning; 2019. p. 193–209.
36. Aguilera J, Farías DIH, Ortega-Mendoza RM, Montes-y-Gomez M. Depression and anorexia detection in social media as a one-class classification problem. Appl Intell. 2021. https://doi.org/10.1007/s10489-020-02131-2.
37. Aguilera J, Gonzalez LC, Montes-y-Gomez M, Rosso P. A new weighted k-nearest neighbor algorithm based on Newton's gravitational force. In: Vera-Rodriguez R, Fierrez J, Morales A, editors. Progress in pattern recognition, image analysis, computer vision, and applications. Cham: Springer International Publishing; 2018. p. 305–13.
38. Pu G, Wang L, Shen J, Dong F. A hybrid unsupervised clustering-based anomaly detection method. Tsingua Sci Technol. 2021;2(2):146–53.
39. Casas P, Mazel J, Owezarski P. Unsupervised network intrusion detection systems: detecting the unknown without knowledge. Comput Commun. 2012;35(7):772–83.
40. Favarelli E, Giorgetti A. Machine learning for automatic processing of modal analysis in damage detection of bridges. IEEE Trans Instrum Measure. 2021;70:1–3.
41. Mahfouz AM, Abuhussein A, Venugopal D, Shiva D. Network intrusion detection model using one-class support vector machine. In: Patnaik S et al. (eds.), Advances in machine learning and computational intelligence, algorithms for intelligent systems; 2021.
42. Zaidi SFA, Lee C. One-class classification based bug triage system to assign a newly added developer. In: IEEE international conference on information networking (ICOIN); 2021.
43. Lee SR, Heo MJ, Lee CG, Kim M, Jeong G. Applying deep learning based automatic bug triager to industrial projects. In: Proceedings of the 11th joint meeting on foundations of software engineering; 2017, p. 926–31.
44. Mani S, Sankaran A, Aralikatte R. Deeptriage: exploring the effectiveness of deep learning for bug triaging. In: Proceedings of the ACM India joint international conference on data science and management of data; 2019, p. 171–9.

45. Clifton L, Clifton DA, Watkinson PJ, Tarassenko L. Identification of patient deterioration in vital-sign data using one-class support vector machines. In: 2011 federated conference on computer science and information systems (FedCSIS); 2011.
46. Kemmler M, Rodner E, Wacker ES, Denzler J. One-class classification with Gaussian processes. Pattern Recogn. 2013;46(12):3507–18.
47. Beghi A, Cecchinato L, Corazzol C, Rampazzo M, Simmini F, Susto GA. A one-class svm based tool for machine learning novelty detection in hvac chiller systems. IFAC Proc. 2014;47(3):1953–8.
48. Domingues R, Michiardi P, Zouaoui J, Filippone M. Deep Gaussian Process autoencoders for novelty detection. Mach Learn. 2018;107:1363–83.
49. Sadooghi MS, Khadem SE. Improving one class support vector machine novelty detection scheme using nonlinear features. Pattern Recogn. 2018;83:14–33.
50. Yin L, Wang H, Fan W. Active learning-based support vector data description method for robust novelty detection. Knowl Based Syst. 2018;153:40–52.
51. Rad Mohammadian N, Van Laarhoven T, Furlanello C, Marchiori E. Novelty detection using deep normative modeling for imu-based abnormal movement monitoring in Parkinson's disease and autism spectrum disorders. Sensors. 2018;18(10):3533.
52. Sabokrou M, Khalooei M, Fathy M, Adeli E. Adversarially learned one-class classifier for novelty detection. In: IEEE conference on computer vision and pattern recognition; 2018.
53. Shorten C, Khoshgoftaar TM. A survey on image data augmentation for deep learning. J Big Data. 2019;6(1):1–48.
54. Sabokrou M, Fathy M, Zhao G, Adeli E. Deep end-to-end one-class classifier. IEEE Trans Neural Netw Learn Syst. 2021;32(2):675–84.
55. Oosterlinck D, Benoit DF, Baecke P. From one-class to two-class classification by incorporating expert knowledge: novelty detection in human behaviour. Eur J Oper Res. 2020;282(3):1011–24.
56. Xing HJ, Liu W. Robust AdaBoost based ensemble of one-class support vector machines. Inf Fus. 2020;55:45–58.
57. Rafieipour H, Zadeh AA, Moradan A, Salekshahrezaee Z. Study of genes associated with Parkinson disease using feature selection. J Bioeng Res. 2020;2(4):1–12.
58. Perera P, Nallapati R, Xiang B. OCGAN: one-class novelty detection using GANS with constrained latent representations. In: IEEE/CVF conference on computer vision and pattern recognition (CVPR); 2019, p. 2893–901.
59. Zhang Y, Zhou B, Ding X, Ouyang J, Cai X, Gao J, Yuan X. Adversarially learned one-class novelty detection with confidence estimation. Inf Sci. 2021;552:48–64.
60. Kim S, Choi Y, Lee M. Deep learning with support vector data description. Neurocomputing. 2015;165:111–7.
61. Erfani SM, Rajasegarar S, Karunasekera S, Leckie C. High-dimensional and large-scale anomaly detection using a linear one-class SVM with deep learning. Pattern Recogn. 2016;58:121–34.
62. Sun J, Shao J, He C. Abnormal event detection for video surveillance using deep one-class learning. Multimedia Tools Appl. 2019;78(3):3633–47.
63. Altaher A, Salekshahrezaee Z, Zadeh AA, Rafieipour H, Altaher A. Using multi-inception CNN for face emotion recognition. J Bioeng Res. 3(1):1-12
64. Mahadevan V, Li W, Bhalodia V, Vasconcelos N. Anomaly detection in crowded scenes. In: 2010 IEEE computer society conference on computer vision and pattern recognition; 2010.
65. Cong Y, Yuan J, Liu J. Sparse reconstruction cost for abnormal event detection. In: 24th IEEE conference on computer vision and pattern recognition (CVPR 2011); 2011.
66. Xu D, Ricci E, Yan Y, Song J, Sebe N. Learning deep representations of appearance and motion. In: Proceedings of the British machine vision conference (BMVC2015); 2015.
67. Adam A, Rivlin E, Shimshoni I, Reinitz D. Robust real-time unusual event detection using multiple fixed-location monitors. IEEE Trans Pattern Anal Mach Intell. 2008;30(3):555–60.
68. Zhang M, Wu J, Lin H, Yuan P, Song Y. The application of one-class classifier based on CNN in image defect detection. Procedia Comput Sci. 2017;114:341–8.
69. Gutoski M, Ribeiro M, Aquino NMR, Lazzaretti AE, Lopes HS. A clustering-based deep autoencoder for one-class image classification. In: IEEE Latin American conference on computational intelligence (LACCI); 2017.
70. Ruff L, Vandermeulen R, Goernitz N, Deecke L, Siddiqui SA, Binder A, Muller E, Kloft M. Deep one-class classification. In: International conference on machine learning; 2018.
71. Kingma D, Ba J. Adam: a method for stochastic optimization, v1; 2015. https://arxiv.org/abs/1412.6980.
72. Brendel W, Rauber J, Bethge M. Decision-based adversarial attacks: Reliable attacks against black-box machine learning models, v1; 2017. https://arxiv.org/abs/1712.04248.
73. Stallkamp J, Schlipsing M, Salmen J, Igel C. The German traffic sign recognition benchmark: a multi-class classification competition. In: Proceedings of the international joint conference on neural networks; 2011.
74. Schlegl T, Seeboock P, Waldstein SM, Schmidt-Erfurth U, Langs G. Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. In: International conference on information processing in medical imaging; 2017.
75. Chalapathy R, Menon AK, Chawla S. Anomaly detection using one-class neural networks, v1; 2018. https://arxiv.org/abs/1802.06360.
76. Schlachter P, Liao Y, Yang B. Deep one-class classification using intra-class splitting. In: IEEE data science workshop (DSW), June 2019, p. 100–4, Minneapolis, MN, https://doi.org/10.1109/DSW.2019.8755576.
77. Perera P, Patel VM. Learning deep features for one-class classification. IEEE Trans Image Process. 2019;28:5450–63.
78. Pan SJ, Yang Q. A survey on transfer learning. IEEE Trans Knowl Data Eng. 2009;22(10):1345–59.
79. Mahbub U, Sarkar S, Patel VM, Chellappa R. Active user authentication for smartphones: A challenge data set and benchmark results. In: IEEE 8th international conference on biometrics theory, applications and systems (BTAS); 2016.
80. Burlina P, Joshi M, Billings S, Wang IJ, Albayda J. Deep embeddings for novelty detection in myopathy. Comput Biol Med. 2019;105:46–53.
81. Maaten LVD, Hinton G. Visualizing data using t-SNE. J Mach Learn Res. 2008;9:2579–605.

82. Ghafoori Z, Leckie C. Deep multi-sphere support vector data description. In: SIAM international conference on data mining; 2020.
83. Lloyd S. Least squares quantization in PCM. IEEE Trans Inf Theory. 1982;28:129–37.
84. Liu F, Liu H, Zhang W, Liu G, Shen L. One-class fingerprint presentation attack detection using auto-encoder network. IEEE Trans Image Process. 2021;30:2394–406.
85. Cao J, Dai H, Lei B, Yin C, Zeng H, Kummert A. Maximum correntropy criterion-based hierarchical one-class classification. IEEE Trans Neural Netw Learn Syst. 2020. https://doi.org/10.1109/TNNLS.2020.3015356.
86. Fontenla-Romero O, Pérez-Sánchez B, Guijarro-Berdiñas B. DSVD-autoencoder: a scalable distributed privacy-preserving method for one-class classification. Int J Intell Syst. 2020;26:177–99.
87. Moustafa N, Keshk M, Choo KR, Lynar T, Camtepe S, Whitty M. DAD: a distributed anomaly detection system using ensemble one-class statistical learning in edge networks. Futur Gener Comput Syst. 2021;118:240–51.
88. Pourreza M, Mohammadi B, Khaki M, Bouindour S, Snoussi H, Sabokrou M. G2D: generate to detect anomaly. In: Workshop on applications of computer vision; 2021.
89. Chong P, Ruff L, Kloft M, Binder A. Simple and effective prevention of mode collapse in deep one-class classification. arXiv preprint; 2020. https://arxiv.org/abs/2001.08873.
90. Tan WR, Chan CS, Aguirre HE, Tanaka K. ArtGAN: artwork synthesis with conditional categorical GANs. In: IEEE international conference on image processing (ICIP); 2017.
91. Golan I, El-Yaniv R. Deep anomaly detection using geometric transformations. In: Advances in neural information processing systems; 2018.
92. Ruff L, Vandermeulen RA, Görnitz N, Binder A, Müller E, Müller KR, Kloft M. Deep semi-supervised anomaly detection. arXiv preprint; 2020. https://arxiv.org/abs/1906.02694.
93. Goyal S, Raghunathan A, Jain M, Simhadri HV, Jain P. DROCC: deep robust one-class classification. arXiv preprint; 2020. https://arxiv.org/abs/2002.12718.
94. Pimentel MA, Clifton DA, Clifton L, Tarassenko L. A review of novelty detection. Signal Process. 2014;99:215–49.
95. Saki M, Abolhasan M, Lipman J. A novel approach for big data classification and transportation in rail networks. IEEE Trans Intell Transp Syst. 2019;21(3):1239–49.
96. Jintawatsakoon S, Charoenruengkit W. Novelty detection of beverage bottle images based on transfer learning. In: IEEE 5th international conference on information technology; 2020, p. 87–91.
97. Tsai CF, Lin WC. Feature selection and ensemble learning techniques in one-class classifiers: an empirical study of two-class imbalanced datasets. IEEE Access. 2021. https://doi.org/10.1109/ACCESS.2021.3051969.
98. Lesouple J, Tourneret JY. Incorporating user feedback into one-class support vector machines for anomaly detection. In: 28th European signal processing conference (EUSIPCO 2020); 2020, p. 1609–12, Amsterdam.
99. Jaber M, Nasser A, Charara N, Mansoury A, Yao KC. One-class based learning for hybrid spectrum sensing in cognitive radio. In: 28th European signal processing conference (EUSIPCO 2020); 2020, p. 1683–6, Amsterdam.

Publisher's Note