Machine Learning (2024) 113:5495–5531
https://doi.org/10.1007/s10994-023-06375-0

Survey on extreme learning machines for outlier detection

Rasoul Kiani[1] · Wei Jin[2] · Victor S. Sheng[1]

Received: 7 June 2022 / Revised: 9 May 2023 / Accepted: 17 July 2023 / Published online: 23 January 2024
© The Author(s), under exclusive licence to Springer Science+Business Media LLC, part of Springer Nature 2024

Abstract

In a two-class classification task, if the number of examples of one class (majority) is much greater than that of another class (minority), then the classification is said to be class imbalanced. It can occur among many real-world applications, such as intrusion detection, medical diagnosis, etc. The class imbalance issue can make learning difficult since learning opts to bias towards the majority class. Outliers are cases with anomalous behaviors and are extreme cases of class imbalance. Despite late advances in extreme learning machines (ELMs), there are not many experimental investigations in the field of ELM with outlier detection. In this survey, we provide a comprehensive overview of existing ELMs to address the problem of outlier detection under a unified perspective. Firstly, we describe the background of our work, which includes a brief overview of previous surveys and a detailed description of the enhanced ELMs. Next, available studies regarding why ELMs are used to tackle the class imbalance problem are reviewed. Furthermore, cutting-edge algorithms are surveyed for improved ELMs to detect outliers. We classify these methods under three different machine learning perspectives (i.e., supervised, unsupervised, and semi-supervised approaches). In addition, we explore the developments of existing solutions based on three standardized quality metrics (i.e., accuracy, robustness, and speed) and other performance metrics (e.g., mean absolute percentage error and mean absolute error). After that, related datasets are detailed to facilitate future studies in this field. Last but the most important, this study concludes with discussions, challenges, and suggestions to guide future research.

Keywords Class imbalance · Outlier detection · Anomaly detection · Extreme learning machine

---

Editors: Nuno Moniz, Paula Branco, Luís Torgo, Nathalie Japkowicz, Michal Wozniak, Shuo Wang.

Extended author information available on the last page of the article

1 Introduction

The class imbalance occurs in many real-world applications where the class distributions of data are significantly imbalanced (Ling & Sheng, 2008). This happens when one class, which is a minority group, has significantly fewer samples than the other class, which is a majority group (Johnson & Khoshgoftaar, 2019). That is to say, the majority class makes up most of the dataset, whereas the minority class is often considered the class of interest (Leevy et al., 2018; Bauder & Khoshgoftaar, 2018; Zhang et al., 2014). When the instances belonging to the minority group are misclassified more often than those belonging to the majority group (Johnson & Khoshgoftaar, 2019), this poses a difficulty for learning algorithms, since they will be biased towards the majority group (Krawczyk, 2016; Li et al., 2020a, b). Equation (1) shows the imbalance ratio (IR) to describe the imbalance extent of a dataset (Johnson & Khoshgoftaar, 2019):

$$\rho = \frac{max_i\,\{|C_i|\}}{min_i\,\{|C_i|\}} \tag{1}$$

where C_i is a set of examples in class *i*, and $max_i\,\{|C_i|\}$ and $min_i\,\{|C_i|\}$ return the maximum and minimum class size over all *i* classes, respectively.

Outliers are cases with anomalous behaviors in a domain. They show a higher deviation and are not in line with the behavior of general cases, which could cause unexpected results in analytics (Albuquerque Filho et al., 2022; Kiani et al., 2020; Aggarwal, 2017). A simple illustrative two-dimensional example that depicts an outlier status is shown in Fig. 1. Suppose a dataset contain two large sections, S_1 and S_2. P_1, P_3, P_4, and a small section with very few data points P_2, are referred to as outliers since they are far away from the two large sections (S_1 and S_2) (Wang et al., 2019a, b). Outlier detection has been known as an extreme case of class imbalance learning. Detecting outliers is a significant problem that has been studied in various research and application areas.

By identifying outliers, researchers can obtain vital knowledge which assists in making better decisions (Wang et al., 2019a, b). Also, detecting outliers translates to significant

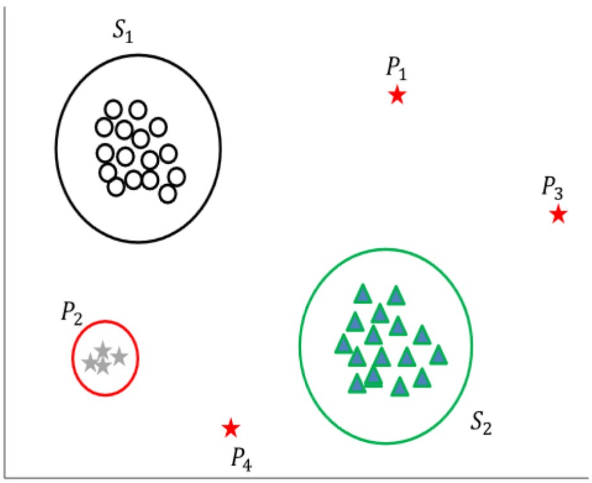

Fig. 1 An example of outliers in a two-dimensional dataset in which S_1 and S_2 show normal data, P_2 presents outliers (collective anomalies), and P_1, P_3, and P_4 indicate outliers (point anomalies)

actionable information in a wide variety of applications such as fraud detection (Anjaneyulu & Kishore, 2019; Jiang et al., 2018; Kalid et al., 2020; Sekar, 2022), intrusion detection in cyber security (Ariafar and Kiani 2017; Sahu et al. 2021; Zhou et al., 2015; Tama et al., 2019), and health diagnosis (Devika et al., 2022; Su et al., 2021; Muñoz-Ramírez et al., 2022; Dashdondov & Kim, 2021). Moreover, outlier detection has been studied in a variety of domains with different data modes including high dimensional data (Aggarwal and Yu 2001), uncertain data (Aggarwal and Yu 2008), streaming data (Aggarwal et al. 2011), network data, and time series data (Gupta et al., 2013).

There are two types of outliers: noises and anomalies. Noises are weak outliers but anomalies are strong outliers. Different aspects of outliers are displayed in Fig. 2. The boundary between noises and anomalies is not clear but can be determined through different analytical methods (Kiani et al., 2020; Aggarwal, 2017; Kiani et al. 2019). Outlier detection problems are among other important factors, along with their key detection parameters important to data analysts, which challenge supervised, semi-supervised, and unsupervised learning methods (Fernández et al., 2022; Adeli et al., 2018; Lee et al., 2021). In supervised learning, normal and outlier cases are labeled. In the semi-supervised learning, only few examples of normal and outlier cases are labeled. In the unsupervised learning, no cases are labeled at all (Chakraborty et al., 2022; Bawono and Bachtiar 2019).

There are three types of anomalies as follows: (a) point anomalies are at a considerable distance from others such as points P_1, P_3, and P_4 in Fig. 1; (b) collective anomalies are a set of correlated cases with a deviation from others such as small section P_2 in Fig. 1; and (c) conditional or contextual anomalies that could be considered anomalous in some specific condition or context. The point and collective anomalies are subsets of conditional anomalies (Aggarwal, 2017; Kiani et al. 2019).

Recently, Wang et al. (2016a, b) proposed a new definition of outlier called cluster-based outlier. Continuously, new types of anomalies called Collective Normal Anomaly (CNA) and Collective Point Anomaly (CPA) were defined by Kiani et al. (2020) to improve a much better detection of the thin boundary between different types of anomalies.

CPA is a subset of Point Anomaly (PA) and there is a thin boundary between PA and CPA. Equation (2) defines CPA that the neighborhood radius of CPA is less than average neighborhood radius of PA:

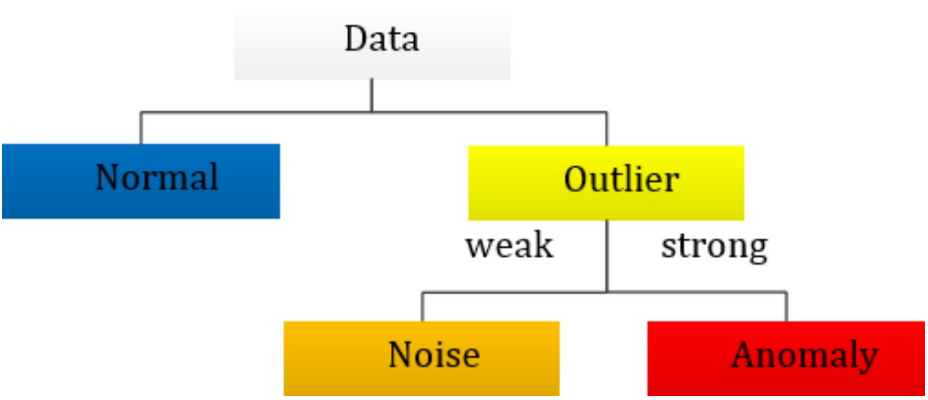

Fig. 2 Different aspect of outliers

$$CPA = \{P_i \in CPA | P_i \in PAandOut\_Rad_{p_i} < Out\_Rad_{PA}\} \tag{2}$$

where $Out\_Rad_{p_i}$ is the neighborhood radius of P_i as well as its average distance to PA and Out_Rad_{PA} is the neighborhood radius of PA (See Kiani et al., 2020 for more details).

CNA is a cluster whose standard deviation density is greater than or equal to the threshold for standard deviation of all clusters. Equation (3) defines the relationship between CNA and Normal Data (ND):

$$CNA = \{C_i \in C | C_i \in NDand\sigma\_Den_{C_i} < Th\_\sigma_{Den_C}\} \tag{3}$$

where C_i is one of the detected clusters, C is the set of all clusters, $\sigma\_Den_{C_i}$ is the standard deviation of the density of cluster C_i, and $Th\_\sigma_{Den_C}$ is the threshold of the standard deviation of all clusters (See Kiani et al., 2020 for more details).

In recent years, ELM has gained lots of research interest due to its simplified algorithm and fast learning speed (Zhang et al., 2020; Lu et al., 2017; Janakiraman and Nielsen 2016). An ELM is a single hidden layer feed-forward model whose input layer parameters are assigned using random numbers and fixed during training (Huang et al., 2006a, b; Thammasakorn et al. 2018). The ELM algorithm can provide a high generalization performance in most domains and learn thousands of times faster than conventional popular learning algorithms for feed-forward neural networks (Huang et al., 2006a, b; Thammasakorn et al. 2018). As seen in Fig. 3, an ELM model has an input layer, just one hidden layer, and an output layer. The input weights are initialized randomly, only the output weights update during the training phase. While ELM can be implemented to identify different types of anomalies in various fields, its performance might be affected by the random initialization of weights and biases or by the large generated network which might contain unnecessary number of neurons (Eshtay et al., 2020). This challenge may cause unstable classification accuracy, which greatly limits the performance of the ELM networks (Chen et al., 2019).

The loss function for ELM can be explained as follows:

$$E = \sum_{j=1}^{N} \left(\sum_{i=1}^{L} \beta_i g\left(W_i.X_j + b_i\right) - t_j \right)^2 \tag{4}$$

where N is the number of the input neurons, L is the number of hidden neurons, β_i is the weight vector between hidden layer and output layer, *g(x)* is the activation function, W_i is the weight vector between input layer and hidden layer, X_j is the input data, b_i is the threshold, and t_j is the output data (Mi et al., 2017).

The details of methods developed to tackle the ELM challenges will be reviewed in Sect. 3. Moreover, the review of related work shows that the ELM also is not robust to outliers and it is widely used in batch learning (Frénay & Verleysen, 2015). To address these problems, 20 considerably different optimized ELMs in various outlier detection areas are considered and will be explored in Sect. 4. The literature search process was first followed through the Google Scholar, Springer, IEEE Xplore, and Elsevier databases. Keyword searches were then performed including combinations of query terms such as: "imbalanced

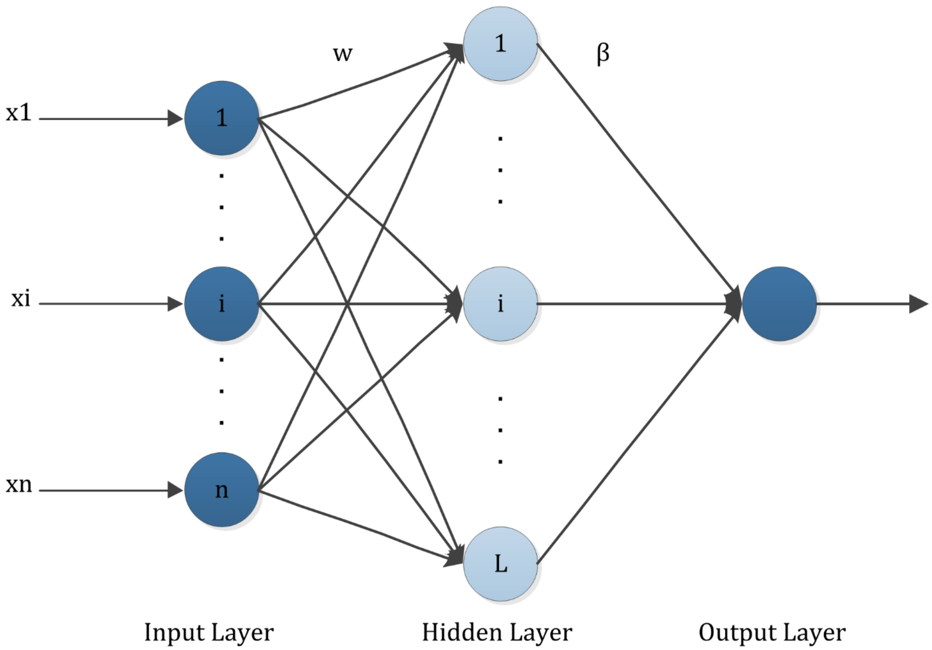

Fig. 3 Structure of the ELM

data", "class imbalance", "outlier detection", "anomaly detection", and "extreme learning machine".

It can be seen that, in recent years, several survey papers such as Wang et al. (2021); Huang et al., 2011a, b; Ding et al. (2014) have provided a comprehensive review on the development of ELM and imbalanced data, including the theoretical analysis, typical variants, recent advances, and real applications. However, the first question is to investigate the suitability of ELMs for outlier detection. For this reason, EMLs have been developed based on different techniques to optimize various metrics such as accuracy, performance, robustness, speed, etc. Experimental results of ELM models showed promising results for real-world applications in outlier detection.

However, the application of ELM for outlier detection is a relatively new area of research. We observed several research gaps during literature review of outlier detection. The first gap is lack of investigation of well-known ELM approaches for outlier detection. Although isolated studies were available as described in Sect. 4, no comprehensive research work is available to fill this gap. The second research gap is why ELMs are used to tackle the imbalanced problem. The third gap turns out to be lack of taxonomy of ELMs amongst themselves using different machine learning techniques and lack of their developments based on standardized quality metrics which is a natural consequence of the previous two gaps. The forth research gap is the use of suitable datasets to facilitate future studies in this field.

The primary contribution of this work is filling the above mentioned research gaps based on state-of-the-art ELMs. The first gap is filled by exploring outlier detection models using ELMs. To the best of our knowledge, the papers investigated in this study have not been analyzed for outlier detection. To fill the second research gap, we survey recent ELMs for class imbalanced data. This investigation can provide a comprehensive survey for researchers to follow in this field. To bridge the third research gap, we classify twenty cutting-edge ELMs under three different machine learning techniques and explore their developments based on popular metrics. To close the forth gap related datasets are detailed in this work.

Therefore, we need a survey with a particular focus on outlier detection, especially in the area of outlier detection and ELMs. In summary, the novel and significant contributions of the paper are as follows:

- We investigate recent advances of ELMs in class imbalanced data and introduce cutting-edge algorithms and discuss them with highlighting more details. To this purpose, 20 papers issued from 2014 to 2021 will be discussed. It can provide valuable information for researchers who are interested in working on class imbalance problems, such as increasing the penalty associated with misclassifying the positive class relative to the negative class, oversampling the majority class and under-sampling the minority class.
- We significantly focus on state-of-the-art ELMs which have been applied for outlier detection in different issues. In this respect, 20 papers published between 2013 and 2021 will be explored. Moreover, we classify these methods under three different machine learning perspectives (i.e., supervised, unsupervised, and semi-supervised approaches). In addition, we explore the developments of existing solutions based on three popular metrics (i.e., accuracy, robustness, and speed) and other performance metrics (e.g., mean absolute percentage error and mean absolute error). This study can be followed by researchers who tend to develop ELMs for outlier detection since to the best of our knowledge, this is the first work which has focused on ELMs in this field.
- We discuss different datasets used for the improvement of ELMs in outlier detection to facilitate future studies in this field, due to the fact that datasets are the basis for facilitating the development in multiple computing domains, robustness, and reliability of results.
- We present some contemporary open challenges to cover significant gaps when ELMs are applied for outlier detection.

The remainder of this paper is organized as follows. Section 2 reviews three surveys on which focused recent ELMs and are among the most cited articles or have been recently published. Twenty most recent published studies that investigate ELMs for addressing the class imbalance are surveyed in Sect. 3. In Sect. 4, twenty most recent published papers that provide readers with a more complex picture of existing approaches for outlier detection using ELMs are discussed. Section 5 provides necessary information on datasets which have been used in this field. Finally, discussions are given, and conclusions are drawn.

2 Literature review

The reasons why extensive research on ELM has been carried out are less manual intervention, higher classification accuracy, and less training time (Wang et al., 2021). In this section, we describe the background for our work, which includes a brief overview of surveys and a detailed description of the enhanced ELMs. In this regard, three surveys which are among the most cited articles or have been recently published are summarized as follows.

As shown in Table 1, we briefly describe the survey paper published by Huang et al. (2011a, b).

ELM and its applications were reviewed by Ding et al. (2014). They first described a brief review of ELM and its different methods and provided the variants of ELM. Some classical applications of ELM are then introduced, which are summarized in Table 2 as follows.

Wang et al. (2021) published a review on ELM. Their paper was focused on theoretical analysis, various improvements of ELM in terms of stability, efficiency, and accuracy, real-time learning tasks, and the applications of ELM in various fields. We summarize ELM applications based on the Wang et al. (2021) paper in Table 3 as follows.

3 ELM for class imbalanced data

3.1 Methods addressing accuracy

Li et al. (2020a, b) developed a method based on Multi-Kernel Extreme Learning Machine Fusion (MKELMF) in order to solve the problem of low classification accuracy in imbalanced data classification. They used the meta-learning algorithm to select a group of Kernel Extreme Learning Machines (KELM) with sufficient classification ability for combined decision making. Thammasakorn et al. (2018) developed a GWO-weighted ELM that is a combination of the weighted ELM integrated with the Gray wolf optimizer. The GWO-ELM improved the accuracy of the weighted ELM model in the classification of imbalanced data, where the original weighted ELM suffers from the pessimal selected regularization parameter. To get fast and efficient classification, a new online sequential extreme learning machine algorithm with the sparse-weighting strategy was proposed by Mao et al. (2017) to increase the accuracy of minority class while reducing the accuracy loss of majority class as much as possible. Experimental results on two kinds of imbalanced datasets, UCI datasets and the real-world air pollutant forecasting dataset, showed that the proposed method has higher prediction accuracy and better numerical stability compared with ELM, OS-ELM, meta-cognitive OS-ELM and weighted OS-ELM.

3.2 Methods addressing robustness

Li et al. (2014) proposed a boosting weighted ELM, which embeds the weighted ELM seamlessly into a modified AdaBoost framework. They modified AdaBoost in two aspects to be more effective for imbalanced learning: (i) the initial distribution weights are set to be asymmetric so that AdaBoost converges at a faster speed; (ii) the distribution weights are then updated separately for different classes to avoid destroying the distribution weights asymmetry. Cheng et al. (2020) proposed an improved algorithm called Boosting Label Weighted-ELM (BLW) which integrates Label Weighted-ELM (LW-ELM) into the boosting ensemble learning framework. They claimed that BLW-ELM is a universal and self-adapting algorithm that can promote the robustness of classification regardless of the data distribution types. Xu et al. (2019) proposed an ensemble classification method called AdaWELM, for fault diagnosis in wastewater treatment. The individual classifiers are built by using the weighted extreme learning machine (WELM), and then combined them with

Table 1 A briefing of survey paper published by Huang et al. (2011a, b)

Method Name	More details
ELM	
• Basic ELM (Huang et al., 2006a, b) • Random hidden layer feature mapping based ELM (Huang et al., 2011a, b) • Kernel ELM(Huang et al., 2011a, b)	The essence of ELM is that the hidden layer of single hidden layer feedforward neural networks (SLFNs) need not be tuned. Compared with those traditional computational intelligence techniques, ELM provides better generalization performance at a much faster learning speed and with least human intervention.
Fully complex ELM (Huang et al., 2008)	The fully complex algorithm can be linearly extended to the complex domain and can obtain much lower symbol error rate (SER).
Online sequential ELM (OS-ELM) (Liang et al., 2006)	In many industrial applications training data may come one by one or chunk by chunk. In these cases, online sequential learning algorithms are preferred over batch learning algorithms as sequential learning algorithms do not require retraining whenever new data are received.
Incremental ELM (I-ELM) (Huang et al., 2006a, b)	I-ELM has no parameters for users to specify except the maximum network architecture and the expected accuracy. I-ELM can work well with a wide range of activation functions no matter whether they are sigmoidal or non-sigmoidal, continuous or non-continuous, and differentiable or non-differentiable.
ELM ensembles	
• Adaptive ensemble models of ELM (Heeswijk et al. 2009) • ELM ensemble for large scale regression applications (Van Heeswijk et al., 2011) • Ensemble of Online Sequential ELM (EOS-ELM) (Lan et al., 2009)	Network ensemble consists of a few of single networks that may have different adaptabilities to the new data. Some of the networks in the ensemble may adapt faster and better to the new data than others, which could make the ensemble overcome the problem of networks that could not adapt well to the new data.
Pruning ELM (P-ELM) (Rong et al., 2008)	
• Optimally-pruned ELM (OP-ELM)	P-ELM was proposed as a systematic and automated method for the ELM classifier network design. It starts with a large network and then eliminates the hidden nodes that have low relevance to the class labels by using statistical criteria. OP-ELM is applicable for both regression and classification applications.
Constructive ELM	
• Error minimized ELM (EM–ELM) (Feng et al., 2009)	EM-ELM is an error minimization based method in which the number of hidden nodes can grow one-by-one or group-by-group until optimal. The approach can significantly reduce the computational complexity and its convergence was proved as well.
• Stepwise forward selection based constructive ELM for regression (Lan et al., 2010a, b)	The fast construction algorithm (FCA) is a constructive hidden node selection method for ELM based on orthogonal least squares (OLS). OLS selects a suitable set of variables to form the subset model from a large set of candidates. By modifying the classic forward selection algorithm, a constructive hidden nodes selection method for ELM (CS-ELM) was proposed, which is less greedy and without any matrix decompositions.
• Two-stage ELM for regression (TS-ELM) (Lan et al., 2010a, b)	The first stage attempts to select hidden nodes by the forward recursive algorithm and the selection is terminated by the final prediction error (FPE) criterion; while the second stage is a backward refinement phase that removes the insignificant hidden nodes by applying the Leave-One-Out (LOO) method.

Table 2 A briefing of classical applications of ELM published by Ding et al. (2014)

Application	Why ELM
Classification	
• Mobile object index (Wang et al., 2012)	ELM was used to classify the region dynamically to adapt to changes in environment.
• Text categorization (Zheng et al., 2013)	RELM was developed, including the uni-label and multi-label situations.
• Electrocardiogram signals (Karpagachelvi et al., 2012)	ELM was used to classify ECG signals.
• Effective accuracy performance (Kim et al., 2009)	ELM was used to propose a robust arrhythmia classification algorithm.
• Machine control (Lee et al., 2009)	ELM was used to classify machine control commands out of time series of spike trains of ensembles of CAI hippocampus neurons of a rat.
Regression	
• Regression	The Parallel ELM (PELM) was used to develop a good speedup, scale-up, and size-up performance on very large scale datasets (He et al., 2013).
• Regression	The Ridge Regression ELM (RRELM) was developed to avoid the adverse effects caused by the perturbation or the multi-collinearity (Li & Niu, 2013).
• e-insensitive regression	ELM was formulated in 2-norm as the unconstrained optimization problem in primal variables (Balasundaram, 2013).
• Regression	A novel ELM and an evolutionary algorithm were introduced to ensure that the better hidden nodes to survive in the next generation (Feng et al., 2012).
Pattern recognition	
• Multi-label face recognition (Zong, & Huang, 2011)	The performance of the one against- all (OAA) and one-against-one (OAO) ELM was studied.
• Human face recognition (Mohammed et al., 2011)	A bidirectional two-dimensional principal component analysis (B2DPCA) and ELM were introduced.
• Human actions recognition (Minhas et al., 2010)	ELM was applied for development of a framework based on visual vocabularies.
• Handwritten character recognition (Chacko et al., 2012)	ELM was developed to classify the features of handwritten characters to accelerate the speed of leaning.
• Speaker recognition (Lan et al., 2013)	ELM was enhanced to examine the verification task.
• 3D object recognition (Nian et al., 2013)	ELM was developed to identify the inherent distribution and the dependence structure for each 3D object.
Forecasting and diagnosis	
• Retail industry (Chen & Ou, 2011)	The Gray ELM (GELM) was used to construct a forecasting model.
• Fashion retailing (Sun et al., 2008)	ELM was introduced to investigate the relationship between the sales amount and some significant factors.
• Hydraulic tube tester data (Hu et al., 2008)	Multi-stage ELM was used to improve the accuracy of clustering.
• Clinical applications (Daliri, 2012)	Fuzzy ELM was used to diagnose the lung cancer.
• Electric power systems (Xu et al., 2013)	ELM was used to assess real-time frequency stability of predictors.
Image processing	
• Surface reconstruction (Zhou et al., 2013)	An improved ELM called a polyharmonic extreme learning machine (P-ELM) was used to reconstruct a smoother surface.

Table 2 (continued)

Application	Why ELM
• Image segmentation (Pan et al., 2012a, b)	ELM classifier was trained online to simulate the visual neuron system and then extracted pixels of object from the image.
• Color image segmentation (Pan et al., 2012a, b)	ELM was used to train the pixels classifier based on the RGB color to extract object regions and provide a reference boundary of objects.

the Adaboost ensemble. The weight matrix in WELM could be updated adaptively along with the iteration of Adaboost learning.

3.3 Methods addressing speed

A novel algorithm was proposed by Zhai et al. (2017) for classification of imbalanced large datasets. The proposed algorithm included four stages: (1) alternately over-sample p times between positive class instances and negative class instances; (2) construct balanced data subsets based on the generated positive class instances; (3) train component classifiers with the ELM algorithm on the constructed balanced data subsets; (4) integrate the ELM classifiers with the simple voting approach. The experimental results showed that the algorithm can obtain promising speed-up and scalability. Yu et al. (2018a, b) presented an alternative to Weighted ELM (WELM), which is called label-WELM (LW-ELM). Unlike WELM, LW-ELM coped with class imbalance learning problems by tuning the training error of each class label. Their method provided stronger tolerance to training errors of the minority-class instances. They designed two types of weight allocation strategies, both of which are based on the class-imbalance ratio (CIR). In contrast with WELM, LW-ELM is fast and flexible.

3.4 Methods addressing other performance metrics

Imbalanced learning problem in big data was addressed by different researchers. Wang et al. (2017a, b, c) proposed a Distributed and Weighted ELM (DW-ELM) algorithm, which is based on the MapReduce framework. The paper was focused on the feasibility of parallel computation. Wang et al. (2020a, b) focused on integrating area under the Receiver Operating Characteristic (ROC) maximization into the ELM framework to tackle imbalanced binary classification tasks. They developed a new AUC-based ELM called AUC-ELM for imbalanced binary classification, which essentially is revealed to be equivalent to an ELM on another transformed data space. The proposed AUC-ELM has fewer parameters to tune. A regularized weighted circular complex-valued extreme learning machine was proposed by Shukla and Yadav (2015). In this paper, a regularized weighted circular complex-valued ELM called RWCC-ELM was developed, which incorporated the strength of both CC-ELM and WELM. The method is evaluated using imbalanced datasets taken from the Keel repository. Raghuwanshi and Shukla (2021) developed the SMOTE based class-specific kernelized extreme learning machine (SMOTE-CSKELM), which uses the Gaussian kernel function to map the input data to the feature space. The proposed work has the advantage of both the minority oversampling and the class-specific regularization coefficients.

Yu et al. (2018a, b) proposed an efficient ELM classification model, called active online-weighted ELM (AOW-ELM), to cope with the low performance or high time consumption problem. In this regard, the Weighted ELM (WELM) is selected as the basic classifier to

Table 3 Application fields of ELM (Wang et al., 2021)

Applications	Why ELM
Medical application	
• Magnetic Resonance Imaging (MRI)	ELM was used as an attention-deficit/hyperactivity disorder detection method (Peng et al., 2013). ELM was developed to detect a cocaine dependency scheme (Termenon et al., 2013). A hierarchical ELM was used to detect attention deficit/hyperactivity disorder (Qureshi et al., 2016). ELM was enhanced to detect Alzheimer's disease (Termenon et al., 2016). ELM was explored as a pathological brain detection system for magnetic resonance images (Lu et al., 2018). Compare three different classification algorithms including, support vector machine (SVM), import vector machine (IVM), and ELM, to extract features from brain MRIs (Lama et al., 2017). ELM was employed for classification (Qureshi et al., 2017). ELM was used to deal with the class imbalanced problem (Zhang et al., 2018a, b). A utilized multilayer ELM was used as the classifier and trained the network with multiclass pathological brain images (Nayak et al., 2020). ELM was used to identify Alzheimer's disease (Nguyen et al., 2019). A regularized ELM was used for brain tumor identification (Gumaei et al., 2019).
• Computerized Tomography (CT)	ELM was used to propose a 3D liver CT segmentation method (Huang et al. 2012). ELM was used as an autoencoder for image feature pre-processing (Huang et al. 2014). ELM was applied for lung disease detection (Ramalho et al., 2014). ELM was used to develop liver tumor detection and segmentation system (Zhu et al., 2016).
• Ultrasound	ELM was used for thyroid nodules classification (Xia et al., 2017). ELM was developed for thyroid nodules classification (Cai et al., 2019).
• RNA classification	An ELM-based approach was used for classification between cirRNAs and ln-cRNAs (Chen et al., 2018). An ELM was proposed for recognition of cirRNAs (Niu et al., 2020).
• Electroencephalo-gram (EEG)	ELM was proposed as an approach for epileptic EEG classification (Yuan et al., 2011). ELM was proposed as an approach for automated detection of seizure (Song et al., 2019). ELM was used to construct seizure detection (Song et al., 2016). ELM was applied for epileptic EEG classification combined with cellular automata (Zhou et al., 2018).
• Mammogram	ELM was applied for abnormality detection in Mammograms (Vani et al. 2010). ELM was developed to analyze microcalcification in digitized mammograms (Malar et al., 2012). ELM was used to detect breast cancer (Wang et al., 2014). ELM was applied as a feature selection method (Xie et al., 2016). ELM was introduced as a breast tumor detection method (Wang et al., 2016a, b). Elm was applied for feature extraction from mammograms (Hu et al., 2017). ELM was applied for breast cancer detection (Wang et al., 2019a, b).
Chemistry application	

Table 3 (continued)

Applications	Why ELM
	ELM was used to predict variables in chemical processes (Geng et al., 2017a, b). ELM was used to evaluate the green management in power generation enterprises in China (Qin et al., 2019). Elm was introduced into prediction of the nonlinear optical property (Wang et al., 2013). ELM was employed in prediction of protein-protein interactions (PPIs) (You et al., 2016; Lei et al., 2018; Li et al., 2020a, b). ELM was used to predict the toxicity of ionic liquids (Cao et al., 2018). ELM was used to predict the Henry's law constant (HLC) of CO2 (Kang et al., 2018). Elm was used to measure NOx in vehicle exhaust (Ouyang et al., 2020).
Economy application	
	ELM was used to evaluate the internationalization success of companies (Landa-Torres et al., 2012). ELM was proposed to predict economic growth (Sokolov-Mladenović et al., 2016; Marković et al., 2017) ELM was used to predict economic growth with data of agriculture, manufacturing and industry (Milačić et al., 2017). ELM was used to forecast economic growth based on information technology levels of Nations (Rakic et al., 2019). ELM was applied into analyzing CO2 emission to predict economic development (Marjanović et al., 2016; Shukla et al. 2018; Sun et al., 2017). ELM was used to forecast economic growth based on demand and price from the energy resources market (Cogoljević et al., 2018; Sánchez-Oro et al., 2016; Sun & Zhang, 2018). ELM was applied for credit risk assessment (Shoumo et al. 2019).
Transportation application	
	ELM was used to improve lifetime of transportation system and increased its reliability (Sun et al., 2011). ELM was applied for building the driver distraction detection system (Liu et al., 2015). ELM was used to predict road surface temperature (Liu et al. 2017). ELM was applied for predicting dynamic delay of large-scale railway network (Oneto et al., 2017). ELM was used to recognize the traffic sign (Zeng et al., 2016). ELM was used to analyze traffic accidents based on video data (Zhang et al. 2017). ELM was used to help taxi drivers search best routes (Wang et al., 2017a, b, c). ELM was used to predict traffic flow for drivers and governments (Liu et al., 2018; Xing et al. 2018).
Robotics application	
	ELM was used to perform object grasping detection (Sun et al. 2015). ELM was applied for robotic arms control (Alcin et al. 2016). ELM was used to classify electroencephalographic (EEG) signals (She et al., 2018). ELM was applied for robotic motion control (Duan et al., 2017).
IOT application	
• Intrusion Detection Systems	ELM was applied for detection of cyber-attack (Rathore & Park, 2018). ELM was used to assign bug fixing jobs to engineers (Yin et al., 2018). ELM was used to analyze distributed denial-of-service (DDoS) attacks (Li et al. 2019).
Geography application	

Table 3 (continued)

Applications	Why ELM
	ELM was used to generate the landslide susceptibility indexes (Huang et al., 2017). ELM was proposed to evaluate the stability of rubble mound breakwaters (Wei et al., 2019). ELM was used to predict the sediment-carrying capacity (Wei et al., 2019). ELM was applied for deriving the operation rule of hydro power reservoir (Niu et al., 2019). ELM was used to extract local features (Li et al., 2015).
Food industry application	
	ELM was proposed as a food safety monitoring System (Geng et al., 2017a, b). ELM was applied as a classifier for detecting amino acid nitrogen content in soy sauce (Ouyang et al., 2013). ELM was applied for classification of different kinds of wines (da Costa et al., 2018). ELM was applied for prediction during large-scale food sampling analysis (Liu et al. 2017). ELM was applied for prediction of dairy food safety (Zhang et al., 2018a, b). ELM was used to differentiate kernel-damaged wheat from kernel-undamaged wheat (Guo et al., 2019).

guarantee the impurity of instance selection in the procedure of active learning, and an efficient real-time updated model of WELM was deduced theoretically. In order to improve the learning performance of classical ELMs for imbalanced data learning, Ri et al. (2020) presented a novel variant of the ELM algorithm based on a hybrid cost function which employs the probability of given training samples belonging to each class to calculate the G-mean. An enhanced kernel-based multilayer fuzzy weighted extreme learning machine (EML-KFWELM) was proposed by Wang et al. (2020a, b) to deal with the disadvantages of kernel-based multilayer extreme learning machines (ML-KELM). They embedded the weighted strategy into ML-KELM to enhance the classification performance of the minority class, and proposed fuzzy membership to eliminate classification error caused by outliers and noise samples. Then, an enhanced grey wolf optimization (EGWO) method was developed to perform the parameters optimization and improve the generalization performance of ML-KELM. Zhang et al. (2022) proposed an extreme learning machine algorithm with output weight adjustment called OWA-ELM, which can make the decision boundary of ELM move to majority classes, and improve the classification performance of imbalanced data.

4 ELM for outlier detection

4.1 Methods addressing accuracy in outlier detection

Zhan and Luo (2015) addressed the outlier robustness of ELM (ORELM) regression problems, where a norm loss function is used to enhance the robustness. Zhang et al. (2018a, b) established the neural network model for prediction of silicon content in hot metal based on ELM algorithm. In this respect, an outlier detection method based on W-ELM was proposed from a statistical view and the ordinary ELM and W-ELM algorithms are modified to reduce the interference of outliers.

A novel wind speed forecasting based on hybrid decomposition and online sequential outlier robust ELM (OSORELM) was proposed by Zhang et al. (2019). During the data preprocessing phase, the wind speed is decomposed by hybrid mode decomposition (HMD). The Crisscross Algorithm (CSO) is applied to optimize OSORELM input weights and hidden layer biases that affect predictive performance. A novel optimized predictive model for detecting anomalies in aerospace using the Grey Wolf Optimization (GWO) algorithm and an ELM, called GWO-ELM was introduced by Abdelghafar et al. (2019). The proposed GWO-ELM is used to find anomalous events by comparing the actual observed values with the predicted intervals of telemetry data; the GWO is applied to optimize the ELM's input weights and the bias parameters of hidden neurons to improve its prediction accuracy and ability to detect anomalies. Oikawa et al. (2020) studied the effectiveness of OS-ELM based anomaly driving behavior detector using sensor data of vehicles and compared the performance of it with a Hidden Markov Model (HMM) based and traditional Long Short-Term Memory (LSTM) based methods. Altunay et al. (2021) studied anomaly-based intrusion detection systems consisting of convolutional neural network (CNN), autoencoder (AE), deep belief network (DBN), long short-term memory network (LSTM), or various combinations of these methods on the supervisory control and data acquisition networks (SCADA).

4.2 Methods addressing robustness in outlier detection

Horata et al. (2013) proposed the Extended Complete Orthogonal Decomposition (ECOD) method to solve the computational problem in ELM weights computing via ECODLS algorithm. They also proposed the other three algorithms, i.e. the iteratively reweighted least squares (IRWLS-ELM), ELM based on the multivariate least-trimmed squares (MLTS-ELM), and ELM based on the one-step reweighted MLTS (RMLTS-ELM) to solve the outlier robustness problem. Barreto and Barros (2016) developed a robust ELM (RELM) for pattern classification with outliers. The RELM is designed using M-estimators to compute the output weights instead of the standard ordinary least square (OLS) method.

Naseer et al. (2018) proposed conventional machine learning-based intrusion detection models implemented using well-known classification techniques, including extreme learning machine, nearest neighbor, decision-tree, random-forest, support vector machine, naive-bays, and quadratic discriminant analysis. Zhang et al. (2020) proposed robust ELM (RELM) to improve the modeling ability and robustness of Gaussian and non-Gaussian noise. RELM uses Gaussian mixing (MoG) to create objective functions modified to fit the noise. Furthermore, a solution for new objective function is developed based on the Expected Value Maximization Algorithm (EM). A novel anomaly detection framework based on one-class extreme learning machine (OC-ELM) for the multimode system was presented by Chen et al. (2021).

4.3 Methods addressing speed in outlier detection

Janakiraman and Nielsen (2016) used ELM's fast training and good generalization properties to develop a scalable anomaly detection algorithm for very large datasets. They tuned unsupervised ELM algorithms such as autoencoders and embedded models to detect anomalies. Hashmi and Ahmad (2019) proposed an optimal Replicator Neural Network, called Garson-pruned extreme learning machine based replicator neural network (GP-ELM-RRN), which is optimized using ELM learning and Garson algorithm for anomaly detection. From the experiments it is evident that Garson-pruned ELM-RNN on TensorFlow is the best approach for anomaly detection in terms of accuracy as well as speed. The ELM-AD method proposed by Sridhar and Sanagavarapu (2021) is to identify market manipulation in the market price and volume data.

This section is briefly summarized as Table 4, which shows ELM for outlier detection including its applications, metrics, and techniques. It can be seen that ELMs have been employed in considerably different areas of outlier detection, including intrusion detection systems, health monitoring in space, wind speed forecasting, stock markets, etc. Moreover, researchers have applied various metrics to evaluate the proposed ELMs, where the area under the curve (AUC), which provides a good reference to quantify the model, receiver operating characteristics (ROC) curve, which measures the detection performance of models, and root-mean-square error (RMSE) were among the most popular metrics.

4.4 Methods addressing other performance metrics in outlier detection

Yan (2016) adopted ELMs to a new application domain - industrial machine condition monitoring. More specifically, one-class ELMs are applied for more accurate anomaly detection of gas turbine combustors. The overall combustor anomaly detection system includes two primary functions, deep feature learning and one-class classification. A new definition of outlier, called cluster based outlier was proposed by Wang et al. (2016a, b). The method is more suitable for the complicated datasets that consist of many clusters with different densities. Their method was implemented using unsupervised ELMs.

A probabilistic regularized ELM (RELM) was proposed by Lu et al. (2017) to improve modeling performance with data containing non-Gaussian noise and/or outliers. Wang et al. (2017a, b, c) proposed a new distributed algorithm for the cluster based outlier detection (DACB). On the master node, they collected a small number of points from the slave nodes to obtain a threshold. On each slave node, they designed a new filtering method which uses the threshold to efficiently speed up the computation. They also proposed a ranking method to optimize the order of cluster scanning. Hybrid architecture was proposed by Mi et al. (2017) to forecast wind speed. Regarding the architecture, the wavelet domain denoising is adopted to reduce the noise of the original wind speed series, and a secondary decomposing algorithm is used to reduce the intermittency of the original wind speed series. Furthermore, the auto regressive moving average and ELM models are employed to complete the multi-step forecasting computation for the decomposed stationary sub-layers and intrinsic mode functions, respectively. In addition, the new outlier correction method is developed to guarantee the robustness of the built auto regressive moving average and ELM models during their forecasting computation.

Siqi et al. (2019) proposed a novel Recurrent Adaptive Reconstruction ELM (RAR-ELM) to eliminate outliers from noisy data in a fast and unsupervised manner. The RAR-ELM includes three components: reconstruction ELM, adaptive labeling and recurrent training. They also proposed an Online Sequential RAR-ELM (OS-RAR-ELM) which implements with an online or sequential mode and makes RAR-ELM easily applicable to massive noisy data or online sequential data.

Table 5 shows the taxonomy of cutting-edge ELMs to detect outliers. These methods are classified under three different machine learning perspectives which are supervised, unsu-

Table 4 A brief description of ELM for outlier detection

Application	Metric	Technique
Intrusion detection systems (IDS) (Naseer et al., 2018)	AUC, Precision-Recall Curve, mean average precision and accuracy of classification	Conventional machine learning intrusion detection system models were implemented with different well-known classification techniques, including ELM, k-NN, Decision-Tree, Random- Forest, Support Vector Machine, Naive-Bays, and QDA.
Unknown and complex systems (Lu et al., 2017)	RMSE	The method constructed a new objective function to minimize both mean and variance of the modeling error.
Multimode system (Chen et al., 2021)	ROC, AUC	In the training phase, the multimode dataset was divided into several subsets according to the identified operation modes. Then the corresponding detection models were established respectively under different operation modes by OC-ELM.
Industrial machine condition monitoring (gas turbine combustor) (Yan 2016)	ROC, AUC	Deep learning technology was applied to learn features out of the raw sensor measurements. Specifically stacked denoising autoencoder was used. Additionally, for the gas turbine combustor anomaly detection application concerned in the paper, two different one-class ELMs namely, non-kernel one-class ELM and kernel one-class ELM, were evaluated.
Cluster based outliers (Wang et al., 2016a, b)	Runtime, disk IO cost	To detect cluster-based outliers, an unsupervised extreme learning machine was used to cluster the data in the given set. Then, a pruning method to reduce the k-nearest neighbors algorithm (k-NN) searching space was designed.
Silicon content in hot metal (Zhang et al., 2018a, b)	Accuracy, delay time under the condition of maximum correlation, and mean square error (MSE)	Two schemes were presented to deal with outliers. One was focused on the outlier detection from a statistical point, and the other was proposed to reduce the interference of outliers.
Unsupervised outlier removal in practical applications (Siqi et al., 2019)	Precision, recall, and $f1$-mean	RAR-ELM recurrently learned to reconstruct data and automatically excluded those data with high reconstruction errors as outliers by a novel adaptive labeling mechanism.
Distributed environments (Wang et al., 2017a, b, c)	Time cost, and the network transmission quantity (NTQ)	The algorithm adopted master slave architecture. The master node can monitor the points with large weights on each slave node and computed a threshold. An optimization method was used to improve the performance of the threshold.
Supervised pattern classification (Barreto & Barros, 2016)	Batch and recursive learning rules	A model selection strategy based on PSO was introduced to find an optimal architecture for datasets contaminated with non-Gaussian noise and outliers.
Anomaly detection in large datasets (Hashmi & Ahmad, 2019)	Accuracy, specificity, and execution speed	ELM-based RNN solved the problem of determining the number of hidden layers in RNN, and the problem to determine the number of hidden layer neurons was solved by Garson's algorithm.
Health monitoring in space (Abdelghafar et al., 2019)	RMSE, mean absolute error (MAE), confusion matrix, the CPU time, and mean and standard deviation	To obtain the optimized predictive model GWO was applied to find the optimal ELM parameters. GWO can find the best set of input weights and biases over a set of iterations. Then, the best prediction performance was achieved by finding the optimal output weights and hidden layer output matrix values to obtain the optimal target output.

Table 4 (continued)

Application	Metric	Technique
Wind speed forecasting (Mi et al., 2017)	MAE, RMSE, and Mean Absolute Percentage Error (MAPE)	A new hybrid method was proposed for the wind speed multi-step forecasting. The wavelet decomposition was adopted to reduce the noise of the original data. The different decomposing algorithms were used to decompose the original data. The different forecasting models were built to predict the pre-processed data. The outlier correction method was proposed to correct the wrong predictions.
Safety driving (anomaly behavior of drivers or vehicles) (Oikawa et al. 2020)	Accuracy, sequential learning speed, anomaly score	An anomaly driving behavior detection technique using OS-ELM was proposed and the anomaly detection accuracy of the method was compared with an ordinary Incremental HMM (IncHMM) and LSTM based method.
Outlier robustness problem (Horata et al., 2013)	The mini-max probability machine regression (MPMR), accuracy, and training time	The outlier robustness problem was enhanced using three algorithms: IRWLS-ELM, MLTS-ELM, and RMLTS-ELM.
Outlier robustness of ELM in regression problems (Zhan and Luo 2015)	Outlier robustness, computational efficiency, and RMSE	An augmented Lagrange multiplier based ELM algorithm, namely ORELM was proposed for solving the outlier robust regression problems. The ORELM method not only keeps the advantage of fast training speed but also shows notable performance in dealing with outliers.
Unknown noise in gas utilization and iron making process (Zhang et al., 2020)	RMSE	Robust objective functions were developed based on MoG to improve modeling capabilities with complex and unknown noise. The ELM algorithm was implemented to help get the optimal parameters.
Wind speed forecasting (Zhang et al., 2019)	MAE, RMSE, and MAPE	The HMD an effective method was used to deeply decompose the original wind speed. Then, an online robust model OSORELM was developed to forecast the multi-step wind speed with better performance, and the parameters of OS-ORELM were adapted by CSO.
Aviation safety problem (Janakiraman and Nielsen 2016)	AUC and training speed	ELM based anomaly detection was considered and three extensions were developed.
supervisory control and data acquisition (Altunay et al. 2021)	Accuracy	Architectures use softmax function, extreme learning machine, deep belief networks, and multilayer detectors in the classification process.
Stock markets (Sridhar and Sanagavarapu 2021)	Accuracy, Precision, Recall, F1-measure, ROC, and AUC	ELM-AD was proposed for detection of anomaly real-time anomalies in stock market manipulations field. To this purpose, the feature selection method was proposed for price and volume manipulation.

pervised, and semi-supervised approaches. It can be observed that a significant majority of methods are focused on supervised learning, while small minorities of approaches use unsupervised and semi-supervised techniques.

Table 6 depicts the developments of ELMs to detect outliers based on the most popular standardized quality metrics. It can be seen that performance, accuracy, robustness, and speed were among the most popular metrics, respectively.

5 Datasets for outlier detection using ELMs

Datasets have laid the foundation for training, scoring, and benchmarking machine learning models and have played a fundamental role in progress in this area (Paullada et al., 2021) since they are the basis for facilitating the development of multiple computing domains and providing range, robustness, and reliability of results (Dekker, 2006). In this section, we survey recent issues pertaining to datasets in imbalanced data research, focusing primarily on work in ELM and outlier detection. The selected datasets are conducted by researchers whose papers are discussed in the previous section.

5.1 Datasets for improving accuracy in outlier detection

The ORELM method proposed by Zhan and Luo (2015) is conducted on Abalone and Housing for two regression tasks as well as breast cancer diagnosis for one binary classification task. The Abalone dataset is used for predicting the age of abalone from eight physical measurements. The housing dataset concerned the relationship between housing values and some attributes such as average of rooms per dwelling and per capita crime rate by town.

Table 5 Taxonomy of cutting-edge ELMs to detect outliers based on machine learning techniques

Ref	Method Name	Supervised	Unsupervised	Semi-supervised
Naseer et al. (2018)	Anomaly detection based on different deep learning NNs	✓		
Lu et al. (2017)	RELM	✓		
Chen et al. (2021)	OC-ELM		✓	
Yan (2016)	One-Class ELM	✓		
Wang et al. (2016a, b)	Cluster based outliers		✓	
Zhang et al. (2018a, b)	E-ELM & EW-ELM	✓		
Siqi et al. (2019)	RAR-ELM		✓	
Wang et al. (2017a, b, c)	DACB		✓	
Barreto and Barros (2016)	RELM/PSO	✓		
Hashmi and Ahmad (2019)	GP-ELM-RNN		✓	✓
Abdelghafar et al. (2019)	GWO-ELM	✓		
Mi et al. (2017)	WDD-WPD-ARMA(SS) EMDELM (NS)-OCM	✓		
Oikawa et al. (2020)	OS-ELM			✓
Horata et al. (2013)	ECOD	✓		
Zhan and Luo (2015)	Outlier-robust ELM	✓		
Zhang et al. (2020)	R-ELM	✓		
Zhang et al. (2019)	HMD-OSORELM-C	✓		
Janakiraman and Nielsen (2016)	L-ELMAD		✓	
Altunay et al. (2021)	Deep learning methods in SCADA systems		✓	
Sridhar and Sanagavarapu (2021)	ELM-AD	✓		

Table 6 Taxonomy of cutting-edge ELMs to detect outliers based on quality metrics

Ref	Accuracy	Robustness	Speed	Other performance metrics
Naseer et al. (2018)		✓		
Lu et al. (2017)		✓		✓
Chen et al. (2021)		✓		
Yan (2016)				✓
Wang et al. (2016a, b)				✓
Zhang et al. (2018a, b)	✓			✓
Siqi et al. (2019)				✓
Wang et al. (2017a, b, c)				✓
Barreto and Barros (2016)		✓		
Hashmi and Ahmad (2019)	✓		✓	
Abdelghafar et al. (2019)	✓			✓
Mi et al. (2017)				✓
Oikawa et al. (2020)	✓		✓	✓
Horata et al. (2013)		✓		
Zhan and Luo (2015)	✓	✓		
Zhang et al. (2020)		✓		✓
Zhang et al. (2019)	✓	✓		✓
Janakiraman and Nielsen (2016)	✓		✓	
Altunay et al. (2021)	✓			
Sridhar and Sanagavarapu (2021)	✓		✓	

Naseer et al. (2018) trained deep models on NSLKDD training datasets (NSLKDDTrain20p and NSLKDDTrain+) and tested on NSLKDD test datasets (NSLKDDTest+and NSLKDDTest21). The NSLKDD Dataset is available in four partitions. Two partitions namely NSLKDDTrain20p and NSLKDDtrain+serve as training data for model learning and provide 25,192 and 125,973 training records respectively. Remaining two partitions called NSLKDDTest+and NSLKDDTest21 are available for performance evaluation of trained models and provide 22,543 and 11,850 instances respectively. Additionally, NSLKDDTest21 contains records for attack types not available in other NSLKDD train and test datasets. These attack types include processtable, mscan, snmpguess, snmpgetattack, saint, apache2, httptunnel, back and mailbomb. Zhang et al. (2018a, b) presented the simulation results based on the real iron-making data. The real production data collected in a blast furnace with 2500 m^3 are employed in the experiment. They chose 1205 sets of silicon content data. Additionally, they employed the proposed frameworks into two real-world applications: one regression application for the Abalone dataset and one classification dataset for the breast cancer dataset. Both datasets are available via UCI machine learning repository.

The GWO-ELM method proposed by Abdelghafar et al. (2019) is conducted on the NASA shuttle valve dataset to evaluate the efficiency of the proposed model for anomaly detection from space application telemetry data. The shuttle valve dataset is compiled from health monitoring measurements for electromagnetic valves in the space shuttle under various conditions; voltage, temperature, and impedance. The dataset is divided into training and test sets, where the training set includes the normal samples that had been recorded at the normal range of voltage, temperature, and poppet impedance, then tested on some normal samples merged with abnormal samples which are recorded at outlier voltage, high temperature (69–71 ∘C) or poppet impedance values at 9 or 45 mils. The normal range of voltage is 18, 20, 22, 24, 26, 28 or 30 V, while the outlier is considered at 14, 16 and 32 V. The OSORELM method proposed by Zhang et al. (2019) was conducted on a real-world dataset. Simulations using raw wind speed data (latitude: 31.19 N, longitude: 102.24 W) from a wind farm in Texas, USA, collected at 10-minute intervals to validate the performance of the proposed hybrid model and use for testing. Since wind speed data varies greatly from season to season, the months of January, April, July, and October are randomly selected for each of the 2004 seasons.

Oikawa et al. (2020) created a new dataset, called the Wheel-DriveSet with a powered wheelchair for anomaly behavior detection experiments. This dataset provides wheelchair driving data obtained from three different drivers with three different routes. All drivers are male and 22–24 years old. They used an electric wheelchair of WHILL Model CR which is controlled by a joystick equipped on the front of the right arm support. The driving courses are flat pedestrian paths on a university campus. The drivers made six laps per course; abnormal driving behaviors are defined as zigzag driving and intense joystick operations. Altunay et al. (2021) studied two different datasets including NSL-KDD and The Secure Water Treatment tested (SWaT).

5.2 Datasets for improving robustness in outlier detection

The experiments of proposed method by Horata et al. (2013) are conducted on both toy and real-world datasets. The experiments are divided into two categories. The first category is conducted on a generated dataset and an example of a real-world dataset, Abalone. Furthermore, the problems in the second are divided into two types, toy and the real-world regression problems in which Abalone, Boston, and Protein datasets are used.

The proposed method by Barreto and Barros (2016) is simulated on synthetic and real-world datasets. They considered a synthetic two-dimensional dataset generated according to a pattern of two intertwining moons. Pattern vectors from Class 1 receives "+1" labels, while the ones from Class 2 are tagged with "−1" labels. The first time they were trained with the outlier-free dataset with $N=1000$ samples. The second time, they were trained with $N_{out} = 20$ outliers added to the original dataset. It should be noted that all data samples are used for training the classifiers. Next, the robustness of the RELM-B classifier is evaluated on two real-world benchmarking datasets, Ionosphere and the Wisconsin Diagnostic Breast Cancer (WDBC), which are publicly available for download from the UCI Machine Learning Repository website. The Ionosphere dataset describes a classification task where radar signals target two types of electrons in the ionosphere. The WDBC dataset describes patients diagnosed with breast cancer or not. Additionally, they changed the type of data to evaluate the method. In this respect, the Yale-A and Sussex face image databases are chosen. The Yale-A face image database includes 165 grayscale images in GIF format of 15 individuals. There are 11 images per subject, one per different facial expression or configuration: center-light, with glasses, happy, left-light, without glasses, normal, right light, sad, sleepy, surprised, and wink. The Sussex face image database includes 100 images of 384×287 pixels. These images are in grayscale in the Sun Standard *Rasterfile* format. This database includes 10 subjects and each subject is seen under 10 different poses.

Zhang et al. (2020) used several selected benchmark datasets to validate the proposed RELM. The selected three regression datasets, including Auto-MPG, Abalone, California Housing, and a time series prediction, i.e., Mackey–Glass, are available via the UCI Machine Learning Repository. Next, they evaluated the effectiveness of RELM using two real-world applications, including gas utilization (GUR) prediction and hot metal silicon content (HMSC) prediction in the blast furnace iron making process. The method proposed by Chen et al. (2021) is verified on a public dataset of aircraft engines. The effectiveness of the proposed method is verified on the Commercial Modular Aero-Propulsion System Simulation (C-MAPSS) dataset. This dataset consists of 260 multivariate time series. It contains four subsets, and the subset involving a single failure mode and 6 operating conditions (FD002) is adopted here. This subset consists of a training set (train_FD002) and a testing set (test_FD002).

5.3 Datasets for improving speed in outlier detection

The proposed methods by Janakiraman and Nielsen (2016) were conducted on a real aviation safety benchmark problem. The aviation data including radar measurements which are recorded at Denver Terminal Radar Approach Control Facility (TRACON) were stored by the Performance Data Analysis and Reporting System (PDARS) program. PDARS also provides additional features to enhance the study, including: runway detection and flight separation function calculation. The GP-ELM-RNN method proposed by Hashmi and Ahmad (2019) are conducted on openly available UCI/ODDS datasets: lymphography (outliers ratio: 6/142), wisconsin breast cancer (outlier ratio: 39/444), post-operative (outliers ratio: 26/64), pageblocks (outliers ratio: 140/4913), credit card fraud detection (outliers ratio: 492/284,315), forest cover (outliers ratio: 2747/283,301) and kddcup99 (outliers ratio: (2211/567,497). Sridhar and Sanagavarapu (2021) conducted the ELM-AD method on a daily trading data which were extracted from the Bombay Stock Exchange (BSE) based on the affidavit information provided by the Security and Exchange Board of India (SEBI) for companies that involves in stock market manipulations. Stock market manipulations include court orders, show cause notices, release orders, adjudication orders, etc.

To facilitate the future studies in this quite promising field, we will list the related information about datasets detailed in this section. For convenience, Table 7 summarizes commonly used datasets and provides the brief information for each dataset (such as source, availability, and URL). It shows that proposed methods were employed in forty different datasets for evaluation.

Turning to Fig. 4, the pie chart details the availability of datasets for outlier detection using ELMs. Overall, approximately three-quarters of them are widely available while almost one-quarter of the pie chart is accounted for Unavailable datasets. Moreover, a tiny fraction of data is generated synthetically, and their details are accessible.

5.4 Datasets for improving other performance metrics in outlier detection

Yan (2016) used a database which includes several years of thermal couples (TC) measurements sampled at once-per-minute for a fleet of gas turbines. For demonstration purpose approximately a year's worth of data collected for one specific turbine that has 27 combustor chambers is used. In the study, 13,791 normal samples for unsupervised feature learning are used, and the rest of data (300 abnormal and 47,575 normal samples) for training and testing the one-class classifiers considered are used.

Five regression datasets from UCI, including Servo, Stock, California Housing, Airfoil Self-Noise, and Energy Efficiency, were used by Lu et al. (2017) to evaluate the probabilistic RELM method. The Servo dataset concerns a robot control problem. Data are from a simulation of a servo system involving a servo amplifier, a motor, a lead screw/nut, and a type of sliding carriage. The Stock dataset is the daily stock prices for ten aerospace companies from January 1988 through October 1991. The California Housing dataset contains details regarding households in California, including median house value, median income, household median age, total rooms, total bedrooms, population, households, latitude, and longitude. The Airfoil Self-Noise dataset comprises different size airfoils at various wind tunnel speeds and angles of attack. The Energy Efficiency provides 12 different building shapes simulated in Ecotect.

Wang et al. (2016a, b, 2017a, b, c) used a synthetic dataset for the experiments. In detail, given the data size N, $N/1000-1$ clusters are generated and randomly assigned each of them a center point and a radius. In average, each cluster includes 1000 points following Gaussian distribution. At last, the remaining 1000 points are scattered into the space. The architecture proposed by Mi et al. (2017) is conducted on five groups of actual wind speed time series including 700 samples. The 1st-600th samples of these groups of wind speed time series data are utilized to build the various forecasting models, while the leaving 601st–700th samples in each group are used to evaluate the performance of the built models. Siqi et al. (2019) demonstrated the effectiveness of RAR-ELM to the task of outlier image removal on five frequently-used data collections: the Caltech101 dataset, the cifar-10 dataset, The Street View House Numbers (SVHN) dataset, the MNIST dataset, and the Fashion dataset.

6 Discussion

We depicted the state-of-the-art ELM algorithms for class imbalance. However, these algorithms involve tuning the training error, mapping features, adapting distribution weights and bias, promoting classification robustness, the feasibility of parallel computing, performance and accuracy improvement, stability, and scalability.

Table 7 Related information about datasets

Dataset	Source	Available	URL
NSL-KDD	Naseer et al. (2018) Altunay et al. (2021)	✓	http://www.di.uniba.it/~andresini/datasets.html
Servo	Lu et al. (2017)	✓	https://archive.ics.uci.edu/ml/datasets/Servo
Stock	Lu et al. (2017)	✓	https://archive.ics.uci.edu/ml/datasets/Stock+portfolio+performance
California Housing	Lu et al. (2017) Zhang et al. (2020)	✓	https://scikit-learn.org/stable/modules/generated/sklearn.datasets.fetch_california_housing.html
Airfoil Self-Noise	Lu et al. (2017)	✓	https://archive.ics.uci.edu/ml/datasets/airfoil+self-noise
Energy Efficiency	Lu et al. (2017)	✓	https://archive.ics.uci.edu/ml/datasets/energy+efficiency
C-MAPSS	Chen et al. (2021)	Currently Unavailable	https://data.nasa.gov/widgets/xaut-bemq
Turbine	Yan (2016)	Unavailable	
Synthetic	Wang et al. (2016a, b) Wang et al. (2017a, b, c)	Details are available	
Iron-making	Zhang et al. (2018a, b)	Unavailable	
Caltech101	Siqi et al. (2019)	✓	https://www.tensorflow.org/datasets/catalog/caltech101
Cifar-10	Siqi et al. (2019)	✓	https://www.cs.toronto.edu/~kriz/cifar.html
Street View House Numbers (SVHN)	Siqi et al. (2019)	✓	http://ufldl.stanford.edu/housenumbers/
MNIST	Siqi et al. (2019)	✓	http://yann.lecun.com/exdb/mnist/
Fashion	Siqi et al. (2019)	✓	https://github.com/zalandoresearch/fashion-mnist
Synthetic	Barreto and Barros (2016)	Details are available	
Ionosphere	Barreto and Barros (2016)	✓	https://archive.ics.uci.edu/ml/datasets/ionosphere
WDBC	Barreto and Barros (2016)	✓	https://archive.ics.uci.edu/ml/datasets/breast+cancer+wisconsin+(diagnostic)
Yale-A	Barreto and Barros (2016)	✓	http://vision.ucsd.edu/content/yale-face-database
Sussex face	Barreto and Barros (2016)	✓	https://link.springer.com/article/10.1007/BF02311576
Lymphography	Hashmi and Ahmad (2019)	✓	https://archive.ics.uci.edu/ml/datasets/Lymphography
Post-Operative	Hashmi and Ahmad (2019)	✓	https://archive.ics.uci.edu/ml/datasets/Post-Operative+Patient

Table 7 (continued)

Dataset	Source	Available	URL
Wisconsin	Hashmi and Ahmad (2019)	✓	https://archive.ics.uci.edu/ml/datasets/breast+cancer+wisconsin+(diagnostic)
Page-blocks	Hashmi and Ahmad (2019)	✓	https://archive.ics.uci.edu/ml/datasets/Page+Blocks+Classification
Credit Card	Hashmi and Ahmad (2019)	✓	https://archive.ics.uci.edu/ml/datasets/default+of+credit+card+clients
Forest Cover	Hashmi and Ahmad (2019)	✓	https://archive.ics.uci.edu/ml/datasets/covertype
Kddcup99	Hashmi and Ahmad, 2019)	✓	http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html
NASA shuttle valve	Abdelghafar et al. (2019)	✓	https://cs.fit.edu/~pkc/nasa/data/
Wind speed time series	Mi et al. (2017)	Unavailable	
Wheel-DriveSet	Oikawa et al. (2020)	Unavailable	
Abalone	Horata et al. (2013) Zhang et al. (2020)	✓	https://archive.ics.uci.edu/ml/datasets/abalone
Boston	Horata et al. (2013)	✓	https://archive.ics.uci.edu/ml/machine-learning-databases/housing/
Protein	Horata et al. (2013)	✓	https://archive.ics.uci.edu/ml/datasets/Protein+Data
Auto-MPG	Zhang et al. (2020)	✓	https://archive.ics.uci.edu/ml/datasets/auto+mpg
Mackey–Glass	Zhang et al. (2020)	✓	https://archive.ics.uci.edu/ml/datasets/glass+identification
Raw wind speed data	Zhang et al. (2019)	Unavailable	
Aviation data (radar measurements)	Janakiraman and Nielsen (2016)	Unavailable	
SWaT	Altunay et al. (2021)	Unavailable	
BSE	Sridhar and Sanagavarapu (2021)	✓	https://www.kaggle.com/ravisane1/5-year-bse-sensex-dataset

The "ELMs for outlier detection" are based on several criteria, such as performance, reducing the interference of outliers, robustness, accuracy, training speed, input weights, bias parameters, computational problems, and scalability. It is challenging to optimize them. More researchers preferred robustness and parameters optimization over other criteria as the main issues. Since the performance of robust methods relies on the accuracy of weight estimation. The proposed methods also impose a high computational cost, which often reduces the effectiveness of ELMs. However, we contend that there are significant gaps, and they could be addressed to effectively cope with major challenges which have affected ELMs. These matters are listed as follows:

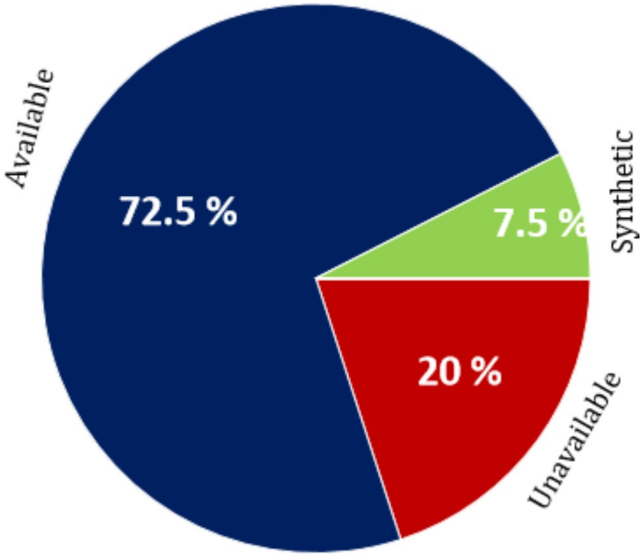

Fig. 4 Availability of datasets for outlier detection using ELMs

1. Limitations that may deteriorate the ELM performance metric should be addressed, e.g., the unbalanced distribution of training and testing datasets. This aspect requires further research on flexible approaches when the learning model deals with large datasets.
2. Although ELM is originally proposed for two or more-class classifications and regression, One-class ELMs have emerged recently. Researchers can introduce one-class ELMs to new application domains, particularly industrial fields. Raw data with imbalanced class distribution can be found almost everywhere, especially in industrial applications.
3. ELM has superior results in terms of robustness and generalization performance than state-of-the-art machine learning approaches, including SVM, k-NN, Naïve Bayes, and Artificial neural Network (ANN). Moreover, it is efficient and effective for both classification and regression. However, it is interesting to study what strategies can be taken to improve the generalization and reliability of the ELM for outlier detection.
4. The ELM avoids many challenges that affect other gradient-based learning methods, such as learning rate, local minima and epoch selection, and termination criteria. What criteria should be considered to select a suitable evolutionary algorithm for optimizing the ELM? It seems that evolutionary algorithms which are used to optimize ELMs can be evaluated based on time complexity and stability. These metrics can be assessed using the CPU time required for training and the mean and the standard deviation of the fitness evaluation, respectively.
5. More attention should be given to improve the performance of ELMs without parameter optimization because parameter optimization can negatively impact model sensitivity.
6. The algorithm for anomaly detection on large datasets should be fast, accurate, and scalable. Whether the accuracy of the learning model is more important than the calculation time or not should be answered. Since in some cases such as intrusion detection, calculation time is as important as the model accuracy.

7. Model error is one of the main obstacles to improved accuracy and reliability. To enhance the robustness of ELMs, the distribution of modeling errors can be evaluated to ensure consistency between the modeling errors and the outlier distribution.
8. Which batch and sequential learning methods can be introduced to build a robust ELM for semi-supervised and unsupervised pattern classification when outliers are present in the data?
9. Deep learning algorithms can be developed for feature extraction between anomalies but there is very limited research on jointly exploiting the challenges of semi-supervised and unsupervised learning approaches regarding deep learning. To resolve this issue, studies may need to extend both semi-supervised and unsupervised cases.
10. Further attention needs to be paid to the development of multimode anomaly detection for industrial processes. To this purpose, reducing time cost, achieving stronger distinguishing performance and greater stability may lead to desired results.
11. In a distributed environment with an increase in the data size, more points need to be scanned to find the outliers, and it may lead to increasing the network transmission quantity. Therefore, the ELM model should be stable after a certain number of calculations and should not be affected by the data size.
12. In a recurrent training process, when the reconstruction error distribution (RED) of normal data and outliers overlap with each other, it may lead to misclassification. Thus, which strategy can be taken to make the RED more separable.

Contributions to outlier detection using ELMs can be categorized based on three main areas: theoretical foundations, knowledge discovery, and imbalanced domains. Table 8 shows the proposed methods make a major contribution to imbalanced domains. In this area, environmental applications and anomaly detection are the most popular topics among researchers, while fraud detection and health applications have been less studied. Alternatively, other imbalanced domains, such as social media applications and deep fake classification, have not yet been explored. Theoretical foundations are the second most popular topic. We realize that pre-processing, deep learning, big data, and one-class learning have been studied in this field. Future work can consider post-processing approaches, feature selection, and transformation methods for outlier detection using ELMs. Knowledge discovery has been less investigated. Imbalanced regression and automated machine learning have been studied while researchers can conduct more thorough studies on improving the remaining topics in this area, such as lifelong machine learning, graph classification, imbalanced time series, and learning with imbalanced data streams.

We illustrate that a few of papers have explored cutting-edge ELMs to detect outliers based on semi-supervised learning. Researchers can explore supervised and un-supervised approaches for outlier detection using ELMs. Moreover, the improvement of ELMs based on supervised and semi-supervised learning can be investigated simultaneously. In addition, we present that there are significant developments of cutting-edge ELMs to detect outliers in terms of accuracy and robustness. Oikawa et al. (2020) have simultaneously developed accuracy, speed, and other performance metrics (e.g., AUC). Moreover, Zhang et al. (2019) have simultaneously developed accuracy, robustness, and other performance metrics (e.g., mean absolute percentage error and mean absolute error). However, there is still no work to simultaneously enhance robustness and speed metrics.

Table 8 Topics of interest for outlier detection using ELMs

Ref	Theoretical foundations	Knowledge discovery	Imbalanced domains
Naseer et al. (2018)			✓
Lu et al. (2017)		✓	
Chen et al. (2021)	✓		
Yan (2016)	✓		✓
Wang et al. (2016a, b)	✓		
Zhang et al. (2018a, b)			✓
Siqi et al. (2019)		✓	
Wang et al. (2017a, b, c)			✓
Barreto and Barros (2016)		✓	
Hashmi and Ahmad (2019)	✓		
Abdelghafar et al. (2019)			✓
Mi et al. (2017)			✓
Oikawa et al. (2020)			✓
Horata et al. (2013)	✓		
Zhan and Luo (2015)		✓	
Zhang et al. (2020)			✓
Zhang et al. (2019)			✓
Janakiraman and Nielsen (2016)			✓
Altunay et al. (2021)	✓		
Sridhar and Sanagavarapu (2021)			✓

7 Conclusion

Recently, ELM techniques have received considerable attention in computational intelligence and machine learning communities, in both theoretical study and applications due to their fast speed, easy implementation, and great potential to resolve regression and classification problems. Despite recent advances in ELM and its growing popularity, there are few empirical studies in the field of ELM with outlier detection.

We investigated the suitability of ELMs-based approaches for outlier detection and the reasons why extensive research on ELMs has been carried out. Continuously, we filled several research gaps during the literature review of the outlier detection problem, which are: (1) lack of investigation of well-known ELM approaches for outlier detection, (2) lack of analysis of why ELMs can be potentially used to tackle the imbalanced data problem, (3) lack of taxonomy of ELMs amongst themselves using different machine learning techniques and lack of their developments based on standardized quality metrics, and (4) lack of pointing out suitable datasets to facilitate future studies in the outlier detection field.

To the best of our knowledge, this survey provides a comprehensive analysis of ELM and its applications for addressing the above-mentioned gaps. In this matter, twenty studies published between 2014 and 2021 are summarized and discussed, exploring several advanced techniques for learning from imbalanced data with ELMs. It has been demonstrated that

ELMs for handling imbalanced data can be successfully extended although the adaptation of distribution weights and biases remains a significant challenge.

The survey also explores twenty papers published between 2013 and 2021 to address state-of-the-art ELMs which have been applied for outlier detection on different issues. Moreover, it classifies these methods under three different machine learning perspectives (i.e., supervised, unsupervised, and semi-supervised approaches). AUC, ROC, and RMSE are among the foremost well-known measurements for assessing the performance of the proposed ELMs.

Continuously, the survey covers studies that critically use datasets, with a focus on ELM as the limiting factor for outlier detection. ELMs have been employed in considerably different areas of outlier detection, including intrusion detection systems, health monitoring in space, wind speed forecasting stock markets, etc.

To conclude, this study can be followed by researchers who tend to develop ELMs for outlier detection since to the best of our knowledge, this is the first work that has focused on ELMs in this field. Alternatively, we assert that several challenges remain open and may be worth absorbing the attention of researchers.

In future work, outlier injection methods will be reviewed. Then a flexible outlier injection process will be proposed to inject various types of anomalies and will tackle the class imbalanced data issue regarding outlier detection using ELM. This may result in a high-quality division of data into training, evaluation, and testing data, thus improving accuracy. Moreover, we presented some contemporary open challenges to cover significant gaps when ELMs are applied for outlier detection. In this context, we will focus on what strategies can be taken to improve the generalization and reliability of the ELM for outlier detection.

Acknowledgements Many thanks for the reviewers' very constructive remarks. We hope the revised manuscript has addressed satisfactorily your concerns.

Authors' contributions Rasoul Kiani and Dr. Victor S. Sheng had the idea for the article. Rasoul Kiani performed the literature search and data analysis. Rasoul Kiani, Dr. Victor S. Sheng, and Dr. Wei Jin drafted and critically revised the work.

Funding No funding was received for conducting this study.

Data Availability Not applicable.

Code Availability Not applicable.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval Not applicable.

Consent to participate Not applicable.

Consent for publication Not applicable.

References

Abdelghafar, S., Darwish, A., Hassanien, A. E., Yahia, M., & Zaghrout, A. (2019). Anomaly detection of satellite telemetry based on optimized extreme learning machine. *Journal of Space Safety Engineering*, *6*(4), 291–298.

Adeli, E., Thung, K. H., An, L., Wu, G., Shi, F., Wang, T., & Shen, D. (2018). Semi-supervised discriminative classification robust to sample-outliers and feature-noises. *IEEE transactions on pattern analysis and machine intelligence*, *41*(2), 515–522.

Aggarwal, C. C. (2017). An introduction to outlier analysis. *Outlier analysis* (pp. 1–34). Springer.

Aggarwal, C. C., & Yu, P. S. (2001, May). Outlier detection for high dimensional data. In *Proceedings of the 2001 ACM SIGMOD international conference on Management of data* (pp. 37–46).

Aggarwal, C. C., & Yu, P. S. (2008, April). Outlier detection with uncertain data. In *Proceedings of the 2008 SIAM International Conference on Data Mining* (pp. 483–493). Society for Industrial and Applied Mathematics.

Aggarwal, C. C., Zhao, Y., & Philip, S. Y. (2011, April). Outlier detection in graph streams. In *2011 IEEE 27th international conference on data engineering* (pp. 399–409). IEEE.

Albuquerque Filho, J. E., Brandão, L. C., Fernandes, B. J., & Maciel, A. M. (2022). A review of neural networks for Anomaly Detection. *IEEE Access: Practical Innovations, Open Solutions*, *10*, 112342–112367.

Alcin, O. F., Ucar, F., & Korkmaz, D. (2016, August). Extreme learning machine based robotic arm modeling. In *2016 21st International Conference on Methods and Models in Automation and Robotics (MMAR)* (pp. 1160–1163). IEEE.

Altunay, H. C., Albayrak, Z., Özalp, A. N., & Çakmak, M. (2021, June). Analysis of anomaly detection approaches performed through deep learning methods in SCADA systems. In *2021 3rd International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)* (pp. 1–6). IEEE.

Anjaneyulu, M., & Kishore, A. P. A. U. (2019). Financial Fraud detection with Anomaly Feature detection on credit card. *International Journal of Scientific Research & Engineering*, *5*(3), 835–837.

Ariafar, E., & Kiani, R. (2017, December). Intrusion detection system using an optimized framework based on data mining techniques. In *2017 IEEE 4th International Conference on Knowledge-Based Engineering and Innovation (KBEI)* (pp. 0785–0791). IEEE.

Balasundaram, S. (2013). On extreme learning machine for ε-insensitive regression in the primal by Newton method. *Neural Computing and Applications*, *22*(3), 559–567.

Barreto, G. A., & Barros, A. L. B. (2016). A robust extreme learning machine for pattern classification with outliers. *Neurocomputing*, *176*, 3–13.

Bauder, R. A., & Khoshgoftaar, T. M. (2018). The effects of varying class distribution on learner behavior for medicare fraud detection with imbalanced big data. *Health information science and systems*, *6*(1), 1–14.

Bawono, A. H., & Bachtiar, F. A. (2019, September). Outlier Detection with Supervised Learning Method. In *2019 International Conference on Sustainable Information Engineering and Technology (SIET)* (pp. 306–309). IEEE.

Cai, Z., Gu, J., Luo, J., Zhang, Q., Chen, H., Pan, Z., & Li, C. (2019). Evolving an optimal kernel extreme learning machine by using an enhanced grey wolf optimization strategy. *Expert Systems with Applications*, *138*, 112814.

Cao, L., Zhu, P., Zhao, Y., & Zhao, J. (2018). Using machine learning and quantum chemistry descriptors to predict the toxicity of ionic liquids. *Journal of hazardous materials*, *352*, 17–26.

Chacko, B. P., Krishnan, V., Raju, V. R., G., & Anto, B., P (2012). Handwritten character recognition using wavelet energy and extreme learning machine. *International Journal of Machine Learning and Cybernetics*, *3*(2), 149–161.

Chakraborty, B., Chaterjee, A., Malakar, S., & Sarkar, R. (2022). An iterative approach to unsupervised outlier detection using ensemble method and distance-based data filtering. *Complex & Intelligent Systems*, 1–16.

Chen, F. L., & Ou, T. Y. (2011). Sales forecasting system based on Gray extreme learning machine with Taguchi method in retail industry. *Expert Systems with Applications*, *38*(3), 1336–1345.

Chen, L., Zhang, Y. H., Huang, G., Pan, X., Wang, S., Huang, T., & Cai, Y. D. (2018). Discriminating circRNAs from other lncRNAs using a hierarchical extreme learning machine (H-ELM) algorithm with feature selection. *Molecular genetics and genomics*, *293*(1), 137–149.

Chen, C., Jin, X., Jiang, B., & Li, L. (2019). Optimizing extreme learning machine via generalized hebbian learning and intrinsic plasticity learning. *Neural Processing Letters*, *49*(3), 1593–1609.

Chen, S., Wu, M., Wen, P., Xu, F., Wang, S., & Zhao, S. (2021). A Multimode Anomaly Detection Method based on OC-ELM for Aircraft Engine System. *IEEE Access: Practical Innovations, Open Solutions*, *9*, 28842–28855.

Cheng, K., Gao, S., Dong, W., Yang, X., Wang, Q., & Yu, H. (2020). Boosting label weighted extreme learning machine for classifying multi-label imbalanced data. *Neurocomputing*, *403*, 360–370.

Cogoljević, D., Alizamir, M., Piljan, I., Piljan, T., Prljić, K., & Zimonjić, S. (2018). A machine learning approach for predicting the relationship between energy resources and economic development. *Physica A: Statistical Mechanics and its Applications*, *495*, 211–214.

da Costa, N. L., Llobodanin, L. A. G., de Lima, M. D., Castro, I. A., & Barbosa, R. (2018). Geographical recognition of Syrah wines by combining feature selection with Extreme Learning Machine. *Measurement*, *120*, 92–99.

Daliri, M. R. (2012). A hybrid automatic system for the diagnosis of lung cancer based on genetic algorithm and fuzzy extreme learning machines. *Journal of medical systems*, *36*(2), 1001–1005.

Dashdondov, K., & Kim, M. H. (2021). Mahalanobis Distance based Multivariate Outlier detection to improve performance of Hypertension Prediction. *Neural Processing Letters*, 1–13.

Dekker, R. (2006). The importance of having data-sets.

Devika, K., Mahapatra, D., Subramanian, R., & Oruganti, V. R. M. (2022). Outlier-based Autism detection using Longitudinal Structural MRI. *IEEE Access: Practical Innovations, Open Solutions*, *10*, 27794–27808.

Ding, S., Xu, X., & Nie, R. (2014). Extreme learning machine and its applications. *Neural Computing and Applications*, *25*(3), 549–556.

Duan, J., Ou, Y., Hu, J., Wang, Z., Jin, S., & Xu, C. (2017). Fast and stable learning of dynamical systems based on extreme learning machine. *IEEE Transactions on Systems Man and Cybernetics: Systems*, *49*(6), 1175–1185.

Eshtay, M., Faris, H., & Obeid, N. (2020). A competitive swarm optimizer with hybrid encoding for simultaneously optimizing the weights and structure of Extreme Learning Machines for classification problems. *International Journal of Machine Learning and Cybernetics*, *11*(8), 1801–1823.

Feng, G., Huang, G. B., Lin, Q., & Gay, R. (2009). Error minimized extreme learning machine with growth of hidden nodes and incremental learning. *IEEE Transactions on Neural Networks*, *20*(8), 1352–1357.

Feng, G., Qian, Z., & Zhang, X. (2012). Evolutionary selection extreme learning machine optimization for regression. *Soft Computing*, *16*(9), 1485–1491.

Fernández, Á., Bella, J., & Dorronsoro, J. R. (2022). Supervised outlier detection for classification and regression. *Neurocomputing*, *486*, 77–92.

Frénay, B., & Verleysen, M. (2015). Reinforced extreme learning machines for fast robust regression in the presence of outliers. *IEEE Transactions on Cybernetics*, *46*(12), 3351–3363.

Geng, Z., Dong, J., Chen, J., & Han, Y. (2017a). A new self-organizing extreme learning machine soft sensor model and its applications in complicated chemical processes. Engineering Applications of Artificial Intelligence, 62, 38–50.

Geng, Z., Zhao, S., Tao, G., & Han, Y. (2017b). Early warning modeling and analysis based on analytic hierarchy process integrated extreme learning machine (AHP-ELM): Application to food safety. Food Control, 78, 33–42.

Gumaei, A., Hassan, M. M., Hassan, M. R., Alelaiwi, A., & Fortino, G. (2019). A hybrid feature extraction method with regularized extreme learning machine for brain tumor classification. *IEEE Access: Practical Innovations, Open Solutions*, *7*, 36266–36273.

Guo, M., Ma, Y., Yang, X., & Mankin, R. W. (2019). Detection of damaged wheat kernels using an impact acoustic signal processing technique based on gaussian modelling and an improved extreme learning machine algorithm. *biosystems engineering*, *184*, 37–44.

Gupta, M., Gao, J., Aggarwal, C. C., & Han, J. (2013). Outlier detection for temporal data: A survey. *IEEE Transactions on Knowledge and data Engineering*, *26*(9), 2250–2267.

Hashmi, A. S., & Ahmad, T. (2019). GP-ELM-RNN: Garson-pruned extreme learning machine based replicator neural network for anomaly detection. *Journal of King Saud University-Computer and Information Sciences*, *34*, 1768–1774.

He, Q., Shang, T., Zhuang, F., & Shi, Z. (2013). Parallel extreme learning machine for regression based on MapReduce. *Neurocomputing*, *102*, 52–58.

Heeswijk, M. V., Miche, Y., Lindh-Knuutila, T., Hilbers, P. A., Honkela, T., Oja, E., & Lendasse, A. (2009, September). Adaptive ensemble models of extreme learning machines for time series prediction. In *International Conference on Artificial Neural Networks* (pp. 305–314). Springer, Berlin, Heidelberg.

Horata, P., Chiewchanwattana, S., & Sunat, K. (2013). Robust extreme learning machine. *Neurocomputing*, *102*, 31–44.

Hu, X. F., Zhao, Z., Wang, S., Wang, F. L., He, D. K., & Wu, S. K. (2008). Multi-stage extreme learning machine for fault diagnosis on hydraulic tube tester. *Neural Computing and Applications*, *17*(4), 399–403.

Hu, K., Yang, W., & Gao, X. (2017). Microcalcification diagnosis in digital mammography using extreme learning machine based on hidden Markov tree model of dual-tree complex wavelet transform. *Expert Systems with Applications*, *86*, 135–144.

Huang, G. B., Chen, L., & Siew, C. K. (2006a). Universal approximation using incremental constructive feedforward networks with random hidden nodes. Ieee Transaction On Neural Networks, 17(4), 879–892.

Huang, G. B., Zhu, Q. Y., & Siew, C. K. (2006b). Extreme learning machine: Theory and applications. Neurocomputing, 70(1–3), 489–501.

Huang, G. B., Li, M. B., Chen, L., & Siew, C. K. (2008). Incremental extreme learning machine with fully complex hidden nodes. *Neurocomputing*, *71*(4–6), 576–583.

Huang, G. B., Wang, D. H., & Lan, Y. (2011a). Extreme learning machines: A survey. International journal of machine learning and cybernetics, 2(2), 107–122.

Huang, G. B., Zhou, H., Ding, X., & Zhang, R. (2011b). Extreme learning machine for regression and multiclass classification. IEEE Transactions on Systems Man and Cybernetics Part B (Cybernetics), 42(2), 513–529.

Huang, F., Yin, K., Huang, J., Gui, L., & Wang, P. (2017). Landslide susceptibility mapping based on self-organizing-map network and extreme learning machine. *Engineering Geology*, *223*, 11–22.

Huang, W., Yang, Y., Lin, Z., Huang, G. B., Zhou, J., Duan, Y., & Xiong, W. (2014, August). Random feature subspace ensemble based extreme learning machine for liver tumor detection and segmentation. In *2014 36th annual international conference of the IEEE engineering in medicine and biology society* (pp. 4675–4678). IEEE.

Huang, W., Tan, Z. M., Lin, Z., Huang, G. B., Zhou, J., Chui, C. K., & Chang, S. (2012, August). A semi-automatic approach to the segmentation of liver parenchyma from 3D CT images with extreme learning machine. In *2012 annual international conference of the IEEE engineering in medicine and biology society* (pp. 3752–3755). IEEE.

Janakiraman, V. M., & Nielsen, D. (2016, July). Anomaly detection in aviation data using extreme learning machines. In *2016 international joint conference on neural networks (IJCNN)* (pp. 1993–2000). IEEE.

Jiang, C., Song, J., Liu, G., Zheng, L., & Luan, W. (2018). Credit card fraud detection: A novel approach using aggregation strategy and feedback mechanism. *IEEE Internet of Things Journal*, *5*(5), 3637–3647.

Johnson, J. M., & Khoshgoftaar, T. M. (2019). Survey on deep learning with class imbalance. *Journal of Big Data*, *6*(1), 1–54.

Kalid, S. N., Ng, K. H., Tong, G. K., & Khor, K. C. (2020). A multiple classifiers system for anomaly detection in credit card data with unbalanced and overlapped classes. *IEEE Access: Practical Innovations, Open Solutions*, *8*, 28210–28221.

Kang, X., Liu, C., Zeng, S., Zhao, Z., Qian, J., & Zhao, Y. (2018). Prediction of Henry's law constant of CO2 in ionic liquids based on SEP and Sσ-profile molecular descriptors. *Journal of Molecular Liquids*, *262*, 139–147.

Karpagachelvi, S., Arthanari, M., & Sivakumar, M. (2012). Classification of electrocardiogram signals with support vector machines and extreme learning machine. *Neural Computing and Applications*, *21*(6), 1331–1339.

Kiani, R., Keshavarzi, A., & Bohlouli, M. (2020). Detection of thin boundaries between different types of anomalies in outlier detection using enhanced neural networks. *Applied Artificial Intelligence*, *34*(5), 345–377.

Kiani, R., Montazeri, M., & Minaei-Bidgoli, B. Intelligent Production and Detection Template of Outlier Dataset Using Clustering. In *2019 5th Conference on Knowledge Based Engineering and Innovation (KBEI)* (pp. 145–152). IEEE.

Kim, J., Shin, H. S., Shin, K., & Lee, M. (2009). Robust algorithm for arrhythmia classification in ECG using extreme learning machine. *Biomedical engineering online*, *8*(1), 1–12.

Krawczyk, B. (2016). Learning from imbalanced data: Open challenges and future directions. *Progress in Artificial Intelligence*, *5*(4), 221–232.

Lama, R. K., Gwak, J., Park, J. S., & Lee, S. W. (2017). Diagnosis of Alzheimer's disease based on structural MRI images using a regularized extreme learning machine and PCA features. *Journal of healthcare engineering*, *2017*, 5485080.

Lan, Y., Soh, Y. C., & Huang, G. B. (2009). Ensemble of online sequential extreme learning machine. *Neurocomputing*, *72*(13–15), 3391–3395.

Lan, Y., Soh, Y. C., & Huang, G. B. (2010a). Constructive hidden nodes selection of extreme learning machine for regression. Neurocomputing, 73(16–18), 3191–3199.

Lan, Y., Soh, Y. C., & Huang, G. B. (2010b). Two-stage extreme learning machine for regression. Neurocomputing, 73(16–18), 3028–3038.

Lan, Y., Hu, Z., Soh, Y. C., & Huang, G. B. (2013). An extreme learning machine approach for speaker recognition. *Neural Computing and Applications*, *22*(3), 417–425.

Landa-Torres, I., Ortiz-Garcia, E. G., Salcedo-Sanz, S., Segovia-Vargas, M. J., Gil-Lopez, S., Miranda, M., & Del Ser, J. (2012). Evaluating the internationalization success of companies through a hybrid grouping harmony search—extreme learning machine approach. *IEEE Journal of Selected Topics in Signal Processing*, *6*(4), 388–398.

Lee, Y., Lee, H., Kim, J., Shin, H. C., & Lee, M. (2009). Classification of BMI control commands from rat's neural signals using extreme learning machine. *BioMedical Engineering OnLine*, *8*(1), 1–10.

Lee, K., Lee, C. H., & Lee, J. (2021). Semi-supervised Anomaly Detection Algorithm using probabilistic labeling (SAD-PL). *IEEE Access: Practical Innovations, Open Solutions*, *9*, 142972–142981.

Leevy, J. L., Khoshgoftaar, T. M., Bauder, R. A., & Seliya, N. (2018). A survey on addressing high-class imbalance in big data. *Journal of Big Data*, *5*(1), 1–30.

Lei, H., Wen, Y., You, Z., Elazab, A., Tan, E. L., Zhao, Y., & Lei, B. (2018). Protein–protein interactions prediction via multimodal deep polynomial network and regularized extreme learning machine. *IEEE journal of biomedical and health informatics*, *23*(3), 1290–1303.

Li, G., & Niu, P. (2013). An enhanced extreme learning machine based on ridge regression for regression. *Neural Computing and Applications*, *22*(3), 803–810.

Li, K., Kong, X., Lu, Z., Wenyin, L., & Yin, J. (2014). Boosting weighted ELM for imbalanced learning. *Neurocomputing*, *128*, 15–21.

Li, W., Chen, C., Su, H., & Du, Q. (2015). Local binary patterns and extreme learning machine for hyperspectral imagery classification. *IEEE Transactions on Geoscience and Remote Sensing*, *53*(7), 3681–3693.

Li, J., Shi, X., You, Z. H., Yi, H. C., Chen, Z., Lin, Q., & Fang, M. (2020a). Using weighted extreme learning machine combined with scale-invariant feature transform to predict protein-protein interactions from protein evolutionary information. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 17(5), 1546–1554.

Li, Q., Song, Y., Zhang, J., & Sheng, V. S. (2020b). Multiclass imbalanced learning with one-versus-one decomposition and spectral clustering. Expert Systems with Applications, 147, 113152.

Li, Z., Wei, L., Li, W., Wei, L., Chen, M., Lv, M., & Gao, N. (2019, October). Research on DDoS attack detection based on ELM in IoT environment. In *2019 IEEE 10th International Conference on Software Engineering and Service Science (ICSESS)* (pp. 144–148). IEEE.

Liang, N. Y., Huang, G. B., Saratchandran, P., & Sundararajan, N. (2006). A fast and accurate online sequential learning algorithm for feedforward networks. *IEEE Transactions on neural networks*, *17*(6), 1411–1423.

Ling, C. X., & Sheng, V. S. (2008). Cost-sensitive learning and the class imbalance problem. *Encyclopedia of machine learning*, *2011*, 231–235.

Liu, T., Yang, Y., Huang, G. B., Yeo, Y. K., & Lin, Z. (2015). Driver distraction detection using semi-supervised machine learning. *IEEE transactions on intelligent transportation systems*, *17*(4), 1108–1120.

Liu, J., Li, Q., Chen, W., Yan, Y., & Wang, X. (2018). A fast fault diagnosis method of the PEMFC system based on extreme learning machine and Dempster–Shafer evidence theory. *IEEE Transactions on Transportation Electrification*, *5*(1), 271–284.

Liu, Y., Li, X., Wang, J., Chen, F., Wang, J., Shi, Y., & Zheng, L. (2017, October). Pattern Discovery from Big Data of Food Sampling Inspections Based on Extreme Learning Machine. In *International Conference on Research and Practical Issues of Enterprise Information Systems* (pp. 132–142). Springer, Cham.

Liu, B., Yan, S., You, H., Dong, Y., Li, J., Li, Y., & Gu, R. (2017, December). An ensembled RBF extreme learning machine to forecast road surface temperature. In *2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA)* (pp. 977–980). IEEE.

Lu, X., Ming, L., Liu, W., & Li, H. X. (2017). Probabilistic regularized extreme learning machine for robust modeling of noise data. *IEEE Transactions on Cybernetics*, *48*(8), 2368–2377.

Lu, S., Lu, Z., Yang, J., Yang, M., & Wang, S. (2018). A pathological brain detection system based on kernel based ELM. *Multimedia tools and applications*, *77*(3), 3715–3728.

Malar, E., Kandaswamy, A., Chakravarthy, D., & Dharan, A. G. (2012). A novel approach for detection and classification of mammographic microcalcifications using wavelet analysis and extreme learning machine. *Computers in biology and medicine*, *42*(9), 898–905.

Mao, W., Wang, J., & Xue, Z. (2017). An ELM-based model with sparse-weighting strategy for sequential data imbalance problem. *International Journal of Machine Learning and Cybernetics*, *8*(4), 1333–1345.

Marjanović, V., Milovančević, M., & Mladenović, I. (2016). Prediction of GDP growth rate based on carbon dioxide (CO2) emissions. *Journal of CO2 Utilization*, *16*, 212–217.

Marković, D., Petković, D., Nikolić, V., Milovančević, M., & Petković, B. (2017). Soft computing prediction of economic growth based in science and technology factors. *Physica A: Statistical Mechanics and its Applications*, *465*, 217–220.

Mi, X. W., Liu, H., & Li, Y. F. (2017). Wind speed forecasting method using wavelet, extreme learning machine and outlier correction algorithm. *Energy Conversion and Management*, *151*, 709–722.

Milačić, L., Jović, S., Vujović, T., & Miljković, J. (2017). Application of artificial neural network with extreme learning machine for economic growth estimation. *Physica A: Statistical Mechanics and its Applications*, *465*, 285–288.

Minhas, R., Baradarani, A., Seifzadeh, S., & Wu, Q. J. (2010). Human action recognition using extreme learning machine based on visual vocabularies. *Neurocomputing*, *73*(10–12), 1906–1917.

Mohammed, A. A., Minhas, R., Wu, Q. J., & Sid-Ahmed, M. A. (2011). Human face recognition based on multidimensional PCA and extreme learning machine. *Pattern recognition*, *44*(10–11), 2588–2597.

Muñoz-Ramírez, V., Kmetzsch, V., Forbes, F., Meoni, S., Moro, E., & Dojat, M. (2022). Subtle anomaly detection: Application to brain MRI analysis of de novo parkinsonian patients. *Artificial Intelligence in Medicine*, *125*, 102251.

Naseer, S., Saleem, Y., Khalid, S., Bashir, M. K., Han, J., Iqbal, M. M., & Han, K. (2018). Enhanced network anomaly detection based on deep neural networks. *IEEE access*, *6*, 48231–48246.

Nayak, D. R., Das, D., Dash, R., Majhi, S., & Majhi, B. (2020). Deep extreme learning machine with leaky rectified linear unit for multiclass classification of pathological brain images. *Multimedia Tools and Applications*, *79*(21), 15381–15396.

Nguyen, D. T., Ryu, S., Qureshi, M. N. I., Choi, M., Lee, K. H., & Lee, B. (2019). Hybrid multivariate pattern analysis combined with extreme learning machine for Alzheimer's dementia diagnosis using multi-measure rs-fMRI spatial patterns. *PloS one*, *14*(2), e0212582.

Nian, R., He, B., & Lendasse, A. (2013). 3D object recognition based on a geometrical topology model and extreme learning machine. *Neural Computing and Applications*, *22*(3), 427–433.

Niu, W. J., Feng, Z. K., Feng, B. F., Min, Y. W., Cheng, C. T., & Zhou, J. Z. (2019). Comparison of multiple linear regression, artificial neural network, extreme learning machine, and support vector machine in deriving operation rule of hydropower reservoir. *Water*, *11*(1), 88.

Niu, M., Zhang, J., Li, Y., Wang, C., Liu, Z., Ding, H., & Ma, Q. (2020). CirRNAPL: A web server for the identification of circRNA based on extreme learning machine. *Computational and structural biotechnology journal*, *18*, 834–842.

Oikawa, H., Nishida, T., Sakamoto, R., Matsutani, H., & Kondo, M. (2020, September). Fast semi-supervised anomaly detection of drivers' behavior using online sequential extreme learning machine. In *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)* (pp. 1–8). IEEE.

Oneto, L., Fumeo, E., Clerico, G., Canepa, R., Papa, F., Dambra, C., & Anguita, D. (2017). Dynamic delay predictions for large-scale railway networks: Deep and shallow extreme learning machines tuned via thresholdout. *IEEE Transactions on Systems Man and Cybernetics: Systems*, *47*(10), 2754–2767.

Ouyang, Q., Chen, Q., Zhao, J., & Lin, H. (2013). Determination of amino acid nitrogen in soy sauce using near infrared spectroscopy combined with characteristic variables selection and extreme learning machine. *Food and Bioprocess Technology*, *6*(9), 2486–2493.

Ouyang, T., Wang, C., Yu, Z., Stach, R., Mizaikoff, B., Huang, G. B., & Wang, Q. J. (2020). NOx measurements in vehicle exhaust using advanced deep ELM networks. *IEEE Transactions on Instrumentation and Measurement*, *70*, 1–10.

Pan, C., Park, D. S., Lu, H., & Wu, X. (2012a). Color image segmentation by fixation-based active learning with ELM. Soft Computing, 16(9), 1569–1584.

Pan, C., Park, D. S., Yang, Y., & Yoo, H. M. (2012b). Leukocyte image segmentation by visual attention and extreme learning machine. Neural Computing and Applications, 21(6), 1217–1227.

Paullada, A., Raji, I. D., Bender, E. M., Denton, E., & Hanna, A. (2021). Data and its (dis) contents: A survey of dataset development and use in machine learning research. *Patterns*, *2*(11), 100336.

Peng, X., Lin, P., Zhang, T., & Wang, J. (2013). Extreme learning machine-based classification of ADHD using brain structural MRI data. *PloS one*, 8(11), e79476.

Qin, Y., Li, M., De, G., Huang, L., Yang, S., Tan, Q., & Zhou, F. (2019). Research on green management effect evaluation of power generation enterprises in China based on dynamic hesitation and improved extreme learning machine. *Processes*, *7*(7), 474.

Qureshi, M. N. I., Min, B., Jo, H. J., & Lee, B. (2016). Multiclass classification for the differential diagnosis on the ADHD subtypes using recursive feature elimination and hierarchical extreme learning machine: Structural MRI study. *PloS one*, 11(8), e0160697.

Qureshi, M. N. I., Oh, J., Cho, D., Jo, H. J., & Lee, B. (2017). Multimodal discrimination of schizophrenia using hybrid weighted feature concatenation of brain functional connectivity and anatomical features with an extreme learning machine. *Frontiers in neuroinformatics*, *11*, 59.

Raghuwanshi, B. S., & Shukla, S. (2021). Classifying imbalanced data using SMOTE based class-specific kernelized ELM. *International Journal of Machine Learning and Cybernetics*, *12*(5), 1255–1280.

Rakic, G., Milenkovic, D., Vujovic, S., Vujovic, T., & Jović, S. (2019). Information system for e-GDP based on computational intelligence approach. *Physica A: Statistical Mechanics and its Applications*, *513*, 418–423.

Ramalho, G. L. B., Rebouças Filho, P. P., Medeiros, F. N. S. D., & Cortez, P. C. (2014). Lung disease detection using feature extraction and extreme learning machine. *Revista Brasileira de Engenharia Biomédica*, *30*(3), 207–214.

Rathore, S., & Park, J. H. (2018). Semi-supervised learning based distributed attack detection framework for IoT. *Applied Soft Computing*, *72*, 79–89.

Ri, J. H., Tian, G., Liu, Y., Xu, W. H., & Lou, J. G. (2020). Extreme learning machine with hybrid cost function of G-mean and probability for imbalance learning. *International Journal of Machine Learning and Cybernetics*, *11*(9), 2007–2020.

Rong, H. J., Ong, Y. S., Tan, A. H., & Zhu, Z. (2008). A fast pruned-extreme learning machine for classification problem. *Neurocomputing*, *72*(1–3), 359–366.

Sahu, S. K., Mohapatra, D. P., & Ray, N. K. (2021, December). An Ensemble-based Outlier Detection Approach on Intrusion Detection. In *2021 19th OITS International Conference on Information Technology (OCIT)* (pp. 404–409). IEEE.

Sánchez-Oro, J., Duarte, A., & Salcedo-Sanz, S. (2016). Robust total energy demand estimation with a hybrid Variable Neighborhood Search–Extreme Learning Machine algorithm. *Energy Conversion and Management*, *123*, 445–452.

Sekar, M. (2022). Fraud and Anomaly Detection. In *Machine Learning for Auditors* (pp. 193–202). Apress, Berkeley, CA.

She, Q., Hu, B., Gan, H., Fan, Y., Nguyen, T., Potter, T., & Zhang, Y. (2018). Safe semi-supervised extreme learning machine for EEG signal classification. *IEEE Access: Practical Innovations, Open Solutions*, *6*, 49399–49407.

Shoumo, S. Z. H., Dhruba, M. I. M., Hossain, S., Ghani, N. H., Arif, H., & Islam, S. (2019, October). Application of machine learning in credit risk assessment: a prelude to smart banking. In *TENCON 2019–2019 IEEE Region 10 Conference (TENCON)* (pp. 2023–2028). IEEE.

Shukla, S., & Yadav, R. N. (2015). Regularized weighted circular complex-valued extreme learning machine for imbalanced learning. *IEEE Access: Practical Innovations, Open Solutions*, *3*, 3048–3057.

Shukla, A. K., Kumar, S., Jagdev, R., Muhuri, P. K., & Lohani, Q. D. (2018, July). Interval Type-2 Fuzzy weighted Extreme Learning Machine for GDP Prediction. In *2018 International Joint Conference on Neural Networks (IJCNN)* (pp. 1–8). IEEE.

Siqi, W., Qiang, L., Xifeng, G., En, Z., & Jianping, Y. (2019). Fast and unsupervised outlier removal by recurrent adaptive reconstruction extreme learning machine. *International Journal of Machine Learning and Cybernetics*, *10*(12), 3539–3556.

Sokolov-Mladenović, S., Milovančević, M., Mladenović, I., & Alizamir, M. (2016). Economic growth forecasting by artificial neural network with extreme learning machine based on trade, import and export parameters. *Computers in Human Behavior*, *65*, 43–45.

Song, J. L., Hu, W., & Zhang, R. (2016). Automated detection of epileptic EEGs using a novel fusion feature and extreme learning machine. *Neurocomputing*, *175*, 383–391.

Song, T., Li, D., Liu, Z., & Yang, W. (2019). Online ADMM-based extreme learning machine for sparse supervised learning. *IEEE Access: Practical Innovations, Open Solutions*, *7*, 64533–64544.

Sridhar, S., & Sanagavarapu, S. (2021, September). ELM-AD: Extreme Learning Machine Framework for Price and Volume Anomaly Detection in Stock Markets. In *2021 International Conference on Computing and Communications Applications and Technologies (I3CAT)* (pp. 44–51). IEEE.

Su, J., Shen, H., Peng, L., & Hu, D. (2021). Few-shot domain-adaptive anomaly detection for cross-site brain images. *IEEE Transactions on Pattern Analysis and Machine Intelligence, doi: 10.1109/TPAMI.2021.3125686*

Sun, W., & Zhang, C. (2018). Analysis and forecasting of the carbon price using multi—resolution singular value decomposition and extreme learning machine optimized by adaptive whale optimization algorithm. *Applied energy*, *231*, 1354–1371.

Sun, Z. L., Choi, T. M., Au, K. F., & Yu, Y. (2008). Sales forecasting using extreme learning machine with applications in fashion retailing. *Decision Support Systems*, *46*(1), 411–419.

Sun, Z. L., Ng, K. M., Soszynska-Budny, J., & Habibullah, M. S. (2011). Application of the LP-ELM model on transportation system lifetime optimization. *IEEE Transactions on intelligent transportation systems*, *12*(4), 1484–1494.

Sun, W., Wang, C., & Zhang, C. (2017). Factor analysis and forecasting of CO2 emissions in Hebei, using extreme learning machine based on particle swarm optimization. *Journal of cleaner production*, *162*, 1095–1101.

Sun, C., Yu, Y., Liu, H., & Gu, J. (2015, December). Robotic grasp detection using extreme learning machine. In *2015 IEEE International Conference on Robotics and Biomimetics (ROBIO)* (pp. 1115–1120). IEEE.

Tama, B. A., Comuzzi, M., & Rhee, K. H. (2019). TSE-IDS: A two-stage classifier ensemble for intelligent anomaly-based intrusion detection system. *IEEE Access: Practical Innovations, Open Solutions*, *7*, 94497–94507.

Termenon, M., Graña, M., Barrós-Loscertales, A., & Ávila, C. (2013). Extreme learning machines for feature selection and classification of cocaine dependent patients on structural MRI data. *Neural processing letters*, *38*(3), 375–387.

Termenon, M., Graña, M., 2, A., Akusok, A., Miche, Y., Björk, K. M., & Lendasse, A. (2016). Brain MRI morphological patterns extraction tool based on Extreme Learning Machine and majority vote classification. *Neurocomputing*, *174*, 344–351.

Thammasakorn, C., Chiewchanwattana, S., & Sunat, K. (2018, July). Optimizing weighted elm based on gray wolf optimizer for imbalanced data classification. In *2018 10th International Conference on Information Technology and Electrical Engineering (ICITEE)* (pp. 512–517). IEEE.

Van Heeswijk, M., Miche, Y., Oja, E., & Lendasse, A. (2011). GPU-accelerated and parallelized ELM ensembles for large-scale regression. *Neurocomputing*, *74*(16), 2430–2437.

Vani, G., Savitha, R., & Sundararajan, N. (2010, December). Classification of abnormalities in digitized mammograms using extreme learning machine. In *2010 11th International Conference on Control Automation Robotics & Vision* (pp. 2114–2117). IEEE.

Wang, B., Wang, G., Li, J., & Wang, B. (2012). Update strategy based on region classification using ELM for mobile object index. *Soft Computing*, *16*(9), 1607–1615.

Wang, J. N., Jin, J. L., Geng, Y., Sun, S. L., Xu, H. L., Lu, Y. H., & Su, Z. M. (2013). An accurate and efficient method to predict the electronic excitation energies of BODIPY fluorescent dyes. *Journal of computational chemistry*, *34*(7), 566–575.

Wang, Z., Yu, G., Kang, Y., Zhao, Y., & Qu, Q. (2014). Breast tumor detection in digital mammography based on extreme learning machine. *Neurocomputing*, *128*, 175–184.

Wang, X., Shen, D., Bai, M., Nie, T., Kou, Y., & Yu, G. (2016a). Cluster-based outlier detection using unsupervised extreme learning machines. In Proceedings of ELM-2015 Volume 1 (pp. 135–146). Springer, Cham.

Wang, Z., Qu, Q., Yu, G., & Kang, Y. (2016b). Breast tumor detection in double views mammography based on extreme learning machine. Neural Computing and Applications, 27(1), 227–240.

Wang, R., Chow, C. Y., Lyu, Y., Lee, V. C., Kwong, S., Li, Y., & Zeng, J. (2017a). TaxiRec: Recommending road clusters to taxi drivers using ranking-based extreme learning machines. IEEE Transactions on Knowledge and Data Engineering, 30(3), 585–598.

Wang, X., Bai, M., Shen, D., Nie, T., Kou, Y., & Yu, G. (2017b). A distributed algorithm for the cluster-based outlier detection using unsupervised extreme learning machines. Mathematical Problems in Engineering, 2, 1–12.

Wang, Z., Xin, J., Yang, H., Tian, S., Yu, G., Xu, C., & Yao, Y. (2017c). Distributed and weighted extreme learning machine for imbalanced big data learning. Tsinghua Science and Technology, 22(2), 160–173.

Wang, H., Bah, M. J., & Hammad, M. (2019a). Progress in outlier detection techniques: A survey. Ieee Access, 7, 107964–108000.

Wang, Z., Li, M., Wang, H., Jiang, H., Yao, Y., Zhang, H., & Xin, J. (2019b). Breast cancer detection using extreme learning machine based on feature fusion with CNN deep features. IEEE Access: Practical Innovations, Open Solutions, 7, 105146–105158.

Wang, G., Wong, K. W., & Lu, J. (2020a). AUC-based extreme learning machines for supervised and semi-supervised imbalanced classification. IEEE Transactions on Systems Man and Cybernetics: Systems, 51(12), 7919–7930.

Wang, Y., Wang, A. N., Ai, Q., & Sun, H. J. (2020b). Enhanced Kernel-Based multilayer fuzzy Weighted Extreme Learning Machines. IEEE Access: Practical Innovations, Open Solutions, 8, 166246–166260.

Wang, J., Lu, S., Wang, S. H., & Zhang, Y. D. (2021). A review on extreme learning machine. *Multimedia Tools and Applications*, 1–50.

Wei, X., Liu, H., She, X., Lu, Y., Liu, X., & Mo, S. (2019). Stability Assessment of Rubble Mound Breakwaters using Extreme Learning Machine Models. *Journal of Marine Science and Engineering*, *7*(9), 312.

Xia, J., Chen, H., Li, Q., Zhou, M., Chen, L., Cai, Z., & Zhou, H. (2017). Ultrasound-based differentiation of malignant and benign thyroid nodules: An extreme learning machine approach. *Computer methods and programs in biomedicine*, *147*, 37–49.

Xie, W., Li, Y., & Ma, Y. (2016). Breast mass classification in digital mammography based on extreme learning machine. *Neurocomputing*, *173*, 930–941.

Xing, Y. M., Ban, X. J., & Liu, R. (2018, January). A short-term traffic flow prediction method based on kernel extreme learning machine. In *2018 IEEE International Conference on Big Data and Smart Computing (BigComp)* (pp. 533–536). IEEE.

Xu, Y., Dai, Y., Dong, Z. Y., Zhang, R., & Meng, K. (2013). Extreme learning machine-based predictor for real-time frequency stability assessment of electric power systems. *Neural Computing and Applications*, *22*(3), 501–508.

Xu, Y., Mo, H., Sun, C., & Luo, F. (2019, July). Imbalanced Learning of Weighted Extreme Learning Machines Ensemble Algorithm in Wastewater Treatment Plant Fault Diagnosis. In *2019 Chinese Control Conference (CCC)* (pp. 7528–7533). IEEE.

Yan, W. (2016, July). One-class extreme learning machines for gas turbine combustor anomaly detection. In *2016 international joint conference on neural networks (ijcnn)* (pp. 2909–2914). IEEE.

Yin, Y., Dong, X., & Xu, T. (2018). Rapid and efficient bug assignment using ELM for IOT software. *IEEE Access: Practical Innovations, Open Solutions*, *6*, 52713–52724.

You, Z. H., Zhou, M., Luo, X., & Li, S. (2016). Highly efficient framework for predicting interactions between proteins. *IEEE transactions on cybernetics*, *47*(3), 731–743.

Yu, H., Sun, C., Yang, X., Zheng, S., Wang, Q., & Xi, X. (2018a). LW-ELM: A fast and flexible cost-sensitive learning framework for classifying imbalanced data. IEEE Access: Practical Innovations, Open Solutions, 6, 28488–28500.

Yu, H., Yang, X., Zheng, S., & Sun, C. (2018b). Active learning from imbalanced data: A solution of online weighted extreme learning machine. IEEE transactions on neural networks and learning systems, 30(4), 1088–1103.

Yuan, Q., Zhou, W., Li, S., & Cai, D. (2011). Epileptic EEG classification based on extreme learning machine and nonlinear features. *Epilepsy research*, *96*(1–2), 29–38.

Zeng, Y., Xu, X., Shen, D., Fang, Y., & Xiao, Z. (2016). Traffic sign recognition using kernel extreme learning machines with deep perceptual features. *IEEE Transactions on Intelligent Transportation Systems*, *18*(6), 1647–1653.

Zhai, J., Zhang, S., & Wang, C. (2017). The classification of imbalanced large data sets based on MapReduce and ensemble of ELM classifiers. *International Journal of Machine Learning and Cybernetics*, *8*(3), 1009–1017.

Zhang, K., & Luo, M. (2015). Outlier-robust extreme learning machine for regression problems. *Neurocomputing*, *151*, 1519–1527.

Zhang, X., & Qin, L. (2022). An improved Extreme Learning Machine for Imbalanced Data classification. *IEEE Access: Practical Innovations, Open Solutions*, *10*, 8634–8642.

Zhang, J., Wu, X., & Sheng, V. S. (2014). Imbalanced multiple noisy labeling. *IEEE Transactions on Knowledge and Data Engineering*, *27*(2), 489–503.

Zhang, H., Zhang, S., Yin, Y., & Chen, X. (2018a). Prediction of the hot metal silicon content in blast furnace based on extreme learning machine. International Journal of Machine Learning and Cybernetics, 9(10), 1697–1706.

Zhang, Y. D., Zhao, G., Sun, J., Wu, X., Wang, Z. H., Liu, H. M., & Li, J. (2018b). Smart pathological brain detection by synthetic minority oversampling technique, extreme learning machine, and Jaya algorithm. Multimedia Tools and Applications, 77(17), 22629–22648.

Zhang, D., Peng, X., Pan, K., & Liu, Y. (2019). A novel wind speed forecasting based on hybrid decomposition and online sequential outlier robust extreme learning machine. *Energy conversion and management*, *180*, 338–357.

Zhang, J., Li, Y., Xiao, W., & Zhang, Z. (2020). Robust extreme learning machine for modeling with unknown noise. *Journal of the Franklin Institute*, *357*(14), 9885–9908.

Zhang, X., He, T., Lu, L., Yue, S., Cheng, D., & Xu, X. (2017, November). Video analysis of traffic accidents based on projection extreme learning machine. In *2017 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS)* (pp. 149–154). IEEE.

Zhang, Z., Cai, Y., Gong, W., Liu, X., & Cai, Z. (2020, July). Graph Convolutional Extreme Learning Machine. In *2020 International Joint Conference on Neural Networks (IJCNN)* (pp. 1–8). IEEE.

Zheng, W., Qian, Y., & Lu, H. (2013). Text categorization based on regularization extreme learning machine. *Neural Computing and Applications*, *22*(3), 447–456.

Zhou, Z. H., Zhao, J. W., & Cao, F. L. (2013). Surface reconstruction based on extreme learning machine. *Neural Computing and Applications*, *23*(2), 283–292.

Zhou, C., Huang, S., Xiong, N., Yang, S. H., Li, H., Qin, Y., & Li, X. (2015). Design and analysis of multimodel-based anomaly intrusion detection systems in industrial process automation. *IEEE Transactions on Systems Man and Cybernetics: Systems*, *45*(10), 1345–1360.

Zhou, T., Tang, L., Hang, H., & Yang, P. (2018). Classification of epileptic electroencephalograms signal based on improved extreme learning machine. *Journal of Medical Imaging and Health Informatics*, *8*(1), 33–37.

Zhu, W., Huang, W., Lin, Z., Yang, Y., Huang, S., & Zhou, J. (2016). Data and feature mixed ensemble based extreme learning machine for medical object detection and segmentation. *Multimedia Tools and Applications*, *75*(5), 2815–2837.

Zong, W., & Huang, G. B. (2011). Face recognition based on extreme learning machine. *Neurocomputing*, *74*(16), 2541–2551.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Authors and Affiliations

Rasoul Kiani[1] · **Wei Jin**[2] · **Victor S. Sheng**[1]

✉ Victor S. Sheng
victor.sheng@ttu.edu

Rasoul Kiani
rkiani@ttu.edu

Wei Jin
wei.jin@unt.edu

1 Department of Computer Science, Texas Tech University, Lubbock, TX 79409, USA

2 Department of Computer Science and Engineering, University of North Texas, Dallas, TX 75201, USA