Multilayer extreme learning machine: a systematic review

Ravneet Kaur[1] (ID) **· Rajendra Kumar Roul[1] · Shalini Batra[1]**

Abstract

Majority of the learning algorithms used for the training of feedforward neural networks (FNNs), such as backpropagation (BP), conjugate gradient method, etc. rely on the traditional gradient method. Such algorithms have a few drawbacks, including slow convergence, sensitivity to noisy data, local minimum problem, etc. One of the alternatives to overcome such issues is Extreme Learning Machine (ELM), which requires less training time, ensures global optimum and enhanced generalization in neural networks. ELM has a single hidden layer, which poses memory constraints in some problem domains. An extension to ELM, Multilayer ELM (ML-ELM) performs unsupervised learning by utilizing ELM autoencoders and eliminates the need of parameter tuning, enabling better representation learning as it consists of multiple layers. This paper provides a thorough review of ML-ELM architecture development and its variants and applications. The state-of-the-art comparative analysis between ML-ELM and other machine and deep learning classifiers demonstrate the efficacy of ML-ELM in the niche domains of Computer Science which further justifies its competency and effectiveness.

Keywords Machine learning · Deep learning · ELM · ELM-autoencoder · Feature space · Multilayer ELM

1 Introduction

Artificial neural networks (ANN) refer to the computational models which mimic biological nervous systems such as the human brain. Feedforward neural networks (FNNs) are among

✉ Ravneet Kaur
  rkaur1_phd19@thapar.edu

  Rajendra Kumar Roul
  raj.roul@thapar.edu

  Shalini Batra
  sbatra@thapar.edu

[1] Department of Computer Science and Engineering, Thapar Institute of Engineering & Technology, Patiala, 147004, Punjab, India

the most successful ANNs in which no cycle exists between the node connections. They are known as feedforward as the information travels only in the forward direction throughout the network. The basic functional unit of FNN is a neuron [41, 99, 152]. The major components of FNN are i) Input layer, which comprises of neurons responsible for receiving the input data and forwarding it to other layers, ii) Hidden layer, having neurons that apply transformations to the input data, iii) Output layer, responsible for producing the final output depending on the model; and iv) Neuron weights, used to represent the strength of a connection existing between two neurons. The flow of information occurs from input nodes to hidden layer nodes and finally through the output layer nodes. The loss in FNN is computed based on actual output and predicted output using the loss function. Gradient descent is a commonly used optimization technique to find the local minimum, but it is quite slow [80, 104]. The local minimum refers to a point in the local region where the loss function attains the minimum value. Backpropagation is an algorithm used to perform supervised learning in FNN. The error is computed by comparing predicted and actual output using a function. Further, the error is propagated backward through the layers, and weights are updated based on their contribution.

FNNs have been quite significant since backpropagation (BP) algorithm came into existence [45, 78, 143]. The significant drawbacks of BP include slow convergence, unable to handle large datasets, and the problem of local minima. Although various improvements have been proposed in the FNN training method, most of them do not provide guarantee for an optimal global solution. Thus, an efficient generalized learning algorithm needs to be developed that can train FNNs faster.

1.1 Evolution of ML-ELM

A proficient approach named extreme learning machine (ELM) was put forward to train single hidden layer FNNs [55]. The initialization of hidden nodes is done randomly in ELM, and the main essence is that hidden layers need not be iteratively tuned. The only learnable parameters are the connections or weights between the hidden layer and the output layer, which are analytically computed using the least-square method [10, 53, 130]. ELM shows significant performance as it reduces the overall training time, is more generalized, easy to implement, and can reach a global optimum. In the recent past, many researchers have worked extensively on the theory and applications of ELM [2, 9, 58, 95, 149], and many extensions of ELM have been developed [3, 5, 12, 105, 139, 153]. Despite the merits of ELM over other FNNs, it also has some limitations:

- It can't achieve a high level of data abstraction due to a single hidden layer.
- It's implementation is difficult due to the requirement of a huge network for highly-modified input data.
- Memory constraints in ELM and the massive computation cost for evaluating the inverse of larger matrices make it challenging to manage extensive learning tasks.

To handle the above limitations, ELM gave birth to a new deep learning (DL) architecture called multilayer ELM [64, 128] that uses ELM autoencoders to perform unsupervised learning. It can represent any complex target function easily compared to the prevalent machine learning (ML) architectures and deep networks. The evolution of ML-ELM is depicted in Fig. 1.
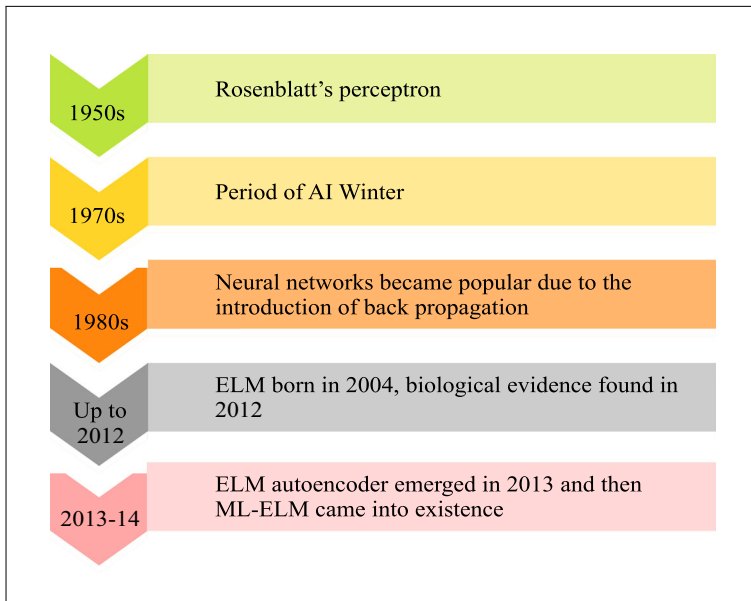
Fig. 1 Evolution of ML-ELM

1.2 Research motivation

The well known DL architectures such as Recurrent Neural Network (RNN), Convolution Neural Network (CNN), Long-short Term Memory (LSTM), etc. [42, 84, 86, 88, 114, 135, 164], suffer from time-consuming training process due to their complicated hierarchical structure and need to fine-tune a large number of parameters. ML-ELM can handle such problems since it does not require iterations during the entire training process, no fine-tuning of parameters is needed, and it maintains a high level of data abstraction [64, 110, 111].

Although many researchers have focused on the survey work of ELM in their research publications [26, 32, 52, 54, 76, 103, 137], ML-ELM is a deep network that requires a thorough review and consideration and is still lacking among the research community. There are very few studies available in the literature [103, 163] which have addressed the survey on ML-ELM architecture. Thus, the major objective or motivation behind this study is to highlight the suitability and effectiveness of ML-ELM, which can further give a new direction and enlighten the opportunities and challenges among the research community.

1.3 Our contributions

As ML-ELM is an emerging field, very few researchers have worked on its extensive review as discussed here. Zhang et al. [163] have presented a review on ML-ELM development, and some commonly used hierarchical structures have been investigated while Parkavi et al. [103] have focused on the recent trends in ELM and ML-ELM. However, our work includes an exhaustive survey on ML-ELM architecture along with discussing its various variants and applications in detail. The comparison of our work with other similar survey papers is presented in Table 1. As it can be seen from Table 1, the present work includes topology

Table 1 Comparison of our work with other similar survey papers

Parameters	Parkavi et al. [103]	Zhang et al. [163]	Our work
Topology of ML-ELM variants	✗	✗	✓
Comparison of ML-ELM with other ML and DL techniques	✗	✗	✓
Applications of ML-ELM	✓	✓	✓
Discussion on ML-ELM feature mapping	✗	✗	✓
Limitations of existing DL architectures	✗	✓	✓
Open issues in ML-ELM	✗	✓	✓
Future directions of research	✓	✗	✓

of ML-ELM variants proposed for better feature learning, handling outliers or noise, optimization of hidden node parameters, reducing multicollinearity, and reducing overfitting; comparison of ML-ELM with other ML and DL techniques and details regarding ML-ELM feature mapping which were not considered in earlier studies. Also, in this paper, the study on different variants and applications of ML-ELM is presented comprehensively. The open issues in ML-ELM have also been discussed in this paper.

Because of the advantages of ML-ELM mentioned earlier, which indicate that it may accelerate the development of DL, a detailed survey work has been carried out starting from the inception of this DL classifier to date. As the paper's main focus is to discuss different variants and application works of ML-ELM, a comprehensive discussion on ELM is presented here. The significant contributions of the paper include the following:

i. A comprehensive discussion on the architecture of ELM, ELM-autoencoder, ML-ELM, and feature mapping of ML-ELM has been done.
ii. An in-depth study on various ML-ELM variants is conducted, and further a topology is defined which helps in understanding the practicality of the existing approaches.
iii. A comparative analysis of ML-ELM with other machine and DL networks is performed.
iv. An extensive survey on different applications of ML-ELM including medical, industrial, academic, security domains etc. is presented.
v. Finally, the open issues in ML-ELM are also listed which can provide future research directions in this field.

1.4 Search criteria for selection of papers

The relevant research articles were extracted from various web domains including Science Direct (https://www.sciencedirect.com/), ACM digital library https://www.acm.org/, Google Scholar (https://scholar.google.com/), Springer (https://link.springer.com/) and IEEE Explore Digital Library (https://ieeexplore.ieee.org) from the year 2006 till 2022 with keywords for search queries being 'Multilayer ELM', 'Deep learning', 'ELM', 'ELM-autoencoder', 'Feature space of ML-ELM', 'ML-ELM variants' and 'ML-ELM applications'. Initially, 1508 search results were retrieved based on the keywords stated earlier. Then, after this step, 560 research articles were excluded with reasons being duplicate entries, different titles, non-relevant abstracts etc. Further, we obtained 142 articles after screening phase which were checked for eligibility. Out of these, 70 papers were excluded
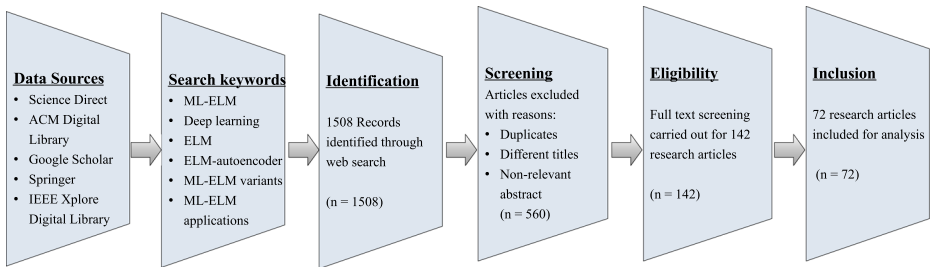
Fig. 2 Search criteria for selection of papers in the current study

after full-text reading as they did not meet the desired outcome. Finally, 72 research articles met the inclusion criteria and thus, were used for analysis in the current study. The above mentioned stages involved in search criteria are described in Fig. 2.

1.5 Paper organization

The rest of the paper is organized as follows: Sections 2 and 3 discuss the basics of ELM and ML-ELM, respectively. In Section 4, different variants of ML-ELM have been discussed according to their topology. A comparison of ML-ELM with other machine and DL architectures is provided in Section 5. Section 6 highlights the application domains of ML-ELM. Finally, the conclusion, open issues, and future enhancements of the work are provided in Section 7. The relationship between various parts of the manuscript is depicted in Fig. 3 and the complete layout is presented in Fig. 4.

2 Fundamentals of extreme learning machine

ELM (shown in Fig. 5) is a shallow network with a single hidden layer [55]. If $\mathbf{X}$ is the input layer with n number of nodes, H is the hidden layer having L number of nodes; then the output layer $\mathbf{Y}$ can be represented by (1).

$$Y_j = \sum_{i=1}^{L} \beta_i g(w_i \cdot x_j + b_i), \ \ j = 1, \cdots N \tag{1}$$

Here, $N \leftarrow$ total number of samples, $w_i \leftarrow$ input weight vector with random initialization connecting input and hidden layer nodes, $b_i \leftarrow$ random bias of ith hidden node,
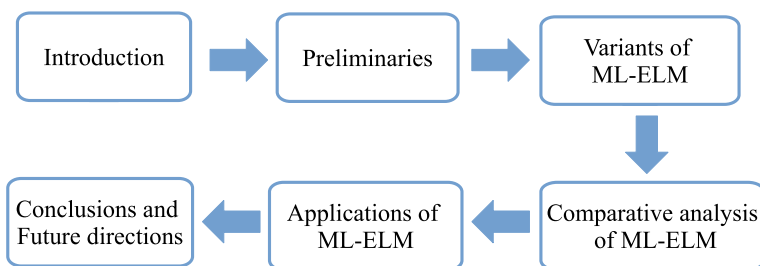

Fig. 3 Relationship between parts of the manuscript

Fig. 4 Layout of the manuscript
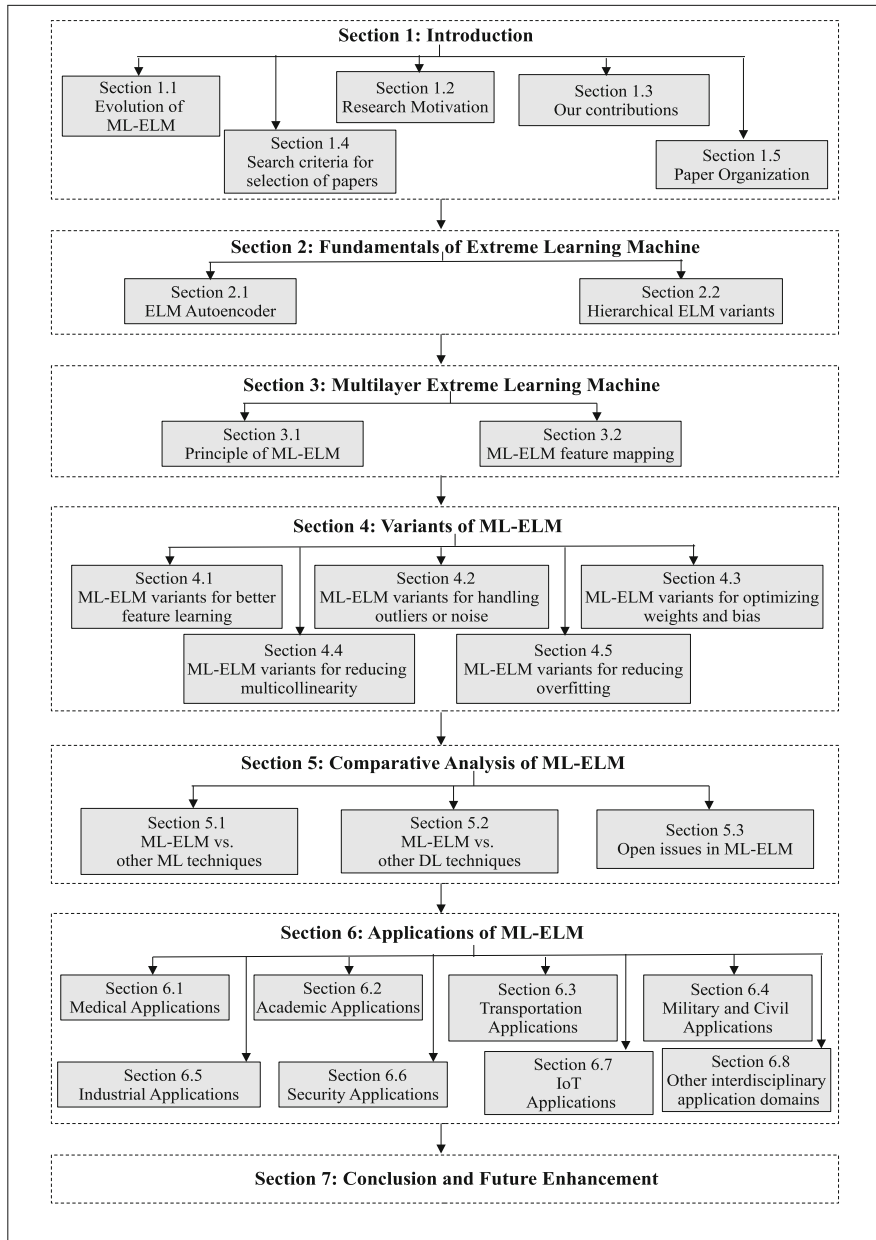
$g \leftarrow$ activation function, and $\beta \leftarrow$ output weight vector connecting output and hidden layer nodes. The relationship among H, β and $\mathbf{Y}$ is represented in (2).
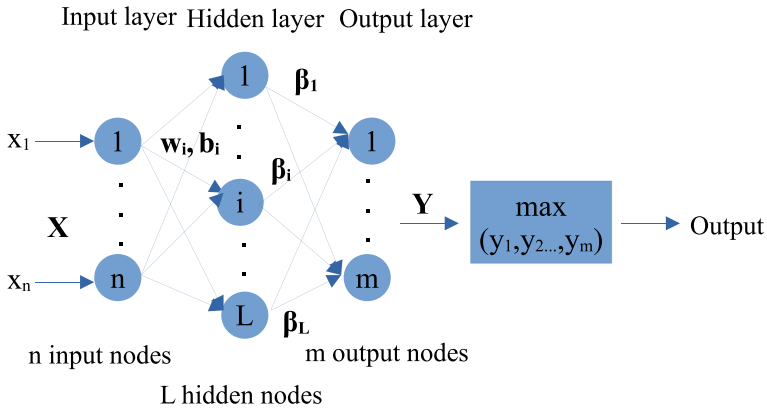
$$H\beta = \mathbf{Y} \tag{2}$$

Fig. 5 Structural diagram of ELM

This implies $\beta = \mathbf{H}^{-1}\mathbf{Y}$, where $\mathbf{H}^{-1}$ is Moore-Penrose inverse [141].

Classic ELM has been modified to improve its performance and make it suitable for real-life problems [5, 11, 12, 44, 105, 139, 153, 162, 165]. ELM finds applications in diverse domains including image processing, computer vision, etc. [2, 3, 6, 13, 16, 58, 91, 95, 128, 149, 150, 157, 157]. Several survey works are available which cover the various variants and applications of ELM in detail [32, 52, 54, 76, 103, 137].

2.1 ELM autoencoder

Zhou et al. [171] proposed ELM autoencoder (ELM-AE), in which the output is a reconstruction of the input supplied to the neural network (Fig. 6) as shown in (3). ELM-AE is a special case of ELM (supervised in nature).

$$\sum_{i=1}^{L} \beta_i g(x_j, w_i, b_i) = x_j, j = 1, 2, \cdots, N \tag{3}$$

Equation (4) represents the relationship between input and output feature vector.

$$\mathbf{H}\beta = \mathbf{X} \tag{4}$$

Unlike ELM, ELM-AE has orthogonal hidden layer weights and biases, enhancing ELM-AE's performance [62] as shown in (5). The output weight vector β, which has three alternative representations, is used to learn the feature space transformation as *equal dimension, sparse and compressed representation*, illustrated in (6), (7) and (8) respectively.

$$\mathbf{h} = g(\mathbf{w}.x + \mathbf{b}), \ \mathbf{w}^T\mathbf{w} = \mathbf{I}, \mathbf{b}^T\mathbf{b} = 1 \tag{5}$$

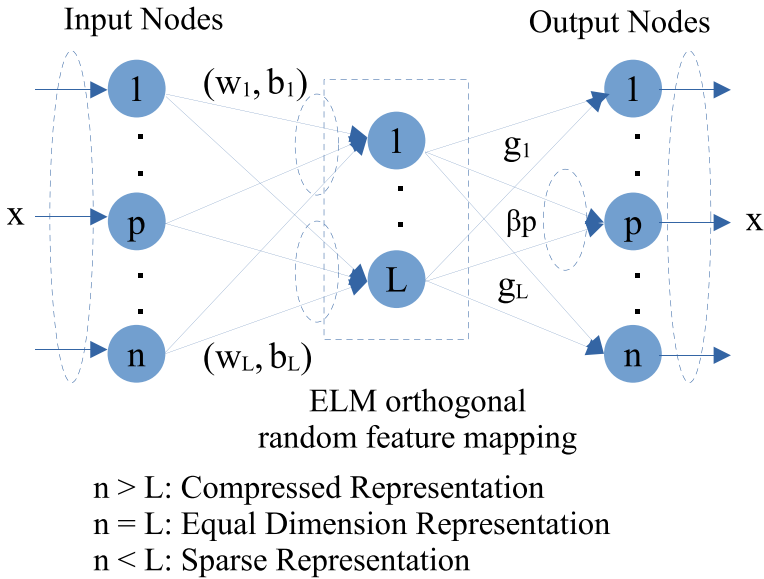ELM-AE, like ELM, is a universal approximator [51].

Fig. 6 Structural diagram of ELM-AE [64]

- *Equal dimension representation*: dimensions of input data and feature space are equivalent, i.e., $n = L$.

$$\beta = \mathbf{H}^{-1}\mathbf{X} \tag{6}$$

- *Sparse representation*: representing features to a higher dimensional feature space from a lower-dimensional input data space, i.e., $n < L$.

$$\beta = \mathbf{H}^T \left(\frac{\mathbf{I}}{C} + \mathbf{H}\mathbf{H}^T \right)^{-1} \mathbf{X} \tag{7}$$

- *Compressed representation*: representing features to a lower dimensional feature space from a higher dimensional input data space, i.e., $n > L$.

$$\beta = \left(\frac{\mathbf{I}}{C} + \mathbf{H}^T\mathbf{H} \right)^{-1} \mathbf{H}^T\mathbf{X} \tag{8}$$

Here, $\mathbf{I}$ represents the unit matrix, and C represents the scale parameter used for adjusting the structural and experimental risk.

Singular value decomposition (SVD) is generally used for feature representation and ELM-AE to represent features based on SVD [8]. The SVD corresponding to (8) is given in (9).

$$\mathbf{H}\beta = \sum_{i=1}^{N} u_i \frac{d_i^2}{d_i^2 + C} u_i^T \mathbf{X} \tag{9}$$

where, d represents the singular values of $\mathbf{H}$ corresponding to the SVD of input data $\mathbf{X}$, and u is the eigenvector of $\mathbf{H}\mathbf{H}^T$. It is hypothesized that the representation of input features is learned by β using singular values, and $\mathbf{H}$ is the projected feature space of the input feature vector $\mathbf{X}$ squashed using a sigmoid function.

2.2 Hierarchical ELM variants

ELM has recently been extended to various multi-layer structures, i.e., multilayer ELM (ML-ELM) [64], hierarchical ELM (H-ELM) [127], and hierarchical local receptive fields ELM (H-LRF-ELM) [48]. ML-ELM is a deep network architecture which performs layer wise unsupervised learning, but requires no backpropagation for parameter tuning, which significantly reduces the computational time. Tang et al. [127] proposed a hierarchical multilayer framework based on ELM named H-ELM, which includes two major components, i.e. feature extraction and supervised classification. It utilizes an ELM-based sparse autoencoder based on l_1-norm optimization. H-ELM is different from ML-ELM on the following lines: 1. ML-ELM uses stacked-layer architecture, whereas H-ELM works by separating the whole network into two distinct subsystems. 2. ML-ELM uses autoencoder based on l_2 norm, whereas H-ELM makes use of l_1 penalty. 3. ML-ELM involves orthogonal initialization of parameters, whereas H-ELM avoids the same. The experimental results showed that H-ELM has a faster learning time and higher learning accuracy. It has significant applications in various domains such as object detection, gesture recognition, etc. Huang et al. [48] put forward the concept of LRF-ELM which uses sparse connections for learning useful representations required in image processing and other related tasks. The enhancement of LRF-ELM to multilayer architecture is termed hierarchical LRF-ELM (H-LRF-ELM). The important components of H-LRF-ELM include feature extractor and ELM.

3 Multilayer extreme learning machine

The main focus of this survey is to conduct an in-depth study of ML-ELM, including its functioning, various architectural works, and applications, which will be discussed in the further sections.

3.1 Principle of ML-ELM

ELM and ELM-AE are combined to form Multilayer ELM, which has more than one hidden layer [64], as shown in Fig. 7. It inherits all the properties from ELM and ELM-AE. Some of the essential features of ML-ELM are:

- The architecture is created by gradually building stacks on the ELM-AE.
- The stacked ELM-AE's first level learns the primary representation of the input data. The higher level of representation in the second level combines the output of the first level, and so on.
- ELM-AE is used for unsupervised training between the hidden layers.
- The hidden layer weights in ML-ELM are initialized using ELM-AE.
- No fine-tuning is required in ML-ELM, unlike other deep networks.
- The output of one trained ELM-AE is used as input for the next trained ELM-AE and so forth.
- Depending on the following two conditions, the hidden layer activation function g is either linear or non-linear:
    - *linear*: if the count of ith and $(i\text{-}1)^{th}$ hidden layer nodes is equivalent.
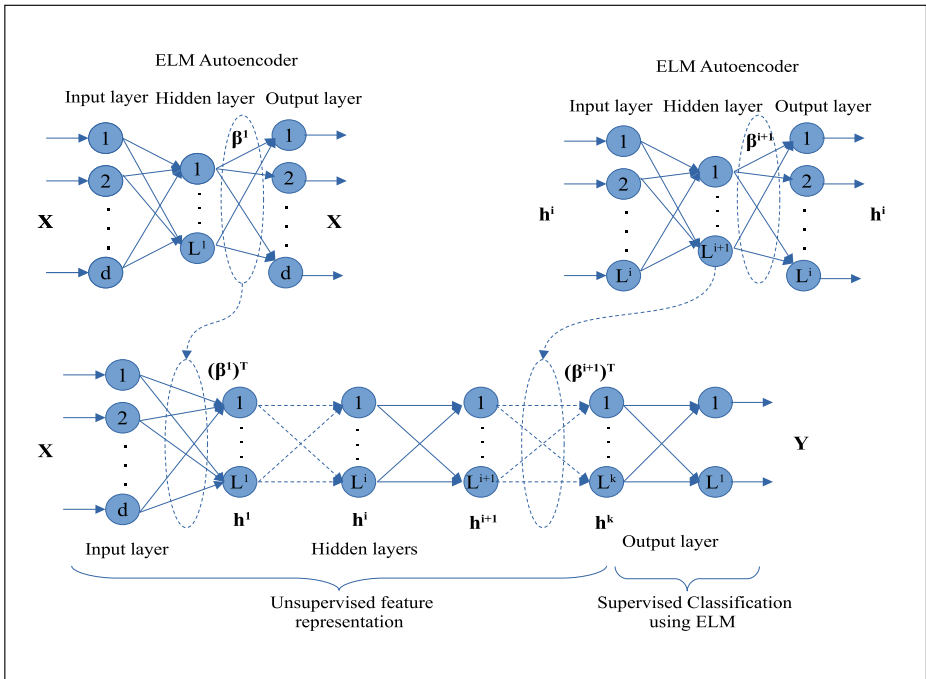   - *non-linear (such as a sigmoidal function)*: if the count of ith and $(i\text{-}1)^{th}$ hidden layer nodes is different.

Fig. 7 Structural diagram of ML-ELM

- The numerical relationship between the ith and $(i\text{-}1)^{th}$ hidden layers is established using (10).

$$H_i = g((\beta_i)^T H_{i\text{-}1}) \tag{10}$$

where H_i and $H_{i\text{-}1}$ are the output and input matrices of the ith hidden layer, and $g(.)$ is the activation function. H_0 is the input layer, and H_1 is the first hidden layer. The output weight β is computed using regularized least squares [109].

- Finally, the network is fine-tuned using ELM.

3.2 ML-ELM feature mapping

The ML-ELM feature mapping (Fig. 8) is discussed further below:

i.   ML-ELM makes extensive use of ELM's universal approximation [54] and classification capabilities [50, 113].

ii. It is well known that ELM is unbiased towards a large number of hidden layer nodes (L) and provides good performance when the size of L is greater than the size of the input vector [47]. This significantly improves the performance of ELM feature mapping.

iii. ML-ELM employs the ELM feature mapping mechanism, which uses sparse representation ($n < L$) property of ELM-AE.
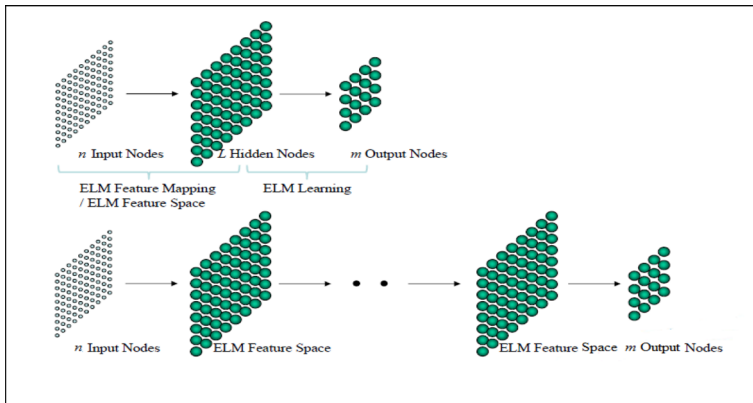
Fig. 8 ML-ELM feature mapping

iv. The features are mapped to higher dimensional space in ML-ELM using (11).

$$
h(\mathbf{x}) = \begin{bmatrix} h_1(\mathbf{x}) \\ h_2(\mathbf{x}) \\ . \\ . \\ . \\ h_L(\mathbf{x}) \end{bmatrix}^T = \begin{bmatrix} g(w_1, b_1, \mathbf{x}) \\ g(w_2, b_2, \mathbf{x}) \\ . \\ . \\ . \\ g(w_L, b_L, \mathbf{x}) \end{bmatrix}^T \tag{11}
$$

where, $h_i(\mathbf{x}) = g(w_i.\mathbf{x} + b_i).h(\mathbf{x}) = [h_1(\mathbf{x}), \cdots, h_i(\mathbf{x}), \cdots, h_L(\mathbf{x})]^T$ can be directly used for feature mapping [49, 51].

v. Kernel techniques are more expensive because they use dot products to find similarities between features in higher-dimensional space. ML-ELM uses ELMs classification capability to do the same without using any kernel technique, which drastically reduces the computational cost.

4 Variants of ML-ELM

A deep-level study of different variants of ML-ELM starting from the commencement to the current date has been carried out in this section. The topology of ML-ELM variants is presented in Fig. 9.

4.1 ML-ELM variants for better feature learning

In ELM with subnetwork nodes, a single hidden node is formed by using other hidden nodes, forming a subnetwork. Yang and Wu [156] put forward an efficient approach for feature learning using subnetwork nodes. The proposed method can be used for different representations, including Dimension Reduction, Expanded Dimension Representation, and Supervised/Unsupervised Learning Mode. Wong et al. [146] put forward a kernel version of ML-ELM named ML-KELM to eliminate the drawbacks of traditional ML-ELM such as unstable performance due to random projection carried out in each layer, huge time consumption due to manual tuning of hidden node count, and slow training speed due to
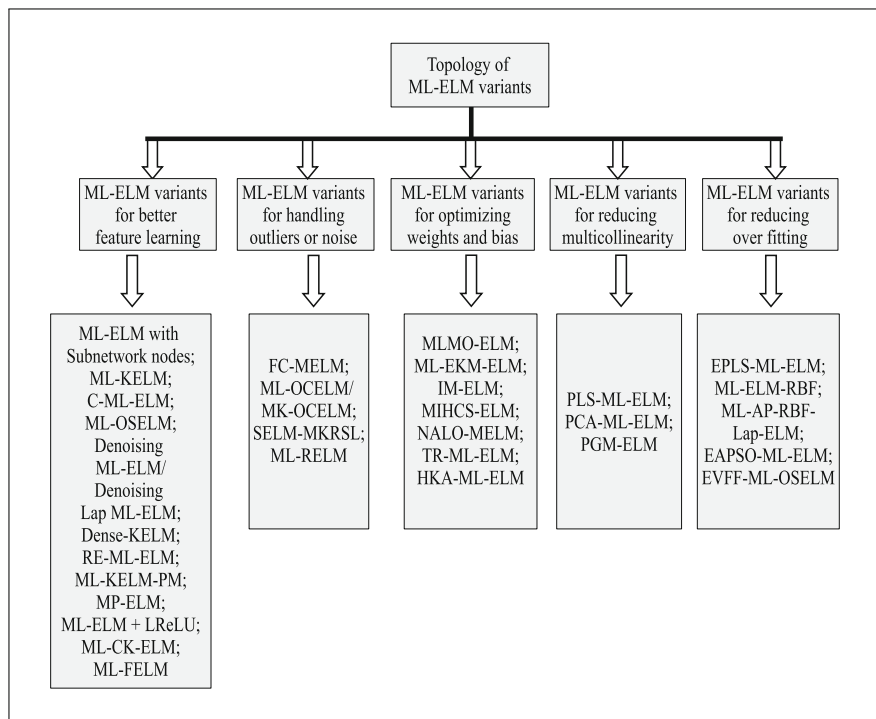
Fig. 9 Topology of ML-ELM variants

more number of hidden layers. The optimization function plays a significant role in deep feature learning methods. The method of optimization adopted in ML-ELM fails to give efficient results while analyzing the variance of input, for example, noise disturbance, image deformation etc. Jia et al. [59] proposed C-ML-ELM, which includes a penalty term in the optimization function whose aim is to minimize the first-order derivative value of output to the input to improve the classification results. This is based on the principle of Contractive Auto-Encoder (CAE), discussed by Rifai et al. [108]. The framework of C-ML-ELM consists of a stack of contractive ELM-AE rather than a traditional auto-encoder at every layer to ensure robust generalization and faster learning capability. Mirza et al. [92] proposed an online sequential version of basic ML-ELM, namely MS-OSELM, which uses online sequential ELM-AE (OS-ELM-AE) to learn robust features from sequential streaming data. It utilizes Online Sequential ELM Auto-Encoder (OS-ELM-AE), which involves random and orthogonal weights and biases. Zhang et al. [160] proposed a variant of ML-ELM, namely Denoising ML-ELM, which uses denoising auto-encoder (ELM-DAE) to incorporate prior knowledge, which further enhances the performance of the classification model. ELM-DAE incorporates denoising criteria into basic ELM-AE to ensure the robustness of the learning features. The inputs to ELM-DAE are the corrupted samples (having noise) represented by $\tilde{\mathbf{X}}$, and the outputs are the original training samples represented by $\mathbf{X}$ where $\mathbf{X} = \{x_i \in R^j\}_{i=1}^{N}$. The framework of Denoising ML-ELM comprises of a stack of ELM-DAE to initialize the weights of the hidden layers. The output weights β are computed using (7) and (8). Zhang et al. [160] introduced the manifold regularization term in the

original cost function corresponding to Denoising ML-ELM and named the variant Denoising Laplacian ML-ELM. Jiang et al. [61] proposed Densely Connected Multilayer Kernel ELM (Dense-KELM) to solve the problems of enormous memory consumption and huge training time faced during the classification of remote sensing image scenes. The training process of Dense-KELM is identical to ML-KELM. Region-enhanced ML-ELM (RE-ML-ELM) proposed by Jia et al. [60] utilizes several input nodes to perform multiplexing of the local significant region. The input data for RE-ML-ELM comprises two main parts: source data and the data obtained from the local significant region. The parameters are computed for each layer by using ELM-AE, and the incorporation of additional information from data helps in improving the representation learning. Fei et al. [29] further presented projective model (PM) based ML-KELM (ML-KELM-PM), which enhances the feature representation as well as classification accuracy of ML-KELM. Zhang et al. [166] presented Multilayer probability ELM (MP-ELM), which automatically enables the extraction of valuable information from the links in Device-free localization (DFL). MP-ELM makes use of ELM autoencoders to maintain the fast learning speed of ELM and returns probabilistic output to enable fast and accurate DFL. Nayak et al. [97] presented an ML-ELM variant which uses LReLU as the activation function. Another ML-KELM variant using combined kernels (ML-CK-ELM) was proposed by Rezaei et al. [107] for multi-label classification. Hernandez et al. [46] proposed Multilayer Fuzzy Extreme Learning Machine (ML-FELM), which consists of stacked Fuzzy Autoencoders to achieve high input feature representation. It also makes use of the Mamdani Fuzzy Logic System for performing classification.

A summary of ML-ELM variants proposed for better feature learning is presented in Table 2.

4.2 ML-ELM variants for handling outliers or noise

Liangjun et al. [77] proposed a correntropy-based ML-ELM, namely FC-ELM, to efficiently classify the datasets corrupted by outliers or noise. Correntropy refers to a non-linear measure that computes the similarity between two random variables. The correntropy for two random variables suppose S and T is defined as:

$$C(S, T) = E[\kappa(S, T)] \tag{12}$$

where E represents expectation and κ represents kernel function which satisfies Mercer's theorem. Using correntropy reconstruction loss instead of mean square error (MSE) makes the ELM autoencoder robust to noise. Dai et al. [21] proposed a multilayer architecture based on OC-ELM (ML-OCELM) and a kernel version of the same (MK-OCELM). The experimental results showed good generalization and less human intervention. Luo et al. [83] put forward a variant of ML-ELM that incorporates the criterion of Kernel risk-sensitive loss (KRSL) to ensure robust performance for training datasets with outliers or noise. KRSL is a measure used for local similarity which utilizes the concept of correntropy [17]. For two random variables, A and B, KRSL is computed using (13).

$$L_\lambda(A, B) = \frac{1}{\lambda} E[\exp(\lambda(1 - \kappa_\sigma(A - B)))]$$
$$= \frac{1}{\lambda} \int \exp(\lambda(1 - \kappa_\sigma(a - b))) dF_{AB}(a, b) \tag{13}$$

where λ represents the risk-sensitive parameter, σ represents the kernel bandwidth, $\kappa_\sigma(.)$ represents the Mercer kernel function [115], $E(.)$ represents the mathematical expectation,

Table 2 Summary of ML-ELM variants for better feature learning

Authors	Name of variant	Features	Results	Application areas	Datasets used
Yang and Wu [156]	ML-ELM with Subnetwork Nodes	Uses subnetwork nodes where a single hidden node is formed by using other hidden nodes, forming a subnetwork.	Exhibits better generalization as compared to other feature learning methods.	Image Recognition	Caltech-101 [30], Caltech-256 [37], MNIST [72] etc.
Wong et al. [146]	ML-KELM	Employs kernel and transformation matrices for better feature representation.	Reduces the storage requirement and model execution time.	Highly non-linear applications	Cmc [79], Yeast [96], Isolet [20], Abalone [140] etc.
Jia et al. [59]	C-ML-ELM	Includes a penalty term in the optimization function whose aim is to minimize the first-order derivative value of output to the input.	Achieves minimum training error while maintaining fast learning speed.	Face recognition, expression recognition etc.	MNIST handwriting dataset [72] and face expression dataset: CAFÉ [22]
Mirza et al. [92]	ML-OSELM	Uses online sequential ELM-AE (OS-ELM-AE) to learn robust features from sequential streaming data.	Performs better than online sequential ELM having a single hidden layer.	Batch processing of data, Image classification	CIFAR10 [67], CALTECH [30], Olivetti faces [112] and UMIST [36]
Zhang et al. [160]	Denoising ML-ELM/ Denoising Lap ML-ELM	Makes use of ELM-DAE and regularization respectively to incorporate prior knowledge in the model.	Shows slightly better classification performance than ML-ELM, and training time is almost similar to that of ML-ELM.	Desirable for supervised as well as semi-supervised ML problems	MNIST dataset [72], G50C [14], and DBWorld e-mails dataset [31]
Jiang et al. [61]	Dense-KELM	Involves highly connected network framework and has a training process similar to ML-KELM.	Shows promising results for classification accuracy compared to ML-KELM, which has less number of parameters and training time.	Useful for feature extraction of remote sensing image scenes	NWPU-RESISC45 dataset [18]

Table 2 (continued)

Authors	Name of variant	Features	Results	Application areas	Datasets used
Jia et al. [60]	RE-ML-ELM	Utilizes several input nodes to perform multiplexing of the local significant region.	Exhibits fast training time compared to various other DL networks.	Image classification	MNIST [72], NORB [73], and ORL [39]
Fei et al. [29]	ML-KELM-PM	Incorporates projective model in ML-ELM, which aids in parameter transfer approach.	Results in better feature representation and more stability for classification.	Breast tumour diagnosis	Bimodal ultrasound image dataset
Zhang et al. [166]	MP-ELM	Makes use of ELM autoencoders to maintain the fast learning speed of ELM and returns probabilistic output to enable fast and accurate Device-free localization (DFL).	Performs better as compared to ELM, hierarchical ELM, ML-ELM, DBM, and deep belief network (DBN).	Suitable for various applications involving uncertainty	Real-time data
Nayak et al. [97]	ML-ELM + LReLU	Uses LReLU as the activation function in ML-ELM	Results in more stability and generalization.	Image classification	MCHD dataset [63]
Rezaei et al. [107]	ML-CK-ELM	Utilizes a combination of different kernels to address the drawbacks of ML-ELM.	Shows better performance as compared to other state-of-the-art approaches.	Multi-label classification problems	Yeast [96], Scene, Computer, Education, Recreation etc.
Hernandez et al. [46]	ML-FELM	Combines Fuzzy Logic theory with ML-ELM model to improve feature representation.	Results in higher accuracy of the model.	Image processing and classification, transport of objects using an unmanned aerial vehicle.	Abalone [140], Letter Recognition [118]etc.

and $F_{AB}(a, b)$ illustrates the distribution of (A,B). Stacked ELM, i.e., multilayer neural network utilizing multiple sub-ELMs, was presented by Luo et al. [83] which utilizes KRSL as the loss function (SELM-MKRSL). The MSE based loss function used in ML-ELM is quite sensitive to outliers and noise. Yu et al. [132] presented a robust multilayer approach named ML-RELM, which incorporates model bias and variance in the loss function to reduce the influence of noise signals. The experimental results showed higher generalization and greater robustness to noise.

A summary of ML-ELM variants proposed for handling outliers or noise is presented in Table 3.

4.3 ML-ELM variants for optimizing weights and bias

The Multi Layer Multi Objective ELM, MLMO-ELM, proposed by Lekamalage et al. [75], uses a multi-objective formulation for learning the hidden layer parameters using Multi Objective ELM-AE (MO-ELMAE). The framework of MLMO-ELM consists of a stack of MO-ELMAE to learn hidden layer parameters. The pth hidden layer output denoted by $\mathbf{H}^p$ is used to learn the parameters of $(p + 1)^{th}$ hidden layer of MLMO-ELM. Ridge regression is used to learn the parameters of the output layer. Vong et al. [134] presented an improved version of ML-KELM, that encodes hidden layer in the form of empirical kernel map (EKM), and named it ML-EKM-ELM. This further leads to its suitability for large-scale problems. Le et al. [71] put forward an ML-ELM variant, namely Incremental ML-ELM (IM-ELM), for efficiently determining the number of hidden layers, i.e., K. The basic idea was to include a layer corresponding to $K + 1$ network to the last hidden layer of ML-ELM (having K hidden layers). The final output of the new network is given by (14).

$$f_{K+1}(x) = \mathbf{H}_{K+1}\beta_{K+1} \tag{14}$$

This can aid in finding the most suitable value of K corresponding to initial input weight $\mathbf{W}_1$ and initial bias $\mathbf{B}_1$. Wu et al. [147] proposed a novel approach called Multilayer Incremental Hybrid Cost Sensitive ELM with Multiple Hidden Output Matrix and Subnetwork Hidden Nodes (MIHCS-ELM). It uses ant clone and multiple greywolf optimization methods for computing optimal hidden node parameters. Zheng et al. [170] presented the usage of a novel ant lion algorithm (NALO) for optimizing random weights and bias involved in ML-ELM (NALO-MELM), which can affect its accuracy. He et al. [43] proposed a tree root algorithm based ML-ELM (TR-ML-ELM), which gives better classification accuracy than ML-ELM. Ma et al. [85] put forward a heuristic kalman algorithm (HKA) based ML-ELM (HKA-ML-ELM) where HKA was used to optimize parameters of ML-ELM. This further leads to an improvement in prediction accuracy.

A summary of ML-ELM variants proposed for optimizing weights and bias is presented in Table 4.

4.4 ML-ELM variants for reducing multicollinearity

Multicollinearity refers to a situation when there is a high correlation between independent variables in a regression problem. This problem is faced in ML-ELM due to the last hidden layer in the network. Su et al. [123] proposed PLS-ML-ELM, which uses partial least square (PLS) to eradicate the problem of multicollinearity. Principal Component Analysis (PCA) is a method adopted for dimensionality reduction. It can identify

Table 3 Summary of ML-ELM variants for handling outliers or noise

Authors	Name of variant	Features	Results	Application areas	Datasets used
Liangjun et al. [77]	FC-MELM	Includes correntropy criterion to efficiently classify the datasets corrupted by outliers or noise.	Exhibits more robust performance as compared to hierarchical ELM (H-ELM) in case of severe noise and has higher classification accuracy than ML-ELM.	Suitable for datasets having noise	MNIST [72], NORB [73], Moore network traffic [94] and Wisconsin Breast Cancer Data [145]
Dai et al. [21]	ML-OCELM/ MK-OCELM	Combines the advantages of ML-ELM and ML-KELM to improve the original OC-ELM.	Exhibits less computational time.	Acoustic classification, Fault detection, Document classification etc.	Abalone [140], Liver [133], Breast [172] etc.
Luo et al. [83]	KRSL based ML-ELM	Incorporates criterion of KRSL to ensure robust performance for training datasets having outliers or noise.	Outperforms other classification methods in terms of accuracy and is quite robust to noise or outliers due to the usage of KRSL.	Problems involving massive amount of data with outliers	USPS [56], ISOLET [20], SPECT heart [69], Internet advertisements [70], Balance scale [66] etc.
Yu et al. [132]	ML-RELM	Makes use of model bias and variance to reduce the effect of noise signals.	Shows high generalization and noise robustness.	Image classification	MNIST handwritten data [72], NORB dataset [73] etc.

Table 4 Summary of ML-ELM variants for optimizing weights and bias

Authors	Name of variant	Features	Results	Application areas	Datasets used
Lekamalage et al. [75]	MLMO-ELM	Makes use of a multi-objective formulation for learning the hidden layer parameters using Multi Objective ELM Auto-Encoder (MO-ELMAE).	Performs better in comparison to ML-ELM, H-ELM, and Deep Boltzmann Machine (DBM).	Object recognition, character recognition etc.	NORB [73] and OCR [131] datasets
Vong et al. [134]	ML-EKM-ELM	Encodes hidden layer in the form of EKM, which is computed from kernel matrix approximation.	Exhibits fast training time and has low storage requirements.	Text categorization or other large-scale applications	NORB [73], CMC [79], Yeast [96], Isolet [20] etc.
Le et al. [71]	IM-ELM	Determines the number of hidden layers to be used in the ML-ELM model.	More stability and higher classification accuracy.	Coal quality exploration	Manually collected dataset
Wu et al. [147]	MIHCS-ELM	Uses optimization methods for hidden node parameters and combines the advantages of Cost Sensitive ELM and sub-network hidden nodes.	Achieves better results as compared to other ELM based methods.	Face recognition, other classification and regression problems.	MNIST [72], Abalone [140], California housing [100] etc.
Zheng et al. [170]	NALO-MELM	Makes use of NALO algorithm for optimizing ML-ELM parameters.	Shows good generalization and better prediction accuracy.	Fault diagnosis methods	Test bearing dataset from Case Western Reserve University
He et al. [43]	TR-ML-ELM	Utilizes a combination of TR algorithm and ML-ELM model to incorporate better stability and accuracy.	Better recognition ability as compared to other methods.	Monitoring changes in coal mines	Real-time data
Ma et al. [85]	HKA-ML-ELM	Uses HKA algorithm to optimize input weights and biases of ML-ELM model.	Results in better prediction accuracy.	Prediction of remaining useful life of Lithium-ion batteries.	NASA [33] and Oxford University battery degradation [7] datasets

the most significant data components by eliminating redundancy [34, 119]. Su et al. [125] put forward a variant of ML-ELM incorporating the PCA method, PCA-ML-ELM, to improve the performance of basic ML-ELM. PCA-ML-ELM, utilizes ELM-AE to obtain the output matrix corresponding to each hidden layer. Zhang et al. [167] put forward an ML-ELM variant, i.e. Self-adaptive ML-ELM model with dynamic generative adversarial net (GAN) [35] (PGM-ELM), which finds application in the classification of biomedical data.

A summary of ML-ELM variants proposed for reducing multicollinearity is presented in Table 5.

4.5 ML-ELM variants for reducing overfitting

ML-ELM might face the problem of overfitting training data. Su et al. [123] proposed an improved version of PLS-ML-ELM by combining it with the ensemble model [124]. Since different simulation results might be generated in various trials, the ensemble model was combined with the algorithm to generate EPLS-ML-ELM which overcame the above-mentioned problems and ensured better generalization. Zhang et al. [161] proposed an ML-ELM variant, namely Radial basis function based ML-ELM (ML-ELM-RBF), for performing multi-label learning by making use of weight uncertainty ELM-AE (WuELM-AE) to solve the overfitting problem. WuELM-AE incorporates weight uncertainty into ELM-AE to ensure robust learning of features. The framework of ML-WuELM consists of a stack of WuELM-AE to learn the parameters of each layer. Further, Xu et al. [151] proposed ML-AP-RBF-Lap-ELM algorithm for multi-label learning, which combines Affinity Propagation (AP) clustering algorithm, ML-RBF, and Lap-ELM. The experimental results showed good stability on various datasets, but the accuracy and generalization capability still improvement. Su et al. [126] proposed EAPSO-ML-ELM, which makes use of adaptive particle swarm optimization (APSO) [117] and ensemble model to enhance ML-ELMs performance. APSO can be used for optimizing input and hidden layer weights and biases that are selected randomly in ML-ELM. The proposed ensemble model contributes towards overcoming the over-fitting problem faced in ML-ELM. The variant EAPSO-ML-ELM is a combination of K ML-ELMs optimized using APSO. The various steps followed are:

Step 1: K optimized ML-ELMs are generated by applying the training phase.
Step 2:   Prediction output matrix O_i is generated through the testing data where $i = 1, \cdots, K$.
Step 3:   The final result is computed by using an average of all K ML-ELMs results (15).

$$O_{final} = \frac{1}{K} \sum_{i=1}^{K} O_i \qquad (15)$$

Su et al. [122] further presented an improved version of MS-OSELM based on variable forgetting factor (VFF) and ensemble model (EVFF-ML-OSELM) where VFF is incorporated to give more emphasis to new incoming data and the ensemble model is used to avoid overfitting.

A summary of ML-ELM variants proposed for reducing overfitting is presented in Table 6. Also, a comparative analysis of various ML-ELM variants based on suitable characteristics is illustrated in Table 7.

Table 5 Summary of ML-ELM variants for reducing multicollinearity

Authors	Name of variant	Features	Results	Application areas	Datasets used
Su et al. [123]	PLS-ML-ELM	Uses PLS to overcome the multicollinearity problem in ML-ELM.	Results in less computation time and higher accuracy.	Prediction model for permeability index	Production data from blast furnace
Su et al. [125]	PCA-ML-ELM	Incorporates PCA to improve the performance of basic ML-ELM.	Exhibits better accuracy as well as robustness.	Prediction model of permeability index	Abalone [140] and California Housing [100]
Zhang et al. [167]	PGM-ELM	Utilizes GAN [35] to generate class samples dynamically for balancing the distribution of various classes.	Shows promising results in terms of ROC curve, F-measure, etc., as compared to other representative methods and exhibits better classification performance.	Suitable for classification problems	Breast cancer [172], Liver patient [133], Diabetic retinopathy [4] and Pima diabetes dataset

Table 6 Summary of ML-ELM variants for reducing overfitting

Authors	Name of variant	Features	Results	Application areas	Datasets used
Su et al. [123]	EPLS-ML-ELM	Used partial least square (PLS) to eradicate the problem of multicollinearity emanating due to the last hidden layer output matrix in ML-ELM.	Exhibits better prediction accuracy and generalization.	Data driven prediction model for blast furnace	Production data from blast furnace
Zhang et al. [161]	ML-ELM-RBF	Performs multi-label learning by making use of weight uncertainty ELM-AE (WuELM-AE) to solve the overfitting problem.	Performs better in comparison to basic ML-ELM and shows good classification results for both single label and multi-label datasets.	Suitable for classification problems	NUS-WIDE dataset [19]
Xu et al. [151]	ML-AP-RBF-Lap-ELM	Integrates AP clustering algorithm, ML-RBF, and Lap-ELM for multi-label learning.	Exhibits good generalization and improved learning ability	Appropriate for problems dealing with multi-label data	Natural Scene [1], Yeast Gene [38], and 20NG dataset [93]
Su et al. [126]	EAPSO-ML-ELM	Makes use of APSO [117] and ensemble model to optimize hidden layer weights and biases and avoid overfitting, respectively.	Shows higher prediction accuracy and better generalization performance compared to various other learning algorithms, including ELM, ML-ELM, etc.	Prediction model for blast furnace	Production data from blast furnace
Su et al. [122]	EVFF-ML-OSELM	Based on the combination of ML-OSELM, VFF and ensemble model.	Exhibits higher prediction accuracy and generalization performance.	Prediction model for hot metal silicon content.	Real-time production data from blast furnace

Table 7 Suitable characteristics of various ML-ELM variants

ML-ELM variant	Kernel based learning	Clustering	Sequential learning	Optimal parameters
ML-ELM with Subnetwork Nodes	✗	✗	✗	✓
ML-KELM	✓	✗	✗	✓
C-ML-ELM	✗	✗	✗	✗
ML-OSELM	✗	✗	✓	✗
Denoising ML-ELM/ Denoising Lap ML-ELM	✗	✗	✗	✗
Dense-KELM	✓	✗	✗	✗
RE-ML-ELM	✗	✗	✗	✗
ML-KELM-PM	✓	✗	✗	✓
MP-ELM	✗	✗	✗	✗
ML-ELM + LReLU	✗	✗	✗	✗
ML-CK-ELM	✓	✗	✗	✓
ML-FELM	✓	✗	✗	✗
FC-MELM	✓	✗	✗	✓
ML-OCELM/ MK-OCELM	✓	✗	✗	✓
KRSL based ML-ELM	✓	✗	✗	✓
ML-RELM	✗	✗	✗	✗
MLMO-ELM	✗	✗	✗	✗
ML-EKM-ELM	✓	✗	✗	✓
IM-ELM	✗	✗	✗	✗
MIHCS-ELM	✗	✗	✗	✓
NALO-MELM	✗	✗	✗	✓
TR-ML-ELM	✗	✗	✗	✗
HKA-ML-ELM	✗	✗	✗	✓
PLS-ML-ELM	✗	✗	✗	✗
PCA-ML-ELM	✗	✗	✓	✗
PGM-ELM	✗	✗	✗	✓
EPLS-ML-ELM	✗	✗	✗	✗
ML-ELM-RBF	✗	✓	✗	✗
ML-AP-RBF-Lap-ELM	✗	✓	✗	✗
EAPSO-ML-ELM	✗	✗	✗	✓
EVFF-ML-OSELM	✗	✗	✓	✗

5 Comparative analysis of ML-ELM

5.1 ML-ELM vs. other ML techniques

Numerous conventional ML classifiers include SVM, k-nearest neighbor, decision trees, etc. A few drawbacks of these classifiers are the requirement of massive resources for functioning, huge unbiased and good quality datasets, high error-susceptibility, and restrictions to approximate a complex function.

ELM is preferred over other ML techniques because it does not depend much on the number of hidden nodes present, requires no iterative tuning, has fast learning speed, exhibits

good generalization performance, and ensures parallel and distributed computation which makes it appropriate for real-time problems.

SVM is one of the well-established classifiers which has been used for various applications by the research communities from time to time [90, 144, 158]. A comparative analysis of SVM, ELM, and ML-ELM is presented in Table 8.

5.2 ML-ELM vs. other DL techniques

This section discusses the limitations of different DL architectures [81] and the advantages of ML-ELM over them.

5.2.1 Limitations of existing DL techniques

The existing DL techniques involve long training time, high computational cost, and massive training data. Some of the limitations of state-of-the-art DL classifiers are:

i. Limitations of Convolution Neural Network (CNN):

- CNN sends all low-level neurons details to higher level neurons, and these higher level neurons again perform convolutions by replicating the knowledge across all the different neurons for checking whether certain features are present or not. This is a time-consuming job.
- CNNs depend on the initial parameter tuning to prevent local optima. Hence, CNNs need to perform many computations for initialization depending upon the problem at hand.
- As the convolution is very slow for forward and backward operation, deep networks require a lot of time for training.
- CNNs have many parameters, and hence, it often exhibits overfitting for small datasets.
- Hyperparamter tuning in CNN is non-trivial.
- It is not spatially invariant to the input data.
- Using traditional CPUs, training CNNs is time-consuming and expensive. Hence, good GPUs are needed for faster training.

ii. Limitations of Recurrent Neural Network (RNN):

- RNN cannot be stacked for an intense deep model and cannot sustain long-term dependencies.
- RNN is prone to gradient vanishing and exploding during backpropagation. It makes the training of RNN difficult in two ways:- a) if *tanh* is used as the activation function, then it cannot process very long sequences b) if ReLu is used as the activation function, then it becomes more unstable.
- Due to its recurrent nature, the computation is slow.
- Like CNN, RNN also needs GPUs for faster training.

iii. Limitations of Long Short Term Memory (LSTM):

- As the data must be moved from one cell to another for evaluation, and the cells are rather complicated due to a few extra characteristics such as forget gates, LSTMs cannot eliminate the vanishing gradient problem.

Table 8 Comparative analysis of the properties of SVM, ELM and ML-ELM

Properties	SVM	ELM	ML-ELM
Belief	No such belief	Learning can be performed without carrying out the tuning of hidden layer parameters.	ELM-AE is used to initialize hidden layer weights to perform layer-wise unsupervised learning.
Biological inspired	Yes	Yes	Yes
Network output functions	$f(x) = \sum_{s=1}^{N_s} \alpha_s t_s \phi(x) \cdot \phi(x_s) + b$	$f_L(x) = \sum_{i=1}^{L} \beta_i g(a_i, b_i, x)$	$H_k = g((\beta^k)^T H_{k-1})$
Multiclass classification	Gives indirect solutions depending on binary case ($t_i = 0$ or 1)	Gives direct solutions	Gives direct solutions
Explicit feature mappings	No (Mapping $\phi(x)$ is unknown and only kernel is used.)	Yes (Different types of explicit feature mappings, h(x) or kernels, can be utilized.)	Feature mapping is used but doesn't use the kernel technique as it is costly.
Hidden node types (mathematical model)	Kernels	Wide variety such as kernel, sigmoid, etc.	Wide variety such as kernel, sigmoid, etc.
Hidden node types (biological neurons)	No	Yes	Yes
Domain	Mainly used for real domain (Finds difficulty in handling complex domain directly)	Applicable for real and complex domains	Applicable for real and complex domains
SLFNs	No	Used as generalized SLFN or other types of SLFNs	No, but ELM-AE is based on SLFN
Layer wise feature representation	No (The representation of features in different layers are ignored.)	Yes	Yes
Connectivity	No focus on network connections	Used for fully connected as well as randomly (partially) connected network	Used for fully connected as well as randomly (partially) connected network
Hyperplane constraints in dual problem	Yes (Because of bias b associated with output nodes.)	No (Because of no bias b associated with output nodes.)	No

Table 8 (continued)

Properties	SVM	ELM	ML-ELM
Universal approximation and classification capability	No theoretical proofs provided	Theoretical proofs provided for various types of non-linear piecewise nodes / neurons	Theoretical proofs provided for various types of non-linear piecewise nodes / neurons
Ridge regression theory	No (The concept of Maximal margin is a specific case of ridge regression utilized in binary classification.)	Yes (Exhibits consistent performance in case of feature learning, regression, clustering, and binary/ multiclass classification.)	Yes
Learning capability	Faces difficulty while handling auto-encoders	Quite efficient for feature learning (auto-encoders) as well as clustering	Quite efficient for feature learning, classification, and dimension reduction
Solutions	Non-closed form	Closed-form as well as non-closed form, sequential, online, and incremental	Sequential, online and incremental

- It requires a high volume of resources such as a large number of tuned parameters and high memory bandwidth for training.
- Random weights initialization affects the performance of LSTMs, and hence it behaves similar to FNN.
- As LSTMs are vulnerable to overfitting, using dropout methods to control this issue is time-consuming.

iv. Limitations of Artificial Neural Network (ANN):

- When the training is finished, the network attempts to optimize the error by generating a specific result, but that is not optimum.
- Like CNN and RNN, ANN also needs good GPUs to avoid slow training.
- There are no proper guidelines or rules to assure an appropriate network structure in ANN. This can only be achieved through trial and error and rich experience.
- In the majority of cases, overfitting persists in the network.
- The result generated by the network does not give a clue as to how and why it is produced, which in turn reduces the trust in ANN.

5.2.2 Advantages of ML-ELM over machine and DL techniques

A detailed comparison of ML-ELM with the above-mentioned state-of-the-art DL mechanisms is provided in Table 9. ML-ELM suitably address the complexities that arise in ML and DL techniques on the following lines:

i. ML-ELM has fewer parameters, no back propagations, and no fine-tuning of hidden node parameters. Also, an increase in the number of hidden nodes and layers won't affect ML-ELM much as compared to other techniques.

ii. ML-ELM involves less training time and has a fast learning speed during the classification process, due to which it doesn't require GPUs.

iii. It performs well on a large volume of datasets.

iv. The classification and approximation capabilities of ELM are the strength of ML-ELM. It can map large datasets to an extended feature space, and can be separated linearly without using any kernel techniques, which can save a lot of resources.

v. Unsupervised training is carried out among the hidden layers using ELM-AE.

vi. Understanding the architecture of ML-ELM is very simple, and it is less complex to implement.

5.3 Open issues in ML-ELM

Some of the open issues in ML-ELM are described below:

i. As the hidden layer parameters are generated randomly, and there is no backpropagation, the training of ML-ELM is quite fast. But some variants of ML-ELM are sensitive to the randomization process, and thus, their changes may degrade the performance of ML-ELM. Further studies are required to handle such problems.

ii. The behavior and the correct number of hidden units, including hidden layer activation function, and the parameters in each layer in ML-ELM, are still debatable and can be reviewed further.

iii. Feature mapping of ML-ELM using universal and classification capabilities of ELM is very robust. However, more research inputs are required to justify such robustness.

Table 9 ML-ELM vs. other DL architectures

| Features	ANN	CNN	RNN	LSTM	ML-ELM
Data	Tabular data	Image data	Sequence data	Sequence data	Both image and sequence data
Backpropagation	Yes	Yes	Yes	Yes	No
Training Time(using CPU)	High	High	High	High	Low
GPU needed	Yes	Yes	Yes	Yes	No
Vanishing and Exploding Gradient	Yes	Yes	Yes	Yes	No
Overfitting	Yes	Yes	Yes	Yes	Yes
Parameter sharing	No	Yes	Yes	Yes	Yes
Recurrent connections	No	No	Yes	Yes	No
Spatial relationship	No	Yes	No	No	Yes
Parameter requirement	High	High	High	High	Low
Memory bandwidth	High	High	High	High	Low
Structural complexity	High	High	High	High	Low
Tuning of hidden parameters	Required	Required	Required	Required	Not required
Implementation	Complex	Complex	Complex	Complex	Simple
Data abstraction	High	High	High	High	High
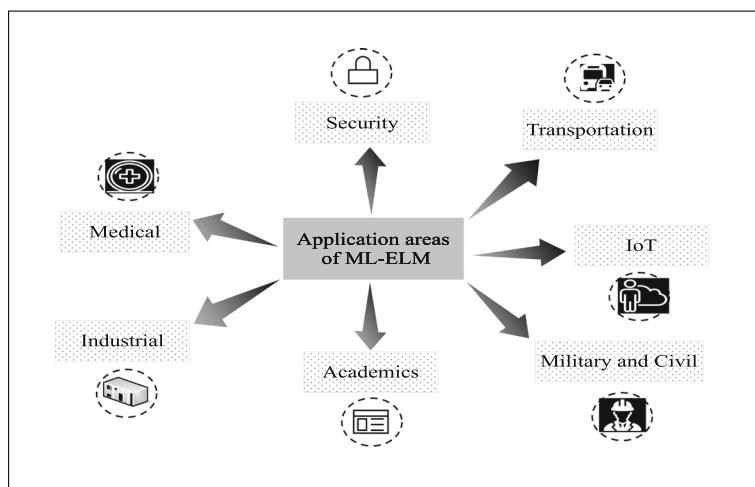
Fig. 10 Application areas of ML-ELM

iv. DL methods such as CNN, RNN, and LSTM need a huge volume of data and many tuned parameters to train the network. Further investigations are required to find whether the combination of such DL architectures with ML-ELM can reduce the requirement of parameters without compromising the performance.

6 Applications of ML-ELM

The advantages of ML-ELM including training speed, accuracy, and generalization make it suitable for various application areas such as medicine, economy, etc. as shown in Fig. 10. This section highlights some of the works in these domains.

6.1 Medical applications

Timely detection of brain tumor can contribute towards saving the lives of a large number of patients. ELMs are found to be quite helpful for classifying tumor images and exhibit better performance than various deep neural network classifiers. As ELM is not much suitable for big data applications, Deepa and Rajalakshmi [25], proposed ML-ELM for classifying tumor images as tumorous or non-tumorous with higher accuracy. Various methods are available for EEG classification based on Bayes classifier, SVM, etc., but most of these algorithms have restrictions for approximating complex functions. An ML-ELM variant proposed by Ding et al. [27], namely Deep ELM (DELM), combines the best of both the approaches, i.e., ML-ELM and KELM. This approach proved successful for EEG classification as it has less training time and high efficiency. The usage of wearable sensors has become quite crucial due to the emergence of smart health facilities. This poses the requirement of an efficient classification algorithm to recognise humans' actions, which can enable them to lead a healthier lifestyle. Chen et al. [15] proposed an algorithm named S-ELM-KRSL for this purpose, which uses stacked ELM (S-ELM) and KRSL similarity. The results achieved showed higher accuracy as compared to other traditional algorithms. The recognition of EEG signals is an essential technology of Brain-Computer Interface

(BCI), which involves feature extraction and classification. Duan et al. [28] proposed an ML-ELM based classification approach for EEG signals to achieve better performance. She et al. [116] employed hierarchical semi-supervised ELM (HSS-ELM) for efficient motor imagery (MI) task classification. The experimental results showed higher classification accuracy and more generalization with least human intervention as compared to other methods used for motor imagery EEG data. It is crucial to identify brain diseases from magnetic resonance (MR) images early stage to avoid serious problems. A variant of ML-ELM, PGM-ELM [167], has been used efficiently for classifying imbalanced medical data. Fei et al. [29] proposed the usage of ML-KELM-PM for Breast Tumor Diagnosis using ultrasound. The algorithm effectively performs transfer learning required in computer-aided diagnosis. An efficient approach for brain images classification was put forward by Nayak et al. [97], which uses ML-ELM to automate the process of feature extraction. The proposed system resulted in better generalization and more robustness. Ijaz et al. [57] presented a model using random forest as classifier to perform early prediction of cervical cancer. Srinivasu et al. [120] used a computationally efficient AW-HARIS algorithm for automated segmentation of CT scan images to identify abnormalities in the human liver. Mandal et al. [87] proposed a tri-stage wrapper-filter-based feature selection method for saving time and cost in disease detection. Dash et al. [24] presented a hybrid method for blood vessel segmentation to improve the performance of curvelet transform. Srinivasu et al. [121] have proposed an efficient deep learning approach based on MobileNet V2 and LSTM for skin disease classification. Dash et al. [23] put forward a joint model having fast guided and matched filters for enhancing vessel extraction in abnormal retinal images. Kumar et al. [68] have presented a comprehensive survey on various artificial intelligence techniques which can be used to diagnose diseases such as cancer, tuberculosis etc.

6.2 Industrial applications

Wang et al. [136] put forward an efficient crack detection model based on ML-ELM, which does not require iterative tuning for learning its parameters. It exhibits phenomenal performance in terms of training efficiency and model accuracy. Identifying coal quality from remote sensing images is a critical task in coal mining. Le et al. [71] proposed an incremental ML-ELM based algorithm named IM-ELM which could more appropriately determine the number of hidden layers. This led to an efficient classification model, which proved to be better in terms of higher speed and accuracy along with low price. A variant of ML-ELM, EPLS-ML-ELM [124], was used to generate a data-driven prediction model for real blast furnace data. Whether or not to adjust the burden distribution matrix is efficiently determined by EPLS-ML-ELM. Su et al. [125] put forward an approach utilizing ML-ELM, PCA, and wavelet transform named W-PCA-ML-ELM for measuring the permeability index. The proposed method proved to be helpful for better generalization and more stability of the prediction model. Reliable and efficient detection of faults is a crucial task for reducing maintenance cost and avoiding unplanned interruption. Yang et al. [155] proposed an ML-ELM based fault diagnosis scheme for wind turbines. The experimental results showed better accuracy and efficiency as compared to other approaches. A variant of ML-ELM put forward by Su et al. [126], namely EAPSO-ML-ELM, was used for hot metal temperature prediction in the blast furnace. The proposed approach proved to be helpful for better generalization and prediction accuracy. Timely and accurate identification of high-quality coal can reduce environmental pollution and increasing production efficiency. A particular variant of ML-ELM based on inertia weight artificial bee colony (ABC), namely IAM-ELM, was put forward by Mao et al. [89]. It proved to be useful for

coal classification as it showed significantly good speed and accuracy compared to various other existing coal classification methods. An efficient approach named NALO-MELM put forward by Zheng et al. [170] was used for fault diagnosis in rotary machines, and it showed higher accuracy than other methods. Lu et al. [82] applied an improved version of ML-OSELM using the evolutionary approach for Stencil Printing Optimization in real-time. The proposed approach contributed towards an increase in prediction accuracy and printing performance. Ma et al. [85] proposed using HKA-ML-ELM to estimate the remaining useful life of Lithium-ion batteries. The experimental results verified the effectiveness of the proposed approach. He et al. [43] used TR-ML-ELM to monitor coal mining areas. The proposed method resulted in high precision and fast speed. Zhao et al. [168] presented an early fault detection approach for analog circuits based on ML-ELM, which showed higher diagnosis accuracy and faster diagnosis speed. The traditional method of measuring the temperature using disposable thermocouples is inefficient and costly for continuous data procurement. Su et al. [122] proposed using EVFF-MLOSELM to predict hot metal silicon content. The simulation results exhibited better accuracy and generalization performance than other algorithms.Gupta et al. [40] presented an efficient content caching strategy for IoT applications which can aid in traffic management at cloud databases. Khan et al. [65] conducted a systematic literature review on security issues/challenges faced by software vendors' organizations. Rani et al. [106] put forward an adapted fault tolerant approach wireless sensor network routing in industrial applications.

6.3 Academic applications

Automatic handwritten digit recognition is a task that has captured a lot of interest in academics and commerce. A novel classification approach based on ELM was put forward by Noushahr et al. [98], which exhibits better generalization and fast speed for learning. The authors have also proposed Multilayer Ensemble ELM (ML-EELM), based on the combination of concepts of CNN, ensemble models, and ELM. This further increases the accuracy of the classifier.

6.4 Security applications

Wang et al. [138] proposed an ML-ELM based approach, which works on the encrypted databases directly and returns the output after converting the multi-class classification problem at hand into the corresponding binary classification problem. This method results in secure and accurate image classification while eliminating the need for decryption. Yang et al. [154] presented a binary decision diagram (BDD) based DL algorithm, i.e., BDD-ML-ELM, for privacy protection in the finger-vein recognition systems. The proposed approach achieved adequate security and prediction accuracy. Traditional security measures, including passwords, tokens, etc., have some shortcomings which can lead to severe problems. Thus the authentication based on keystroke dynamics became more critical. Zhao et al. [169] presented an efficient keystroke dynamics identification approach using ML-ELM, reducing human interaction and manual feature extraction. The proposed method resulted in high accuracy and less time involvement which further justifies its usage for real-time applications. Panigrahi et al. [102] proposed a host-based intrusion detection system using a C4.5-based detector with Consolidated Tree Construction algorithm which works efficiently with class-imbalanced data as well. Panigrahi et al. [101] analyzed the current literature in the field of network intrusion detection, highlighting the various important parameters. Privacy protection is a requirement for applications dealing

with confidential images. Nevertheless, performing decryption before the classification of images increases the computational complexity to a large extent. ML-ELM can be used for efficiently classifying the dataset of encrypted images without decrypting the encrypted image.

6.5 Transportation applications

Intelligent video surveillance (IVS) systems, involving the movement activity of humans, are quite helpful for various applications, including accident detection, patient monitoring, etc. Yu et al. [159] put forward an approach based on ML-ELM and object principal trajectory, namely PTM-ELM. The usage of ML-ELM makes it feasible for frame predictors to adapt to fast-changing features by holding necessary information. The proposed method resulted in excellent image quality and efficient generalization. Lee et al. [74] suggested the usage of ML-RBF-ELM for Terrain Referenced Navigation, which aids in real-time navigation operation. The experimental results justified its use for small unmanned aerial vehicles (UAVs), which don't have huge memory space. Hernandez et al. [46] utilized ML-FELM to classify and transport objects using indoor UAVs. The results showed higher accuracy of ML-FELM for image classification.

6.6 IoT applications

DFL is widely used in the field of wireless localization. An ML-ELM variant, namely MP-ELM [166], proved to be helpful for implementing faster and more accurate DFL. MP-ELM has the advantage of less time and labor consumption to automatically learn important information from the links. Moreover, the validity of this method in DFL has been evaluated in indoor and outdoor environments.

6.7 Military and civil applications

Radar emitter identification mainly finds applications in military and civil fields. Cao et al. [9] proposed an H-ELM based method for identifying radar emitter signals that are quite efficient and can be deployed in various real-life applications.

6.8 Other interdisciplinary application domains

Modality refers to a particular way of experiencing something, and a multimodal research problem is one that involves multiple modalities. Weakly paired multimodal data refers to the situation when the real-world data obtained from various sensors has every modality being partitioned into various groups. Wen et al. [142] proposed ML-ELM based framework to capture the non-linear transformations related to each modality in a simple yet effective manner. Online Incremental Tracking is a tracking technique adopted to learn and adapt a representation that can be used to reflect the changes in the target's appearance. Modeling such variation is a critical task that requires efficient algorithms. H-ELM [129] has been used successfully used for incremental tracking. Clustering is an important task that can solve problems in many domains, including text classification, market research, etc. Wu et al. [148] proposed an evolutionary ML-ELM based approach for data clustering problems. The experimental results revealed better accuracy and robustness of the method. A summary of various ML-ELM applications is listed in Table 10.

Table 10 Summary of ML-ELM applications

Application category	References
Medical Applications	EEG classification [27, 28], Human action recognition [15], MI task classification [116], Imbalanced medical data classification [167], Retinal image enhancement [23], Disease diagnosis [25, 29, 57, 68, 87, 97, 121], Automated segmentation of CT images [120], Blood vessel segmentation [24]
Industrial Applications	Concrete crack detection [136], Coal Industry [71, 89], Blast furnace [122, 124–126], Fault diagnosis [155, 168, 170], Stencil Printing Optimization [82], Lithium-ion batteries useful life prediction [85], Content caching [40], Software security [65], Smart Grid [106]
Academic Applications	Handwritten digit recognition [98]
Security Applications	Encrypted image classification [138], Privacy protection [154], Keystroke dynamics identification [169], Intrusion detection [101, 102]
Transportation Applications	IVS systems [159], Terrain Referenced Navigation [74], UAVs [46]
IoT Applications	Device-Free Localization [166]
Military and Civil Applications	Radar emitter signal identification [9]
Other interdisciplinary application domains	Multimodal research problems [142], Incremental tracking [129], Data clustering [148]

7 Conclusion and future enhancement

This paper reviews architecture of ML-ELM, emphasizing its variants and applications in various domains. The significant advantages of ML-ELM include less training time, random feature mapping, and higher learning accuracy, which accelerate the development of DL. The topology of existing ML-ELM variants till date is also described in the paper which constitutes variants for better feature learning, handling outliers or noise, optimizing weights and bias, reducing multicollinearity and reducing overfitting. Also, a comparative analysis has been performed for ML-ELM with other ML and DL techniques. The latter includes the shortcomings of prevailing state-of-the-art DL architectures such as ANN, CNN, RNN, and LSTM and how ML-ELM can be utilized to handle all such limitations. The open issues in ML-ELM have also been discussed in the paper. Since the variants of ML-ELM have shown promising results in various applications, this learning algorithm can be explored further for various real-life applications such as intrusion detection, fault diagnosis, coal mine area monitoring, etc. The future study on ML-ELM may include the following:

i. The applications of ML-ELM in parallel and distributed computing are open. Also, its effectiveness for big data applications can be investigated further.
ii. More applications of ML-ELM can be studied to verify its generalization capability on massive datasets having noise.
iii. The variance of hidden layer weights remains a topic for future research to understand the in-depth functionality of ELM and ML-ELM.

Data availability Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

Declarations

Conflict of Interests The authors declare no conflict of interest.

References

1. Allen EJ, St-Yves G, Wu Y et al (2022) A massive 7t fmri dataset to bridge cognitive neuroscience and artificial intelligence. Nat Neurosci 25(1):116–126
2. An L, Bhanu B (2012) Image super-resolution by extreme learning machine. In: 2012 19th IEEE international conference on image processing. IEEE, Orlando, pp 2209–2212
3. Andrushia AD, Thangarajan R (2020) Rts-elm: an approach for saliency-directed image segmentation with ripplet transform. Pattern Anal Applic 23(1):385–397
4. Antal B, Hajdu A (2014) UCI machine learning repository. https://archive.ics.uci.edu/ml/datasets/Diabetic+Retinopathy+Debrecen+Data+Set. Accessed 18 Feb 2022
5. Bal PR, Kumar S (2020) Wr-elm: Weighted regularization extreme learning machine for imbalance learning in software fault prediction. IEEE Trans Reliab 69:1355–1375
6. Baradarani A, Wu QJ, Ahmadi M (2013) An efficient illumination invariant face recognition framework via illumination enhancement and dd-dtcwt filtering. Pattern Recogn 46(1):57–72
7. Birkl C (2017) Oxford battery degradation dataset 1. University of Oxford. https://doi.org/10.5287/bodleian:KO2kdmYGg
8. Cambria E, Huang GB, Kasun LLC et al (2013) Extreme learning machines [trends & controversies]. IEEE Intell Syst 28(6):30–59
9. Cao R, Cao J, Mei JP et al (2019b) Radar emitter identification with bispectrum and hierarchical extreme learning machine. Multimed Tools Appl 78(20):28,953–28,970
10. Cao J, Hao J, Lai X et al (2016) Ensemble extreme learning machine and sparse representation classification. J Frankl Inst 353(17):4526–4541
11. Cao J, Lin Z, Huang GB et al (2012) Voting based extreme learning machine. Inf Sci 185(1):66–77
12. Cao F, Yang Z, Ren J et al (2019a) Local block multilayer sparse extreme learning machine for effective feature extraction and classification of hyperspectral images. IEEE Trans Geosci Remote Sens 57(8):5580–5594
13. Chang NB, Han M, Yao W et al (2010) Change detection of land use and land cover in an urban region with spot-5 images and partial lanczos extreme learning machine. J Appl Remote Sens 4(1):043,551
14. Chapelle O, Schölkopf B, Zien A (2006) The geometric basis of semi-supervised learning. In: Semi-Supervised Learning. MIT Press, pp 217–235
15. Chen M, Li Y, Luo X et al (2018) A novel human activity recognition scheme for smart health using multilayer extreme learning machine. IEEE Internet Things J 6(2):1410–1418
16. Chen F, Ou T (2011) Sales forecasting system based on gray extreme learning machine with taguchi method in retail industry. Expert Syst Appl 38(3):1336–1345
17. Chen B, Xing L, Xu B et al (2017) Kernel risk-sensitive loss: definition, properties and application to robust adaptive filtering. IEEE Trans Signal Process 65(11):2888–2901
18. Cheng G, Han J, Lu X (2017) Remote sensing image scene classification: benchmark and state of the art. Proc IEEE 105(10):1865–1883
19. Chua TS, Tang J, Hong R et al (2009) Nus-wide: a real-world web image database from national university of singapore. In: Proceedings of the ACM international conference on image and video retrieval. pp 1–9
20. Cole R, Fanty M (1994) UCI machine learning repository. https://archive.ics.uci.edu/ml/datasets/ISOLET. Accessed 2 March 2022
21. Dai H, Cao J, Wang T et al (2019) Multilayer one-class extreme learning machine. Neural Netw 115:11–22
22. Dailey M, Cottrell G, Reilly J (2001) California facial expressions, cafe. Unpublished digital images University of California. Computer Science and Engineering Department, San Diego
23. Dash S, Verma S, Bevinakoppa S et al (2022) Guidance image-based enhanced matched filter with modified thresholding for blood vessel extraction. Symmetry 14(2):194

24. Dash S, Verma S, Khan M et al (2021) A hybrid method to enhance thick and thin vessels for blood vessel segmentation. Diagnostics 11(11):2017

25. Deepa M, Rajalakshmi M (2016) A fuzzy clustering approach based on multi-layer extreme learning machine for brain tumor detection and classification. International Journal of Advanced Engineering Technology

26. Deng C, Huang G, Xu J et al (2015) Extreme learning machines: new trends and applications. Sci China Inf Sci 58(2):1–16

27. Ding S, Zhang N, Xu X et al (2015) Deep extreme learning machine and its application in eeg classification. Math Probl Eng, 2015

28. Duan L, Bao M, Miao J et al (2016) Classification based on multilayer extreme learning machine for motor imagery task from eeg signals. Procedia Comput Sci 88:176–184

29. Fei X, Zhou W, Shen L et al (2019) Ultrasound-based diagnosis of breast tumor with parameter transfer multilayer kernel extreme learning machine. In: 2019 41st annual international conference of the IEEE engineering in medicine and biology society (EMBC). IEEE, p 933-936, Berlin

30. Fei-Fei L, Fergus R, Perona P (2004) Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories. In: 2004 conference on computer vision and pattern recognition workshop. IEEE, Washington, pp 178–178

31. Filannino M (2011) Dbworld e-mail classification using a very small corpus. Univ Manch 86:6–12

32. Ghosh S, Mukherjee H, Obaidullah SM et al (2018) A survey on extreme learning machine and evolution of its variants. In: International Conference on Recent Trends in Image Processing and Pattern Recognition. Springer, Singapore, pp 572–583

33. Goebel K, Saha B (2010) Dashlink-li-ion battery aging datasets. DAWN MCINTOSH

34. Good RP, Kost D, Cherry GA (2010) Introducing a unified pca algorithm for model size reduction. IEEE Trans Semicond Manuf 23(2):201–209

35. Goodfellow I, Pouget-Abadie J, Mirza M et al (2014) Generative adversarial nets. In: Advances in neural information processing systems. p 2672–2680. Curran Associates, Inc., Red Hook

36. Graham DB, Allinson NM (1998) Characterising virtual eigensignatures for general purpose face recognition. In: Face Recognition. Springer, Berlin, pp 446–456

37. Griffin G, Holub A, Perona P (2007) Caltech-256 object category dataset. California Institute of Technology. https://resolver.caltech.edu/CaltechAUTHORS:CNS-TR-2007-001

38. Güldener U, Münsterkötter M, Kastenmüller G et al (2005) Cygd: the comprehensive yeast genome database. Nucleic Acids Res 33(suppl_1):D364–D368

39. Guo T, Zhang L, Tan X (2017) Neuron pruning-based discriminative extreme learning machine for pattern classification. Cogn Comput 9(4):581–595

40. Gupta D, Rani S, Ahmed SH et al (2021) Edge caching based on collaborative filtering for heterogeneous icn-iot applications. Sensors 21(16):5491

41. Han F, Jiang J, Ling QH et al (2019) A survey on metaheuristic optimization for random single-hidden layer feedforward neural network. Neurocomputing 335:261–273

42. Hassanpour A, Moradikia M, Adeli H et al (2019) A novel end-to-end deep learning scheme for classifying multi-class motor imagery electroencephalography signals. Expert Syst 36(6):e12494

43. He D, Le BT, Xiao D et al (2019) Coal mine area monitoring method by machine learning and multispectral remote sensing images. Infrared Phys Technol 103:103070

44. He Q, Shang T, Zhuang F et al (2013) Parallel extreme learning machine for regression based on mapreduce. Neurocomputing 102:52–58

45. Hemanth JD, Anitha J, Ane BK (2017) Fusion of artificial neural networks for learning capability enhancement: Application to medical image classification. Expert Syst 34(6):e12225

46. Hernandez-Hernandez RA, Martinez-Hernandez U, Rubio-solis A (2020) Multilayer fuzzy extreme learning machine applied to active classification and transport of objects using an unmanned aerial vehicle. In: 2020 IEEE International conference on fuzzy systems (FUZZ-IEEE). IEEE, Glasgow, pp 1–8

47. Huang GB (2012) Extreme learning machine: learning without iterative tuning School of Electrical and Electronic Engineering. Nanyang Technological University, Singapore

48. Huang GB, Bai Z, Kasun LLC et al (2015b) Local receptive fields based extreme learning machine. IEEE Comput Intell Mag 10(2):18–29

49. Huang GB, Chen L (2007) Convex incremental extreme learning machine. Neurocomputing 70(16-18):3056–3062

50. Huang GB, Chen YQ, Babri HA (2000) Classification ability of single hidden layer feedforward neural networks. IEEE Trans Neural Netw 11(3):799–801

51. Huang GB, Chen L, Siew CK et al (2006a) Universal approximation using incremental constructive feedforward networks with random hidden nodes. IEEE Trans Neural Netw 17(4):879–892

52. Huang G, Huang GB, Song S et al (2015a) Trends in extreme learning machines: a review. Neural Netw 61:32–48
53. Huang Z, Lei D, Huang D et al (2019) Boundary moving least square method for 2d elasticity problems. Eng Anal Bound Elem 106:505–512
54. Huang GB, Wang DH, Lan Y (2011) Extreme learning machines: a survey. International journal of machine learning and cybernetics 2(2):107–122
55. Huang GB, Zhu QY, Siew CK (2006b) Extreme learning machine: theory and applications. Neurocomputing 70(1-3):489–501
56. Hull JJ (1994) A database for handwritten text recognition research. IEEE Trans Patt Anal Mach Intell 16(5):550–554
57. Ijaz MF, Attique M, Son Y (2020) Data-driven cervical cancer prediction model with outlier detection and over-sampling methods. Sensors 20(10):2809
58. Jahromi AN, Hashemi S, Dehghantanha A et al (2020) An improved two-hidden-layer extreme learning machine for malware hunting. Comput Secur 89:101655
59. Jia X, Li X, Du H et al (2016) Local invariance representation learning algorithm with multi-layer extreme learning machine. In: International Conference on Neural Information Processing. Springer, Kyoto, pp 505–513
60. Jia X, Li X, Jin Y et al (2019) Region-enhanced multi-layer extreme learning machine. Cogn Comput 11(1):101–109
61. Jiang X, Yan T, Xu Q et al (2018) Remote sensing image scene classification based on densely connected multilayer kernel elm. In: 2018 Australian & New Zealand Control Conference (ANZCC). IEEE, Melbourne, pp 81–86
62. Johnson WB, Lindenstrauss J (1984) Extensions of lipschitz mappings into a hilbert space. Contemp Math 26(189-206):1
63. Johnson KA et al (2001) The whole brain atlas. Harvard Medical School
64. Kasun LLC, Zhou H, Huang GB et al (2013) Representational learning with extreme learning machine for big data. IEEE Intell Syst 28(6):31–34
65. Khan AW, Khan MU, Khan JA et al (2021) Analyzing and evaluating critical challenges and practices for software vendor organizations to secure big data on cloud computing: an ahp-based systematic approach. IEEE Access 9:107309–107332
66. Klahr D, Siegler RS (1978) The representation of children's knowledge. In: Advances in child development and behavior, vol 12. Elsevier, New York, pp 61–116
67. Krizhevsky A, Hinton G et al (2009) Learning multiple layers of features from tiny images. Technical report, University of Toronto
68. Kumar Y, Koul A, Singla R et al (2022) Artificial intelligence in disease diagnosis: a systematic literature review, synthesizing framework and future research agenda. J Ambient Intell Humanized Comput 13:1–28
69. Kurgan LA, Cios KJ, Tadeusiewicz R et al (2001) Knowledge discovery approach to automated cardiac spect diagnosis. Artif Intell Med 23(2):149–169
70. Kushmerick N (1999) Learning to remove internet advertisements. In: Proceedings of the third annual conference on Autonomous Agents. pp 175–181. Association for Computing Machinery, New York
71. Le BT, Xiao D, Mao Y et al (2019) Coal quality exploration technology based on an incremental multilayer extreme learning machine and remote sensing images. IEEE Trans Geosci Remote Sens 57(7):4192–4201
72. LeCun Y, Bottou L, Bengio Y et al (1998) Gradient-based learning applied to document recognition. Proc IEEE 86(11):2278–2324
73. LeCun Y, Huang FJ, Bottou L (2004) Learning methods for generic object recognition with invariance to pose and lighting. In: Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004. IEEE, Washington, pp II–104
74. Lee J, Sung C, Oh J (2019) Terrain referenced navigation using a multilayer radial basis function-based extreme learning machine. Int J Aerosp Eng 2019:13–23
75. Lekamalage CKL, Song K, Huang GB et al (2017) Multi layer multi objective extreme learning machine. In: 2017 IEEE International Conference on Image Processing (ICIP). IEEE, p 1297–1301, Beijing
76. Li L, Sun R, Cai S et al (2019) A review of improved extreme learning machine methods for data stream classification. Multimed Tools Appl 78(23):33375–33400
77. Liangjun C, Honeine P, Hua Q et al (2018) Correntropy-based robust multilayer extreme learning machines. Pattern Recogn 84:357–370
78. Lillicrap TP, Santoro A, Marris L et al (2020) Backpropagation and the brain. Nat Rev Neurosci 21:335–346

79. Lim TS (1997) UCI machine learning repository. https://archive.ics.uci.edu/ml/datasets/Contraceptive+Method+Choice. Accessed 5 March 2022

80. Liu Y, Huangfu W, Zhang H et al (2019) An efficient stochastic gradient descent algorithm to maximize the coverage of cellular networks. IEEE Trans Wirel Commun 18(7):3424–3436

81. Liu W, Wang Z, Liu X et al (2017) A survey of deep neural network architectures and their applications. Neurocomputing 234:11–26

82. Lu H, Wang H, Yoon SW et al (2019) Real-time stencil printing optimization using a hybrid multi-layer online sequential extreme learning and evolutionary search approach. IEEE Transactions on Components. Packag Manuf Technol 9(12):2490–2498

83. Luo X, Li Y, Wang W et al (2020) A robust multilayer extreme learning machine using kernel risk-sensitive loss criterion. Int J Mach Learn Cybern 11(1):197–216

84. Luong NC, Hoang DT, Gong S et al (2019) Applications of deep reinforcement learning in communications and networking: a survey. IEEE Commun Surv Tutorials 21(4):3133–3174

85. Ma Y, Shen D, Wu L et al (2019) The remaining useful life estimation of lithium-ion batteries based on the hka-ml-elm algorithm. Int J Electrochem Sci 14:7737–7757

86. Madessa AH, Dong J, Gan Y et al (2020) A deep learning approach for specular highlight removal from transmissive materials

87. Mandal M, Singh PK, Ijaz MF et al (2021) A tri-stage wrapper-filter feature selection framework for disease classification. Sensors 21(16):5571

88. Mao Q, Hu F, Hao Q (2018) Deep learning for intelligent wireless networks: a comprehensive survey. IEEE Commun Surv Tutorials 20(4):2595–2621

89. Mao Y, Le BT, Xiao D et al (2019) Coal classification method based on visible-infrared spectroscopy and an improved multilayer extreme learning machine. Opt Laser Technol 114:10–15

90. Maulik U, Chakraborty D (2017) Remote sensing image classification: a survey of support-vector-machine-based advanced techniques. IEEE Geosci Remote Sens Mag 5(1):33–52

91. Minhas R, Mohammed AA, Wu QJ (2011) Incremental learning in human action recognition based on snippets. IEEE Trans Circuits Syst Video Technol 22(11):1529–1541

92. Mirza B, Kok S, Dong F (2016) Multi-layer online sequential extreme learning machine for image classification. In: Proceedings of ELM-2015 Volume 1. Springer, Hangzhou, pp 39–49

93. Mitchell T (1999) UCI machine learning repository. https://archive.ics.uci.edu/ml/datasets/Twenty+Newsgroups. Accessed 15 Jan 2022

94. Moore AW, Zuev D (2005) Internet traffic classification using bayesian analysis techniques. In: Proceedings of the 2005 ACM SIGMETRICS international conference on Measurement and modeling of computer systems. Association for Computing Machinery, New York, pp 50–60

95. Mukherjee H, Obaidullah SM, Santosh K et al (2018) Line spectral frequency-based features and extreme learning machine for voice activity detection from audio signal. Int J Speech Technol 21(4):753–760

96. Nakai K (1996) UCI machine learning repository. https://archive.ics.uci.edu/ml/datasets/Yeast. Accessed 25 Jan 2022

97. Nayak DR, Das D, Dash R et al (2020) Deep extreme learning machine with leaky rectified linear unit for multiclass classification of pathological brain images. Multimed Tools Appl 79(21):15381–15396

98. Noushahr HG, Ahmadi S, Casey A (2015) Fast handwritten digit recognition with multilayer ensemble extreme learning machine. In: International Conference on Innovative Techniques and Applications of Artificial Intelligence. Springer, Cambridge, pp 77–89

99. Orimoloye LO, Sung MC, Ma T et al (2020) Comparing the effectiveness of deep feedforward neural networks and shallow architectures for predicting stock price indices. Expert Syst Appl 139:112828

100. Pace K (1997). http://www.dcc.fc.up.pt/ltorgo/Regression/cal_housing.html. Accessed 13 Jan 2022

101. Panigrahi R, Borah S, Bhoi AK et al (2021a) Performance assessment of supervised classifiers for designing intrusion detection systems: a comprehensive review and recommendations for future research. Mathematics 9(6):690

102. Panigrahi R, Borah S, Bhoi AK et al (2021b) A consolidated decision tree-based intrusion detection system for binary and multiclass imbalanced datasets. Mathematics 9(7):751

103. Parkavi RM, Shanthi M, Bhuvaneshwari M et al (2017) Recent trends in elm and mlelm: a review. Adv Sci Technol Eng Syst J 2(1):69–75

104. Poggio T, Liao Q, Banburski A (2020) Complexity control by gradient descent in deep networks. Nat Commun 11(1):1–5

105. Raghuwanshi BS, Shukla S (2018) Class-specific extreme learning machine for handling binary class imbalance problem. Neural Netw 105:206–217

106. Rani S, Koundal D, Ijaz MF et al (2021) An optimized framework for wsn routing in the context of industry 4.0. Sensors 21(19):6474

107. Rezaei Ravari M, Eftekhari M, Saberi Movahed F (2020) Ml-ck-elm: an efficient multi-layer extreme learning machine using combined kernels for multi-label classification. Sci Iran 27(6):3005–3018
108. Rifai S, Vincent P, Muller X et al (2011) Contractive auto-encoders: explicit invariance during feature extraction. In: Proceedings of the 28th international conference on international conference on machine learning. Omnipress, Madison, pp 833–840
109. Rifkin R, Yeo G, Poggio T et al (2003) Regularized least-squares classification. Nato Sci Ser Sub Ser III Comput Syst Sci 190:131–154
110. Roul RK (2018) Detecting spam web pages using multilayer extreme learning machine. Int J Big Data Intell 5(1-2):49–61
111. Roul RK, Asthana SR, Kumar G (2017) Study on suitability and importance of multilayer extreme learning machine for classification of text data. Soft Comput 21(15):4239–4256
112. Samaria FS, Harter AC (1994) Parameterisation of a stochastic model for human face identification. In: Proceedings of 1994 IEEE workshop on applications of computer vision. IEEE, Sarasota, pp 138–142
113. Sandberg IW (1994) General structures for classification. IEEE Trans Circ Syst I Fundam Theory Appl 41(5):372–376
114. Schmidhuber J (2015) Deep learning in neural networks: an overview. Neural Netw 61:85–117
115. Scholkopf B, Smola AJ (2001) Learning with kernels: support vector machines, regularization, optimization, and beyond. MIT Press, Cambridge
116. She Q, Hu B, Luo Z et al (2019) A hierarchical semi-supervised extreme learning machine method for eeg recognition. Med Biol Eng Comput 57(1):147–157
117. Shi Y, Eberhart R (1998) A modified particle swarm optimizer. In: 1998 IEEE international conference on evolutionary computation proceedings. IEEE world congress on computational intelligence (Cat. No. 98TH8360). IEEE, Anchorage, pp 69–73
118. Slate DJ (1991) UCI machine learning repository. https://archive.ics.uci.edu/ml/datasets/letter+recognition. Accessed 20 May 2022
119. Smith LI (2002) A tutorial on principal components analysis. Tech. rep.
120. Srinivasu PN, Ahmed S, Alhumam A et al (2021a) An aw-haris based automated segmentation of human liver using ct images. Comput Mater Contin 69(3):3303–3319
121. Srinivasu PN, SivaSai JG, Ijaz MF et al (2021b) Classification of skin disease using deep learning neural networks with mobilenet v2 and lstm. Sensors 21(8):2852
122. Su X, Sun S, Zhang S et al (2020) Improved multi-layer online sequential extreme learning machine and its application for hot metal silicon content. J Frankl Inst 357(17):12588–12608
123. Su X, Yin Y, Zhang S (2016) Prediction model of improved multi-layer extreme learning machine for permeability index of blast furnace. Control Theory Appl 33(12):1674–1684
124. Su X, Zhang S, Yin Y et al (2018a) Data-driven prediction model for adjusting burden distribution matrix of blast furnace based on improved multilayer extreme learning machine. Soft Comput 22(11):3575–3589
125. Su X, Zhang S, Yin Y et al (2018b) Prediction model of permeability index for blast furnace based on the improved multi-layer extreme learning machine and wavelet transform. J Frankl Inst 355(4):1663–1691
126. Su X, Zhang S, Yin Y et al (2019) Prediction model of hot metal temperature for blast furnace based on improved multi-layer extreme learning machine. Int J Mach Learn Cybern 10(10):2739–2752
127. Tang J, Deng C, Huang GB (2015) Extreme learning machine for multilayer perceptron. IEEE Trans Neural Netw Learn Syst 27(4):809–821
128. Tang J, Deng C, Huang GB et al (2014a) A fast learning algorithm for multi-layer extreme learning machine. In: 2014 IEEE International conference on image processing. IEEE, Paris, pp 175–178
129. Tang J, Deng C, Huang GB et al (2014b) Compressed-domain ship detection on spaceborne optical image using deep neural network and extreme learning machine. IEEE Trans Geosci Remote Sens 53(3):1174–1185
130. Tang L, Lu Y (2020) Study of the grey verhulst model based on the weighted least square method. Phys A Stat Mech Appl 545:123615
131. Taskar B, Guestrin C, Koller D (2003) Max-margin markov networks. Adv Neural Inf Process Syst 16:22–29
132. Tj Yu, Xf Yan (2020) Robust multi-layer extreme learning machine using bias-variance tradeoff. J Cent South Univ 27(12):3744–3753
133. Venkata Ramana B, Prasad Babu S, Venkateswarlu NB (2012) UCI machine learning repository. https://archive.ics.uci.edu/ml/datasets/ILPD+(Indian+Liver+Patient+Dataset). Accessed 28 May 2022
134. Vong CM, Chen C, Wong PK (2018) Empirical kernel map-based multilayer extreme learning machines for representation learning. Neurocomputing 310:265–276

135. Wang X, Han Y, Leung VC et al (2020) Convergence of edge computing and deep learning: a comprehensive survey. IEEE Commun Surv Tutorials 22(2):869–904
136. Wang B, Li Y, Zhao W et al (2019) Effective crack damage detection using multilayer sparse feature representation and incremental extreme learning machine. Appl Sci 9(3):614
137. Wang J, Lu S, Wang SH et al (2021) A review on extreme learning machine. Multimed Tools Appl 81:41611–41660
138. Wang W, Vong CM, Yang Y et al (2017a) Encrypted image classification based on multilayer extreme learning machine. Multidim Syst Signal Process 28(3):851–865
139. Wang Y, Wang A, Ai Q et al (2017b) A novel artificial bee colony optimization strategy-based extreme learning machine algorithm. Prog Artif Intell 6(1):41–52
140. Waugh S (1995) UCI machine learning repository. https://archive.ics.uci.edu/ml/datasets/abalone. Accessed 7 Feb 2022
141. Weisstein EW (2002) Moore-penrose matrix inverse. https://mathworldwolframcom/. Accessed 11 Dec 2021
142. Wen X, Liu H, Yan G et al (2018) Weakly paired multimodal fusion using multilayer extreme learning machine. Soft Comput 22(11):3533–3544
143. Whittington JC, Bogacz R (2019) Theories of error back-propagation in the brain. Trends Cogn Sci 23(3):235–250
144. Widodo A, Yang BS (2007) Support vector machine in machine condition monitoring and fault diagnosis. Mech Syst Signal Process 21(6):2560–2574
145. Wolberg W, Mangasarian O, Coleman T et al (1990) Pattern recognition via linear programming: theory and application to medical diagnosis. In: Large-scale numerical optimization. SIAM Publications, Citeseer, Madison, pp 22–30
146. Wong CM, Vong CM, Wong PK et al (2016) Kernel-based multilayer extreme learning machines for representation learning. IEEE Trans Neural Netw Learn Syst 29(3):757–762
147. Wu D, Qu Z, Guo F et al (2019) Multilayer incremental hybrid cost-sensitive extreme learning machine with multiple hidden output matrix and subnetwork hidden nodes. IEEE Access 7:118422–118434
148. Wu X, Zhou T, Yi K et al (2021) An evolutionary multi-layer extreme learning machine for data clustering problems. In: 2021 40th chinese control conference (CCC). IEEE, Shanghai, pp 1978–1983
149. Xia R, Chen Y, Feng Y (2020) A method to measure thermal conductivity of vacuum insulation panel using enhanced extreme learning machine model. J Therm Sci 29:623–631
150. Xu Y, Dai Y, Dong ZY et al (2013) Extreme learning machine-based predictor for real-time frequency stability assessment of electric power systems. Neural Comput Applic 22(3-4):501–508
151. Xu X, Shan D, Li S et al (2019) Multi-label learning method based on ml-rbf and laplacian elm. Neurocomputing 331:213–219
152. Yaghoubi S, Fainekos G (2019) Worst-case satisfaction of stl specifications using feedforward neural network controllers: a lagrange multipliers approach. ACM Trans Embed Comput Syst (TECS) 18(5s):1–20
153. Yang J, Cao J, Wang T et al (2020) Regularized correntropy criterion based semi-supervised elm. Neural Netw 122:117–129
154. Yang W, Wang S, Hu J et al (2019) Securing deep learning based edge finger vein biometrics with binary decision diagram. IEEE Trans Ind Inform 15(7):4244–4253
155. Yang ZX, Wang XB, Zhong JH (2016) Representational learning for fault diagnosis of wind turbine equipment: a multi-layered extreme learning machines approach. Energies 9(6):379
156. Yang Y, Wu QJ (2015) Multilayer extreme learning machine with subnetwork nodes for representation learning. IEEE Trans Cybern 46(11):2570–2583
157. Yang J, Xie S, Yoon S et al (2013) Fingerprint matching based on extreme learning machine. Neural Comput Applic 22(3-4):435–445
158. Yu W, Liu T, Valdez R et al (2010) Application of support vector machine modeling for prediction of common diseases: the case of diabetes and pre-diabetes. BMC Med Inform Decis Mak 10(1):16
159. Yu H, Wang J, Sun X (2019) Surveillance video online prediction using multilayer elm with object principal trajectory. SIViP 13(6):1243–1251
160. Zhang N, Ding S, Shi Z (2016a) Denoising laplacian multi-layer extreme learning machine. Neurocomputing 171:1066–1074
161. Zhang N, Ding S, Zhang J (2016b) Multi layer elm-rbf for multi-label learning. Appl Soft Comput 43:535–545
162. Zhang W, Ji H (2013) Fuzzy extreme learning machine for classification. Electron Lett 49(7):448–450
163. Zhang J, Li Y, Xiao W et al (2020) Non-iterative and fast deep learning: Multilayer extreme learning machines. J Frankl Inst 357(13):8925–8955

164. Zhang C, Patras P, Haddadi H (2019a) Deep learning in mobile and wireless networking: A survey. IEEE Commun Surv Tutorials 21(3):2224–2287

165. Zhang Y, Wu J, Zhou C et al (2017) Instance cloned extreme learning machine. Pattern Recogn 68:52–65

166. Zhang J, Xiao W, Li Y et al (2019b) Multilayer probability extreme learning machine for device-free localization. Neurocomputing 396:383–393

167. Zhang L, Yang H, Jiang Z (2018) Imbalanced biomedical data classification using self-adaptive multilayer elm combined with dynamic gan. Biomed Eng Online 17(1):181

168. Zhao G, Liu Y, Zhou J et al (2020a) Analog circuit incipient fault diagnosis from raw signals using multi-layer extreme learning machine. In: 2020 11th International Conference on Prognostics and System Health Management (PHM-2020 Jinan). IEEE, Jinan, pp 315–321

169. Zhao G, Wu Z, Gao Y et al (2020b) Multi-layer extreme learning machine-based keystroke dynamics identification for intelligent keyboard. IEEE Sensors J 21(2):2324–2333

170. Zheng L, Wang Z, Zhao Z et al (2019) Research of bearing fault diagnosis method based on multi-layer extreme learning machine optimized by novel ant lion algorithm. IEEE Access 7:89845–89856

171. Zhou H, Huang GB, Lin Z et al (2014) Stacked extreme learning machines. IEEE Trans Cybern 45(9):2013–2025

172. Zwitter M, Soklic M (1988) UCI machine learning repository. https://archive.ics.uci.edu/ml/datasets/Breast+Cancer. Accessed 10 Feb 2022