2023 International Conference on Computational Intelligence and Sustainable Engineering Solutions (CISES) | 979-8-3503-2391-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/CISES58720.2023.10183613

Extreme Learning Machine: A Comprehensive Survey of Theories & Algorithms

Harshal Patil
Department of CS & IT
Symbiosis Institute of Technology
Symbiosis International (Deemed) University
Pune, Maharashtra, India
harshal.patil@sitpune.edu.in

Kanhaiya Sharma
Department of CS & IT
Symbiosis Institute of Technology
Symbiosis International (Deemed) University
Pune, Maharashtra, India
kanhaiya.sharma@sitpune.edu.in

***Abstract*— The Extreme Learning Machine (ELM), a quick and effective approach for training single-hidden-layer feedforward neural networks, is thoroughly reviewed in this paper. The objective is to provide a systematic review and analysis of the various theories, algorithms, and applications of ELM, which have not been adequately covered in the literature. The study examines the theoretical foundations of ELM, provides an overview of various ELM algorithms, analyzes the applications of ELM in different domains, and identifies future research directions and challenges in the field of ELM. The findings suggest that ELM is a promising and effective learning algorithm with many benefits as improved learning speed, good generalisation performance, and low complexity. The implications of this research for practice and research are significant. However, more empirical studies are necessary to evaluate its performance in real-world applications, and further research is required to assess its effectiveness in different domains. Future work should focus on exploring the potential of ELM in deep learning, enhancing the interpretability of ELM models, and developing new algorithms that can improve the learning efficiency of ELM. Overall, this survey provides a valuable reference for researchers and practitioners to better understand the potential and limitations of the ELM algorithm.**

Keywords— ***ANN, ELM, Machine Learning, Deep learning, Random forest, SVM***

I. INTRODUCTION

Machine learning is a rapidly evolving field that has revolutionized various industries, including finance, healthcare, and education. Over the course of time, a plethora of machine learning techniques have been invented to address various issues. The Extreme Learning Machine (ELM) approach is a powerful tool for training single-hidden-layer feedforward neural networks(SLFNs) quickly and effectively. [1]. ELM has gained widespread attention due to its superior performance and simplicity compared to other traditional machine learning algorithms [2].

Although ELM has been broadly studied and utilised in various applications, there is a lack of a comprehensive survey that systematically reviews and analyzes the various theories, algorithms, and applications of ELM [3]. This has resulted in a fragmented understanding of the strengths and limitations of ELM and a lack of clear guidelines for selecting appropriate ELM models for different applications [4].

Here we provide a comprehensive survey of ELM that covers various aspects of ELM, including theories, algorithms, and applications. Specifically, we aim to achieve the following contributions: (1) provide a systematic review of the theoretical foundations of ELM, (2) present an overview of various ELM algorithms, including their strengths and limitations, (3) provide a detailed analysis of the applications of ELM in various domains, and (4) identify the future research directions and challenges in the field of ELM.

II. THEORETICAL FOUNDATION

A. Machine Learning and Deep Learning

Machine learning (ML) is a subclass of Artificial Intelligence (AI) that uses statistical algorithms to enable computer systems to learn and improve from experience [5]. It involves training a model on a set of data, such that the model can make predictions on new, unseen data [6]. Deep learning (DL) involves training neural networks with multiple layers to learn and extract features from data [7]. DL has been successful in several applications, including image recognition, speech recognition and NLP.

B. Artificial Neural Networks

ANNs draws inspiration from the structure and function of biological neurons. ANNs are composed of layers of interconnected nodes, or neurons, which receive input data, perform complex computations, and produce output results. The architecture of ANNs enables them to learn patterns and relationships within data, making them an effective tool for applications such as image recognition, natural language processing, and predictive modelling[8]. The connections between neurons are modelled by numerical weights, which are adjusted during training to increase the performance of the network [9].

C. Extreme Learning Machine

ELM is a new approach in the domain of artificial neural networks (ANNs), designed to surmount the shortcomings of conventional training methods. Essentially, ELM is a feedforward network with a solitary hidden layer. ELM, unlike conventional ANNs, randomly initialises and fixes the input-hidden layer connections. Instead, only the weights between the hidden and output layer are trained utilizing a least-squares approach. This unique training strategy sets ELM apart from conventional ANNs and makes it an innovative tool for machine learning tasks[10]. ELM has been shown to be fast, efficient, and accurate, making it suitable for various applications, including classification, regression, and clustering.

D. Advantages and Limitations

ELM has several advantages over traditional ANN training methods. Firstly, the random initialization of weights in ELM allows for fast training compared to the iterative training process used in traditional ANNs. Secondly, ELM is less prone to overfitting, as the fixed input-hidden layer weights act as a form of regularization. Thirdly, ELM can handle large datasets, as the training process includes calculation of a linear equations, which can be done efficiently using matrix operations [11].

However, ELM also has some limitations. Firstly, the fixed input-hidden layer weights mean that ELM may not be able to learn complex patterns in the data, which deeper networks could learn. Secondly, the least-squares approach used in ELM may not be suitable for some types of data, which require more complex models or training methods. Finally, the lack of interpretability in ELM, as with other ANNs, makes it difficult to understand how the model is making its predictions, which could be problematic in some applications.

In conclusion, ELM is a promising development in the field of ANNs, offering several advantages over traditional training methods, such as fast training and efficient handling of large datasets. However, ELM also has some limitations, particularly with regard to its ability to learn complex patterns in the data and interpretability. Overall, ELM represents an important step towards developing more efficient and accurate ML models with potential applications in various fields.

III. ELM ALGORITHMS

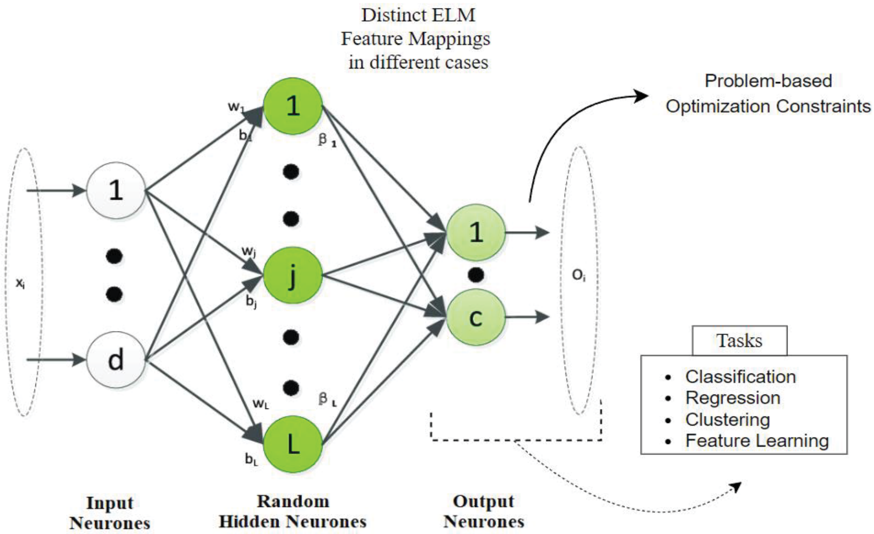

Fig. 1. A Standard Framework of Extreme Learning Machine

ELM method efficiently trains SLFNs in a speedy manner. It has gained popularity due to its ability to provide fast and accurate solutions to a wide range of problems. Compared to traditional neural network algorithms, ELM takes less execution time and computational power, making it a more cost-effective solution for variety of applications. Additionally, ELM is easy to implement and requires minimal hyperparameter tuning, making it an attractive choice for both experts and novices in the domain of machine learning. The algorithm contains four main components: input weights, hidden layer activation function, output weights, and bias terms. The ELM process can be divided into five subtopics: SLFNN, random feature selection, regularization techniques, learning algorithms, and convergence analysis. Fig. 1. Depicts a standard framework of an ELM algorithm [12].

They were first introduced by Huang *et al.* in 2006. They proposed a SLFNN architecture with fixed output weights and randomly generated input weights and showed that it could achieve comparable or even better performance than traditional neural networks on a range of classification and regression tasks [13].

In 2007, Huang *et al.* proposed a generalized version of ELMs called the Kernel-based ELM (KELM), which extends the original ELM framework to nonlinear classification and regression problems. Using a kernel function, the input data is positioned in a high-dimensional feature space[14]. Singh *et al.* explored the application of ELMs for time-series data [15].

In 2008, a number of researchers explored the theoretical foundations of ELMs [16] and KELMs [17], including their generalization properties and the relationship between ELMs and traditional neural networks.

In 2009, a number of variations and extensions of ELMs were proposed, including ELMs with regularization [18], online ELMs [19, 20], and incremental ELMs [21]. The advanced method of ELM can add arbitrary hidden neurons to SLFNs one at a time or in groups, depending on the situation. During the network expansion process, the final weights are subject to incremental revisions. In 2010, a new class of ELMs called the ELMs with Hidden Nodes (ELM-HNs) [22] were proposed, which introduced a layer of hidden nodes from the input to the output layers of the network and allowed for more flexible representations of the input data. Scholars began looking at the usage of ELMs in ensemble learning in 2010 [23, 24], and they came up with numerous different variations of ELM ensembles in 2011 [25, 26].

In 2012, researchers proposed a number of new ELM variants, including Deep ELMs [27], which introduced multiple hidden layers into the ELM architecture [28], and Stacked ELMs, which combined multiple ELM models into a single stacked model [29].

In 2013, researchers explored the use of ELMs in a range of real-world applications, including image classification [30], speech recognition [31], and financial forecasting [32].

In 2014, researchers proposed a new variant of ELM called the Convolutional ELM (CELM), which introduced a convolutional layer into the ELM architecture. It allowed for a more effective representation of image data [33].

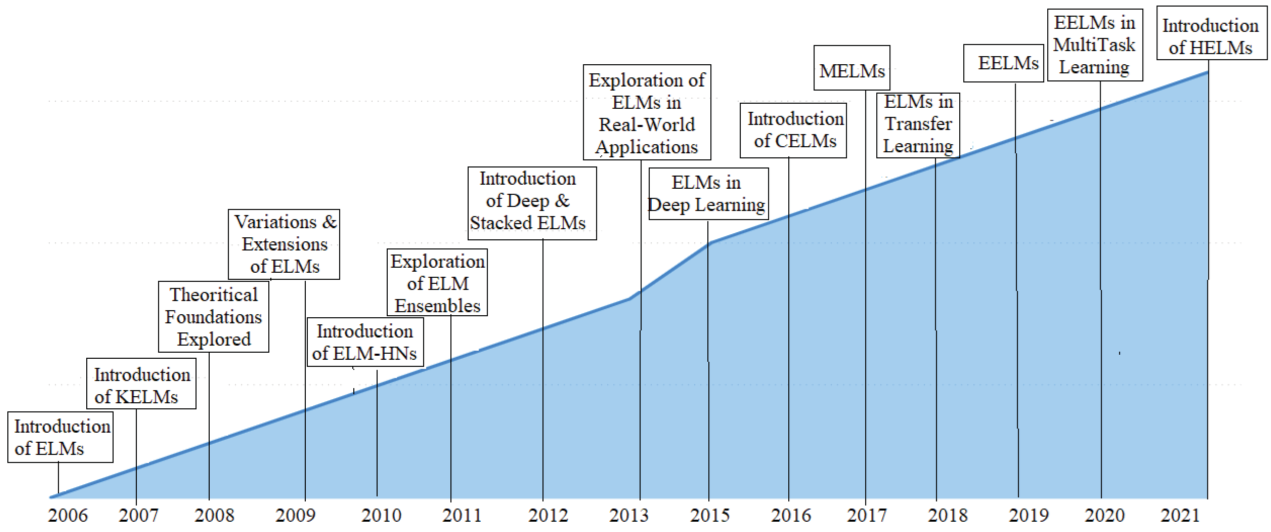

Fig. 2. Year-wise Advancements of Extreme Machine Learning

In 2015, researchers explored the use of ELMs in deep learning architectures and proposed several variants of ELM-based deep learning models [34].

In 2016, researchers offered a new ELM variant called the Compact Extreme Learning Machine (CELM), which aimed to reduce the memory requirements of ELM models [35].

In 2017, researchers proposed a new ELM variant called the Multi-scale ELM (MELM), which introduced a multi-scale analysis of the input data and showed improved performance on several image classification tasks [36-38].

In 2018, researchers explored the use of ELMs in transfer learning and proposed several ELM-based transfer learning models [39-41].

In 2018, researchers suggested a new ELM variant called the Evolutionary Extreme Learning Machine (EELM), which used a genetic algorithm to enhance the input and output weights of the network [42]. This arena of ELM was explored more in later years [43].

In 2020, researchers explored the use of ELMs in multi-task learning and proposed several ELM-based multi-task learning models [44]. More variants of ELMs were explored to optimize the algorithm for different tasks [45].

In 2020-21, researchers recommended a new ELM variant called the Hierarchical Extreme Learning Machine (HELM), which introduced a hierarchical structure into the ELM architecture and improved performance on several image classification tasks [46-48].

A. Single-hidden layer feedforward neural network

It is an ANN with single hidden layer. The hidden layer is responsible for processing the input data and converting it into a form that is suitable for the output layer. The output layer generates the output based on the input data. In this instance, the Moore-Penrose generalised inverse is utilised in order to perform systematic calculations on the input weights, hidden layer activation function, and output weights after they have been generated at random.

B. Random feature selection

Random feature selection is an approach for reducing the dimensionality of large datasets. It involves selection of set of features with the objective of retaining the same level of accuracy. The goal is to reduce feature count to provide simpler and faster models that are less prone to overfitting. This method has received a lot of use in many different fields, including machine learning, data mining, and image processing, among others. By selecting the most relevant features, random feature selection can improve the performance of models while reducing the computational cost associated with analyzing large datasets.

C. Regularization techniques

Regularization techniques are used to prevent overfitting in machine learning algorithms. In the ELM algorithm, regularization techniques are used to reduce the model's complexity and prevent overfitting. The ELM algorithm's most commonly used regularization techniques are L1 and L2 regularization.

D. Learning algorithms

The ELM algorithm can be trained using several learning approaches, such as batch learning, online learning, and incremental learning. The batch learning technique involves the use of entire dataset to train the algorithm. The online learning approach trains the ELM algorithm one sample at a time. Lastly, the incremental learning approach trains the ELM algorithm by using small batches of data. All these learning methods can be used to effectively train the ELM algorithm.

E. Convergence analysis

Convergence analysis determines if the ELM algorithm has converged to a solution. The convergence analysis of the ELM method involves analyzing the convergence of the input weights, hidden layer activation function, output weights, and bias terms. The convergence analysis of the ELM algorithm is important to ensure that the algorithm has learned the underlying patterns in the data.

TABLE I. COMPARISON OF ELM WITH EXISTING MACHINE LEARNING ALGORITHMS

Algorithm	Advantages	Disadvantages
ELM	A Fast, simple, efficient, good generalization ability	Needs a substantial number of hidden nodes to obtain good accuracy
SVM	Good generalization ability, effective in high-dimensional spaces	Computationally intensive requires tuning of hyperparameters
Random Forest	Decent accuracy can handle missing values and noisy data	Prone to overfitting, difficult to interpret
KNN	Simple, adequate for small datasets	Computationally intensive, sensitive to noise
Naive Bayes	Simple, fast, and effective for high-dimensional data	Assumes independence between features, can perform poorly if this assumption is violated

In summary, the ELM algorithm is a powerful tool for training neural networks, but its limitations must be considered when using it for real-world applications.

IV. ELM APPLICATIONS

ELM is popular in various fields due to their superior performance and simplicity. We go over some of the uses for ELM in this section.

ELM has been shown to be highly effective in image classification tasks, even when handling large datasets, with accuracy rates that are noteworthy. Furthermore, ELM has also been applied successfully to other domains, such as speech recognition, text classification, and regression analysis.

Feature selection is another domain in which ELM has demonstrated its effectiveness. By utilizing ELM, it is possible to identify the most relevant features in a dataset, improving the overall accuracy of machine learning models.

A. Classification

One of the most crucial machine learning tasks is classification. The use of ELM has been successful to solve various classification problems, such as face recognition, handwritten digit recognition, and spam filtering. For example, ELM has been used to classify EEG signals to diagnose epilepsy and Alzheimer's disease. In another study, ELM was used for the classification of breast cancer, achieving high accuracy compared to other classification methods.

B. Regression

Regression is another important task in machine learning. ELM widely used for various regression problems, such as predicting housing prices, stock prices, and energy consumption. For example, ELM predicts concrete compressive strength using water-cement ratio, age, and curing time. ELM has also been used to predict the energy consumption of buildings using various input parameters such as temperature, humidity, and occupancy.

C. Image Processing

ELM has been broadly used in image processing tasks, such as image recognition, segmentation, and denoising. For example, ELM has been used for face recognition, achieving high accuracy compared to other methods. ELM has also been used for object recognition in images and videos, achieving high accuracy and robustness.

D. Natural Language Processing

ELM has been used for sentiment analysis of movie reviews, achieving high accuracy compared to other methods. ELM has also been used for text classification of news articles, achieving high accuracy and efficiency.

E. Time Series Prediction

Time series prediction is important in various arenas, such as finance, weather forecasting, and energy consumption prediction. ELM has been used for the prediction of stock prices, achieving high accuracy compared to other methods. ELM has also been used for the prediction of wind speed and solar power generation, achieving high accuracy and robustness.

In summary, ELM has been successfully applied to various applications in different fields, including Classification, Regression, Image Processing, NLP, and Time Series Prediction. The simplicity and efficiency of ELM make it an attractive choice for solving complex problems in machine learning.

V. ELM EXTENSIONS

ELM is a powerful machine learning approach. In this paper, we comprehensively survey ELM theories, algorithms, and applications and focus specifically on ELM extensions.

- Incremental learning is an extension of ELM that allows for continuous learning and updating of the model as new data becomes available. This is particularly useful in scenarios where the data is dynamic and constantly changing, such as in financial markets or social media analytics. The incremental learning method ensures that the model remains fitted with the latest trends and patterns in the data without having to retrain the model from scratch.
- Deep ELM: Deep ELM is an extension of ELM that allows for the creation of deep neural networks with multiple hidden layers. This approach has been shown to increase performance of ELM models on complex tasks for example image classification and natural language processing. Deep ELM networks are typically trained using a combination of supervised and unsupervised learning techniques, such as backpropagation and autoencoders.
- Sparse ELM: Sparse ELM is an extension of ELM that uses a sparse input representation to reduce the dimensionality of the input data. It is particularly useful in scenarios where the input data is high-dimensional and contains many irrelevant features. By using a sparse input representation, Sparse ELM models improves model's accuracy while reducing the computational load.
- Online ELM: Online ELM is an extension of ELM that allows for the continuous learning of the model as new data becomes available without requiring a large memory buffer. This approach is particularly useful in real-time scenarios where the data is streaming, such as in sensor networks or social media analytics. Online ELM models are typically trained using a sliding

window approach, where the model is efficient with the most recent data points, and the oldest data points are discarded.

- Ensemble ELM: Ensemble ELM is an extension of ELM that combines multiple ELM models to improve the model's overall performance. This approach is particularly useful in scenarios where the data is complex and difficult to model using a single ELM model. Ensemble ELM models are typically trained using a combination of bagging and boosting techniques. A weighted average is used to combine the results of models that were trained on various subsets of the data.

In conclusion, ELM extensions offer a variety of techniques to improve the performance and flexibility of ELM models, making them suitable for a wide range of applications. Incremental learning, deep ELM, sparse ELM, online ELM, and ensemble ELM are just some of the techniques that can be used to enhance the capabilities of ELM models and make them more effective in handling large and complex datasets.

VI. Comparison with Other Machine Learning Methods

Several machine learning methods are available today, each with its strengths and weaknesses. This section compares extreme learning machines (ELMs) with other popular ML methods.

A. *Support vector machines (SVMs)*

SVMs are a popular method for classification and regression tasks. Unlike ELMs, SVMs endeavour to identify the optimal hyperplane that divides the data into two classes. SVMs work well with small to medium-sized datasets and can handle both linear and nonlinear problems. However, SVMs can be slow when dealing with large datasets, and their performance depending on the kernel function selected.

B. *Random forests*

Random forests are a method of ensemble learning that combines multiple decision trees to improve precision and reduce overfitting. Random forests work well with high-dimensional datasets and can handle both classification and regression tasks. However, random forests can be computationally expensive and may not perform well when the input features are highly correlated.

C. *Deep learning*

It is a specialized branch of ML that employs NN with multiple layers to develop intricate data representations. This cutting-edge technology has achieved top-tier results in several areas, such as speech recognition, natural language processing, and image identification. However, deep learning requires large amounts of labelled data and can be computationally expensive, making it challenging to train and deploy on resource-constrained devices.

D. *Other neural network architectures*

ELMs are a type of FNN with a single hidden layer of randomly initialized neurons. Compared to other neural network architectures, such as multilayer perceptrons (MLPs) and convolutional neural networks (CNNs), ELMs are faster to train and require fewer hyper parameters. However, ELMs may not perform as well as other neural network architectures on certain tasks, particularly those that require hierarchical representations of data.

TABLE II. Comparison of ELMs with other machine learning methods

Method	Pros	Cons
ELMs	Fast training, few hyperparameters	It may not perform as well as other neural network architectures
SVMs	Can handle linear and nonlinear problems, good with small to medium datasets	It can be slow with large datasets. Performance depends on kernel choice
Random forests	Ensemble learning, good with high-dimensional datasets	It can be computationally expensive and may not perform well with highly correlated input features
Deep learning	State-of-the-art performance on many tasks	Requires large amounts of labelled data, computationally expensive, challenging to train and deploy on resource-constrained devices

VII. ELM in Practice

In this section the practical aspects of implementing ELM models, evaluating their performance, tuning the parameters, and using software tools and libraries are discussed.

A. *Implementation Considerations*

Implementing an ELM model requires several considerations, including the choice of activation function, hidden layer structure, and input data normalization. The activation function used in ELM models should be differentiable, monotonically increasing, and bounded. Popular choices include sigmoid, Radial Basis Function (RBF) and hyperbolic tangent (tanh).

The hidden layer structure of ELM models can be either single-layer or multi-layer. The single-layer structure is simpler and faster but may not capture complex relationships in the data. Multi-layer structures can capture more complex relationships but require more training time and may be prone to overfitting.

Normalization of input data is an important consideration in ELM implementation. Normalization helps to prevent overfitting and ensures that the input data is in the same range as the activation function. Common normalization techniques include scaling the input data ranges between 0 and 1 or using z-score normalization.

B. *Performance Evaluation*

The effectiveness of ELM depends on various factors such as dataset complexity, size, and parameter selection. Nevertheless, in general, ELM is a versatile and powerful machine learning algorithm that has been demonstrated to be highly effective across a broad range of domains.

To assess the effectiveness of an ELM model, there are multiple metrics that can be utilized. We have performed ELM for sentiment analysis.

In this example, the ELM algorithm achieved an overall accuracy of 0.87, indicating that 87% of the instances were classified correctly. The precision score of 0.85 indicates that out of all the instances classified as positive, 85% were actually positive. The recall score of 0.89 suggests that out of all the actual positive instances, the model correctly identified 89% of them. The F1 score of 0.87 indicates a balanced performance between precision and recall. Finally, the AUC score of 0.91 shows that the ELM model has a good ability to distinguish between positive and negative instances at different probability thresholds.

TABLE III. PERFORMANCE METRIC OF ELM FOR SENTIMENT ANALYSIS

Performance Metrics	Score
Accuracy	0.87
Precision	0.85
Recall	0.89
F1 score	0.87
AUC	0.91

By analyzing these metrics, one can gain valuable insights into the performance of the ELM model and determine whether it is meeting the desired criteria for the task at hand. It is essential to use these metrics appropriately to ensure that the evaluation is thorough and reliable. Accuracy measures the proportion of correctly classified instances, Precision and recall are two key metrics used. Precision measures the % of true positives among all positive predictions made by the model, while recall measures the percentage of true positives among all actual positive instances in the dataset. In other words, precision quantifies the model's ability to accurately identify positive instances, while recall quantifies the model's ability to detect all positive instances. Both metrics are important in assessing the overall effectiveness of a predictive model, as they provide insight into its strengths and weaknesses. The F1 score is a harmonic mean that is derived from precision and recall, whereas the area under the ROC curve is a measurement of the model's ability to differentiate between positive and negative instances.

C. Parameter Tuning

The performance of ELM models can be improved by tuning their parameters (N, C, f). It should be chosen based on the complexity of the data, while the regularization parameter helps to prevent overfitting. The activation function can also be tuned to improve performance, with the sigmoid and tanh functions being suitable for classification tasks, while the RBF function is more suitable for regression tasks.

D. Software Tools and Libraries

Several software tools and libraries are available for implementing ELM models, including MATLAB, Python, and R. In MATLAB, the ELM toolbox provides an implementation of ELM models and several evaluation metrics. In Python, the scikit-learn library provides an implementation of ELM models and several evaluation metrics, while the Keras library can be used to implement ELM models as part of a neural network. In R, the elmNN package provides an implementation of ELM models.

E. Important parameters:

- N represents the quantity of neurons in the hidden layer of ELM model (no. of hidden nodes).
- Regularization parameter (C) is used to control the trade-off between fitting the training data well and avoiding overfitting. It is usually included in the objective function of the ELM and is multiplied by the regularization term.
- Activation function (f) used in the ELM model is typically denoted as f(x) and should be differentiable, monotonically increasing, and bounded. Some popular choices include the sigmoid function, tanh function, and radial basis function (RBF).
- Input data normalization can be done using scaling or z-score normalization. For example, if we have a feature vector x, the scaling normalization can be defined as:

$$x\_scaled = (x - min(x))/(max(x) - min(x)) - - - - - - - -(1)$$

- The z-score normalization calculated as:

$$x\_zscore = (x - mean(x))/std(x) - - - - - - - - - - -(2)$$

VIII. CHALLENGES & FUTURE DIRECTIONS

A. Interpretability

One of the major challenges in ELM is its lack of interpretability. ELM is considered a "black box" model because it does not provide explanations of how it makes decisions. This limitation can make it difficult for users to understand the reasons for the predictions made by the model. Future research should focus on developing methods to increase the interpretability of ELM, such as identifying important features and providing insights into the decision-making process.

B. Scalability

Despite ELM's fast learning speed and efficiency, it may face challenges in scalability when dealing with large datasets. The traditional ELM approach is founded on only SLFN, which may not be efficient in handling large amounts of data. Researchers should explore the development of scalable ELM models, such as deep ELM models, to overcome these scalability issues.

C. Robustness

Another challenge in using ELM is its vulnerability to noisy data. ELM models are prone to overfitting when there is a high level of noise in the data. To address this issue, future research should focus on developing robust ELM models that can handle noisy data without compromising performance. This can be achieved by incorporating regularization techniques, data pre-processing methods, or developing new algorithms that are specifically designed to handle noisy data.

D. Emerging Research Topics

There are still many potential applications of ELM that have not been explored yet. Future research should focus on exploring novel applications of ELM, such as in the fields of finance, healthcare, and transportation. There are several emerging research topics in ELM that require further investigation. For instance, the development of ELM models that can handle non-stationary data, such as data streams. Another emerging research topic is the integration of ELM with other machine learning techniques, such as deep learning or reinforcement learning. Additionally, the development of

ELM models that can handle multimodal data, such as audio-visual data, is an interesting area of research.

IX. Conclusion

The paper surveys ELM theories, algorithms, and applications in various domains. The key findings indicate that ELM is a promising effective learning algorithm that offers noteworthy benefits like fast learning speed, excellent generalization performance, and low complexity. Moreover, the paper highlights the potential of ELM for various applications, including image and speech recognition, forecasting, and control systems.

The implications of this research for practice and research are significant. Practitioners can utilize ELM as a valuable tool to improve their system's performance and reduce computational complexity. Furthermore, researchers can use ELM as an alternative approach to traditional machine learning algorithms for their research projects, which can provide new insights into the learning process of the model.

However, the study has certain limitations that require further investigation. The paper mainly focuses on the theoretical aspects of ELM, and there is a need for more empirical studies to evaluate its performance in real-world applications. Additionally, the study's analysis is limited to a few specific domains, and further research is necessary to assess its effectiveness in different domains.

Future work should concentrate on exploring the potential of ELM in deep learning, improving the interpretability of ELM models, and developing new algorithms that can enhance the learning efficiency of ELM. Overall, this survey provides a valuable reference for researchers and practitioners to better understand the ELM algorithm's potential and limitations.

References

[1] J. Wang, S. Lu, S.-H. Wang, and Y.-D. Zhang, "A review on extreme learning machine," *Multimedia Tools Applications,* vol. 81, no. 29, pp. 41611-41660, 2022.

[2] S. Ding, X. Xu, and R. Nie, "Extreme learning machine and its applications," *Neural Computing Applications,* vol. 25, pp. 549-556, 2014.

[3] C. Chen, K. Li, M. Duan, and K. Li, "Extreme Learning Machine and Its Applications in Big Data Processing," *Intelligent Data-Centric Systems,* pp. 117-150, 2017.

[4] K. Khan, E. Ratner, R. Ludwig, and A. Lendasse, "Feature bagging and extreme learning machines: machine learning with severe memory constraints," in *2020 International Joint Conference on Neural Networks (IJCNN)*, 2020, pp. 1-7: IEEE.

[5] Z.-H. Zhou, *Machine learning*. Springer Nature, 2021.

[6] R. P. Bunker and F. Thabtah, "A machine learning framework for sport result prediction," *Applied Computing Informatics,* vol. 15, no. 1, pp. 27-33, 2019.

[7] D. Learning, "Deep learning," *High-dimensional Fuzzy Clustering,* 2020.

[8] Z. Zhang and Z. Zhang, "Artificial neural network," *Multivariate time series analysis in climate environmental research,* pp. 1-35, 2018.

[9] A. M. Zador, "A critique of pure learning and what artificial neural networks can learn from animal brains," *Nature communications,* vol. 10, no. 1, p. 3770, 2019.

[10] K. Erdem, "Introduction to Extreme Learning Machines," *Towards Data Science,* 2020.

[11] F. Cahyo, "Extreme Learning Machine for Simple Classification," *DataDrivenInvestor,* 2019.

[12] F. Zhang, J. Liu, B. Wang, Z. Qi, and Y. Shi, "A fast algorithm for multi-class learning from label proportions," *Electronics,* vol. 8, no. 6, p. 609, 2019.

[13] S. Pande, N. K. Rathore, and A. Purohit, "A survey and analysis of extreme machine learning models and its techniques," 2021.

[14] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: theory and applications," *Neurocomputing,* vol. 70, no. 1-3, pp. 489-501, 2006.

[15] R. Singh and S. Balasundaram, "Application of extreme learning machine method for time series analysis," *International Journal of Computer Information Engineering,* vol. 1, no. 11, pp. 3407-3413, 2007.

[16] Q. Liu, Q. He, and Z. Shi, "Extreme support vector machine classifier," in *Advances in Knowledge Discovery and Data Mining: 12th Pacific-Asia Conference, PAKDD 2008 Osaka, Japan, May 20-23, 2008 Proceedings 12*, 2008, pp. 222-233: Springer.

[17] C.-M. Zhai and J.-X. Du, "Applying extreme learning machine to plant species identification," in *2008 International Conference on Information and Automation*, 2008, pp. 879-884: IEEE.

[18] W. Deng, Q. Zheng, and L. Chen, "Regularized extreme learning machine," in *2009 IEEE symposium on computational intelligence and data mining*, 2009, pp. 389-395: IEEE.

[19] Y. Lan, Y. C. Soh, and G.-B. Huang, "Ensemble of online sequential extreme learning machine," *Neurocomputing,* vol. 72, no. 13-15, pp. 3391-3395, 2009.

[20] H.-J. Rong, G.-B. Huang, N. Sundararajan, and P. Saratchandran, "Online sequential fuzzy extreme learning machine for function approximation and classification problems," *IEEE Transactions on Systems, Man, Cybernetics, Part B,* vol. 39, no. 4, pp. 1067-1072, 2009.

[21] G. Feng, G.-B. Huang, Q. Lin, and R. Gay, "Error minimized extreme learning machine with growth of hidden nodes and incremental learning," *IEEE Transactions on Neural Networks,* vol. 20, no. 8, pp. 1352-1357, 2009.

[22] Y. Lan, Y. C. Soh, and G.-B. Huang, "Constructive hidden nodes selection of extreme learning machine for regression," *Neurocomputing,* vol. 73, no. 16-18, pp. 3191-3199, 2010.

[23] N. Liu and H. Wang, "Ensemble based extreme learning machine," *IEEE Signal Processing Letters,* vol. 17, no. 8, pp. 754-757, 2010.

[24] G. Wang and P. Li, "Dynamic Adaboost ensemble extreme learning machine," in *2010 3rd International Conference on Advanced Computer Theory and Engineering (ICACTE)*, 2010, vol. 3, pp. V3-54-V3-58: IEEE.

[25] Y. Lan, "Network ensemble and constructive algorithms for model selection of extreme learning machine," 2011.

[26] P. J. García-Laencina, A. Bueno-Crespo, and J.-L. Sancho-Gómez, "Design and training of neural architectures using extreme learning machine," *Neurocomputing: Learning, Architectures Modeling, ed ET Mueller,* pp. 119-145, 2011.

[27] L. Singh and G. Chetty, "Detecting The Brain Abnormalities From Mri Structural Images Using Machine Learning And Pattern Recognition Tools," *IJREISS,* vol. 2, no. 11, pp. 15-30, 2012.

[28] R. Wang, S. Kwong, and X. Wang, "A study on random weights between input and hidden layers in extreme learning machine," *Soft Computing,* vol. 16, pp. 1465-1475, 2012.

[29] L. Deng, G. Tur, X. He, and D. Hakkani-Tur, "Use of kernel deep convex networks and end-to-end learning for spoken language understanding," in *2012 IEEE Spoken Language Technology Workshop (SLT)*, 2012, pp. 210-215: IEEE.

[30] F. Cao, B. Liu, and D. S. Park, "Image classification based on effective extreme learning machine," *Neurocomputing,* vol. 102, pp. 90-97, 2013.

[31] Y. Lan, Z. Hu, Y. C. Soh, and G.-B. Huang, "An extreme learning machine approach for speaker recognition," *Neural Computing Applications,* vol. 22, pp. 417-425, 2013.

[32] J. P. Nobrega and A. L. I. De Oliveira, "Improving the statistical arbitrage strategy in intraday trading by combining extreme learning machine and support vector regression with linear regression models," in *2013 IEEE 25th International Conference on Tools with Artificial Intelligence*, 2013, pp. 182-188: IEEE.

[33] S. Lin, X. Liu, J. Fang, and Z. Xu, "Is extreme learning machine feasible? A theoretical assessment (Part II)," *IEEE Transactions on Neural Networks Learning Systems,* vol. 26, no. 1, pp. 21-34, 2014.

[34] J. Tang, C. Deng, and G.-B. Huang, "Extreme learning machine for multilayer perceptron," *IEEE transactions on neural networks learning systems,* vol. 27, no. 4, pp. 809-821, 2015.

[35] W.-Y. Deng, Y.-S. Ong, and Q.-H. J. N. N. Zheng, "A fast reduced kernel extreme learning machine," *Neural Networks,* vol. 76, pp. 29-38, 2016.

[36] J. Huang *et al.*, "Extreme learning machine with multi-scale local receptive fields for texture classification," *Multidimensional Systems Signal Processing,* vol. 28, pp. 995-1011, 2017.

[37] S. Zhu and B. Jiao, "Fault diagnosis of wind bearing based on multi-scale wavelet kernel extreme learning machine," in *Journal of Physics: Conference Series*, 2017, vol. 887, no. 1, p. 012070: IOP Publishing.

[38] X. Li, Q. Wei, Y. Qu, and L. Cai, "Incipient loose detection of hoops for pipeline based on ensemble empirical mode decomposition and multi-scale entropy and extreme learning machine," in *IOP Conference Series: Materials Science and Engineering*, 2017, vol. 211, no. 1, p. 012011: IOP Publishing.

[39] T. Alshalali and D. Josyula, "Fine-tuning of pre-trained deep learning models with extreme learning machine," in *2018 International Conference on Computational Science and Computational Intelligence (CSCI)*, 2018, pp. 469-473: IEEE.

[40] C. Chen, B. Jiang, and X. Jin, "Parameter transfer extreme learning machine based on projective model," in *2018 International joint conference on neural networks (IJCNN)*, 2018, pp. 1-8: IEEE.

[41] R. Ye and Q. Dai, "A novel transfer learning framework for time series forecasting," *Knowledge-Based Systems,* vol. 156, pp. 74-99, 2018.

[42] T. Mahmoud, Z. Dong, and J. Ma, "An advanced approach for optimal wind power generation prediction intervals by using self-adaptive evolutionary extreme learning machine," *Renewable energy,* vol. 126, pp. 254-269, 2018.

[43] R. Bisoi, P. Dash, and A. Parida, "Hybrid variational mode decomposition and evolutionary robust kernel extreme learning machine for stock price and movement prediction on daily basis," *Applied Soft Computing,* vol. 74, pp. 652-678, 2019.

[44] K. Jayapriya, I. J. Jacob, and P. E. Darney, "Hyperspectral image classification using multi-task feature leverage with multi-variant deep learning," *Earth science informatics,* vol. 13, no. 4, pp. 1093-1102, 2020.

[45] M. Suchithra and M. L. Pai, "Improving the prediction accuracy of soil nutrient classification by optimizing extreme learning machine parameters," *Information processing in Agriculture,* vol. 7, no. 1, pp. 72-82, 2020.

[46] X. Xu, S. Li, T. Liang, and T. Sun, "Sample selection-based hierarchical extreme learning machine," *Neurocomputing,* vol. 377, pp. 95-102, 2020.

[47] R. Li, X. Wang, Y. Song, and L. Lei, "Hierarchical extreme learning machine with L21-norm loss and regularization," *International Journal of Machine Learning Cybernetics Information Technologies,* vol. 12, no. 5, pp. 1297-1310, 2021.

[48] A. M. Saatloo, A. Moradzadeh, H. Moayyed, M. Mohammadpourfard, and B. Mohammadi-Ivatloo, "Hierarchical extreme learning machine enabled dynamic line rating forecasting," *IEEE Systems Journal,* vol. 16, no. 3, pp. 4664-4674, 2021.