An Analysis of Extreme Machine Learning Models

Shubhangi Pande

Shri Govindram Seksaria Institute of Technology and Science

Neeraj Rathore (✉ neerajrathore37@gmail.com)

Indira Gandhi National Tribal University

Anuradha Purohit

Shri Govindram Seksaria Institute of Technology and Science

Research Article

Keywords: Backpropagation, extreme learning machine, artificial neural networks, support vector machine, least square-support vector machine, single-layer feed forward networks.

Posted Date: August 10th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-667827/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

An Analysis of Extreme Machine Learning Models

¹Shubhangi Pande, ²Dr. Neeraj Kmar Rathore, ³Dr. Anuradha Purohit

¹PhD Scholar, ^{2,3}Associate Professor

^{1,2,3}Department of Computer Science & Engineering

^{1,3}SGSITS, Indore (M.P.), India, ²Indira Gandhi Natinal Tribal University (IGNTU)-Central University, Amarkantak (M.P.), India.

¹shubhangi.shanu@gmail.com, ²neerajrathore37@gmail.com, ³anuradhapurohit78@gmail.com

Abstract

Machine learning applications employ FFNN (Feed Forward Neural Network) in their discipline enormously. But, it has been observed that the FFNN requisite speed is not up to the mark. The fundamental causes of this problem are: 1) for training neural networks, slow gradient descent methods are broadly used and 2) for such methods, there is a need for iteratively tuning hidden layer parameters including biases and weights. To resolve these problems, a new emanant machine learning algorithm, which is a substitution of the feed-forward neural network, entitled as Extreme Learning Machine (ELM) introduced in this paper. ELM also come up with a general learning scheme for the immense diversity of different networks (SLFNs and multilayer networks). According to ELM originators, the learning capacity of networks trained using backpropagation is a thousand times slower than the networks trained using ELM, along with this, ELM models exhibit good generalization performance. ELM is more efficient in contradiction of Least Square Support Vector Machine (LS-SVM), Support Vector Machine (SVM), and rest of the precocious approaches. ELM's eccentric outline has three main targets: 1) high learning accuracy 2) less human intervention 3) fast learning speed. ELM consider as a greater capacity to achieve global optimum. The distribution of application of ELM incorporates: feature learning, clustering, regression, compression, and classification. With this paper, our goal is to familiarize various ELM variants, their applications, ELM strengths, ELM researches and comparison with other learning algorithms, and many more concepts related to ELM.

Keywords: Backpropagation, extreme learning machine, artificial neural networks, support vector machine, least square-support vector machine, single-layer feed forward networks.

1. Introduction

A kind of artificial neural network, acronym as ANN, is a feed-forward neural network (FFNN), in which no loop exist between the connections of units. This type of network is essentially adapted for supervised learning and it is a biologically inspired classification algorithm [1] (as shown in Fig 1). Primarily, FFNN is implemented by notorious backpropagation (BP) algorithm (1986, Rumelhart, Hinton & Williams), which has been suffered from two massive problems: reluctant convergence and local minimum problem. Thenceforth, numerous approaches are developed to train FFNN in efficient and optimal manner, it comprise second order optimization method (1994, 2010), subset selection method (1991, 2005) and global optimization method (1995, 1993). Prior approaches outperform BP (nevertheless assured to achieve global optimum) in terms of training speed and generalisation performance [2] (as shown in Fig 2).

Fig 1: FFNN (Feed Forward Neural Network)

Fig 2: FFNN past history implementation

G. B. Huang, in 2004, suggested an idea of Extreme Learning Machine, abbreviated as ELM, breaks the obstacle between typical artificial learning techniques and biological learning procedures [3]. According to ELM originators, the learning capacity of networks trained using backpropagation is a thousand times slower than the networks trained using ELM, along with this, ELM models exhibit good generalization performance. ELM consider as a greater capability to achieve global optimum. The fastest speed of ELM is due to the elimination of the need for iterative tuning of hidden node parameters and ELM also satisfy Bartlett's theory (as shown in Fig 3).

Fig 3: Bartlett's theory statement

In ELM, 'Extreme' indicate movement in the direction of brain like learning from typical learning techniques [4]. Initially, ELM employed for SLFNs (Single Layer Feed-Forward

Networks). There are two fundamental problems rectify by Extreme Learning Machine (ELM): 1) Different types of networks (SLFNs and multilayer networks) necessitate different types of learning algorithms 2) Requirement of tuned extensive amount of hidden neurons together with biological neurons (although whose modelling is unrevealed) in learning [5]. Here, different type of SLFNs (Single Layer Feed-Forward networks) includes RBF networks, polynomial networks, sigmoid networks, complex (domain) networks, wavelet networks, Fourier series etc. There are basically three types of approaches in training of feed-forward networks: a) Gradient Descent approach (used by Backpropagation (BP) for multilayer feed-forward networks) b) Standard optimization approach (employed by Support Vector Machine (SVM)) and c) Least Square based approach (used by ELM for generalized SLFNs). The difference in the optimization strategy of implementation in conventional learning methods and ELM as shown in Fig 4. ELM emphasize on mainly three components: real time learning, High accuracy and least user intervention (as shown in Fig 5).

Fig 4: Optimization method used in conventional learning and ELM techniques.

Fig 5: Essential considerations of ELM

As we know that, Rosenblatt perceptron, Baum, RVFL (Random Vector Function Link), Schmidt, Quick Net all are comes under neural network and PCA (Principal Components Analysis), SVM (Support Vector Machine), PSVM (Proximal Support Vector Machine), LS-SVM (Least Square Support Vector Machine), Random projection associated with feature space methods. Prior to ELM theory, these neural networks had some issues, such as in full random nodes, separational capacity is not substantiated, optimization constraints are not adjusted, and the dimensionality of hidden maps is smaller than the amount of training data, universal approximation capabilities not verified and proving.. As in the farmer case, feature space methods have a few flaws, such as a hazy relationship with neural networks and the possibility that universal approximation capability has not been validated. ELM bridged the gap between these neural networks and feature space methods and also filled gap between these methods and biological learning [5] (as shown in Fig 6 and Fig 7).

Fig 6: How ELM fills gap between neural network methods and feature space methods

The most effective ELM application areas are regression, classification, clustering, and feature learning. The additional characteristics exhibit by ELM are: It extends to almost every nonlinear piecewise continuous function, expands to kernels and to a higher degree of mapping of hidden instances, proves the universal approximation and separation capability of generalised SLFNs, employs generalisation theory for learning optimization, and demonstrates the linear independence of random hidden neurons., It also demonstrates that autonomous parameters of hidden node data can be used as a bridge between ridge rectification, system stability, theoretically based neural network generation, maximum margin and additionally network optimization constraints included in the ELM framework.

Fig 7: Resemblance between biological learning and ELM and additive features of ELM

Figure 7 illustrates ELM characteristics on the basis of generalization SLFNs and multi hidden layers of networks, structure risk minimization and ridge regression theories, homogeneous learning algorithms and kernels/random hidden nodes/neurons inherited from ancestors. The hidden neurons in ELM can be allocated randomly and cannot be upgraded or they can be inherited without being modified by their ancestors. Linear models require learning in just one step, in ELM you can learn the output of the hidden nodes in one step. ELM theories show that these hidden nodes rely on training data and that the weight value of the hidden layer should not be adjusted. The ELM original design has three objectives in terms of learning effectiveness: high accuracy, minimal interaction between people and fast learning speed. When ELM trains on SLFNs, there are two major stages: randomised feature mapping and linear parameter solving. On the basis of different parameters, Fig 8 demonstrates ELM characteristics through which understanding of ELM concept is much better.

Fig 8: ELM characteristics on the basis of different parameter

The rest of the paper is organised as follows:: Section 2 discusses the “Extreme Learning Machine”, which comprises how it works, its essence, its learning model, its kernel and non-kernel learning, its pertinent features and its learning principles which describe the basic learning principles on which ELM works. Section 3 tells about the “ELM theories”, which consist of theories satisfied by ELM. Section 4 comprises "application areas of ELM", which describe the application areas on which ELM is applicable. Section 5 represents “ELM approaches and paper contemplate”, which tells about the proactive approaches of ELM and survey of some journal papers. Section 6 consist of “comparison between biological learning, conventional learning and ELM”, which describe the differences among biological learning, conventional learning and ELM, and distinction between deep learning and ELM. Section 7 comprises “comparative analysis of data and suggested models” which describes the data set analysis used by various algorithms and their pictorial representation and also some suggested models which may solve some current problems. Section 8 comprises “conclusion”, which finally concludes the whole survey paper.

2. Extreme Learning Machine

Huang G.B. proposed a new approach to enhancing SLFN generalisation performance, called the Extreme Learning Machine, while avoiding the time-consuming iterative tuning process (ELM) [7]. The classical ELM procedure is as follows: First, it determines the total number of neurons in the network., then it sets the hidden layer's biases and the weights between the input layer and the hidden layer at random; and secondly, It computes the output matrix of the hidden layer and finally, unlike the conventional network learning algorithm, Using the least square method, the Moore-Penrose pseudoinverse computes the weight between the hidden and output layers (as shown in Fig 9).

Fig 9: Step-by-step procedure in classical ELM

Fig 10: The ELM network structure

One of the reason behind for adapting ELM is: its compact network structure (as shown in Fig 10) and easy process of network parameter computation and fast learning speed. During training of SLFNs by ELM, it consist of mainly two phases: a) randomized feature mapping, here randomly initialization of hidden layer is done so that by using some nonlinear feature mapping functions, input space is converted into ELM feature space (as shown in Fig 11) solving of linear parameters. Nonlinear feature mapping, which is not limited to any of the nonlinear piecewise continuous functions (given in Fig 12), are being applied in ELM.

Fig 11: Feature mapping and feature space of ELM

Fig 12: Commonly used mapping function of ELM

The output function of generalised SLFNs can be specified in ELM as -

$$f_L(x) = \sum_{i=1}^n \beta_i G(a_i, b_i, x) = h(x)\beta \dots \dots \dots (1)$$

Where,

$\beta = [\beta_1, \dots, \beta_L]^T$ = output weight vector between hidden layer nodes (L) and output layer nodes ($m \geq 1$)

$h(x) = [h_1(x), \dots, h_L(x)]$ = Nonlinear feature mapping of ELM

$$h(x) = [G(a_1, b_1, x), \dots, G(a_L, b_L, x)] \dots \dots \dots (2)$$

The ELM main idea is to allocate hidden layer weights and biases to random, followed by calculating weight with a least square solution defined by target outputs and the hidden layer.

Mathematically, we can represent ELM model as:

$$\sum_{i=1}^L \beta_i g_i(X_j) = \sum_{i=1}^L \beta_i g_i(W_i X_j + b_i) = t_j, \dots (3)$$

Where,

N = set of unique sample $(X_i, t_i) \dots$

$X_i = [x_{i1}, x_{i2}, \dots, x_{in}]^T \in \mathbb{R}^n$

$t_i = [t_{i1}, t_{i2}, \dots, t_{im}] \in \mathbb{R}^m$

L = hidden layer nodes

$g(x)$ = activation function

$W_i = [W_{i1}, W_{i2}, \dots, W_{in}]^T$ = weight vector that relate hidden node and i^{th} input nodes

$\beta_i = [\beta_{i1}, \beta_{i2}, \dots, \beta_{im}]^T$ = weight vector that relate hidden node and i^{th} output nodes

$b_i = i^{\text{th}}$ hidden nodes threshold

$J = 1, \dots, N$.

We can write compactly all the above equations as

$$H \beta = T \dots \dots \dots (4)$$

Where,

$$H(w_1, \dots, w_N, b_1, \dots, b_N, x_1, \dots, x_N) = \begin{bmatrix} g(w_1, b_1, x_1) & \dots & g(w_L, b_L, x_1) \\ \dots & \dots & \dots \\ g(w_1, b_1, x_N) & \dots & g(w_L, b_L, x_N) \end{bmatrix}$$

$$\beta = \begin{bmatrix} \beta_1^T \\ \dots \\ \beta_L^T \end{bmatrix} \text{ and } T = \begin{bmatrix} t_1^T \\ \dots \\ t_L^T \end{bmatrix}$$

By solving analytically, we can find out output weight (β) by least square solution as:

$$\beta = H^+ T \dots \dots \dots (5)$$

Where,

H^+ = Moore Penrose generalized inverse of H

ELM is a novel type of neural network that begins with a single hidden layer and only advances forward. The above equation declare that the reason of very short training time of ELM i.e. it only require to perform one iteration on the training data [8]. If such a case occurred in which activation function is infinitely differentiable and when the number of

hidden layer nodes ($\tilde{N}$) are corresponds with the different training samples (N) i.e. $\tilde{N} = N$, here H imply a squared matrix calculation of output weights i.e. β can be effectively carried out by inverting H and SLFN's estimated these training samples (N) with zero error i.e. $\sum_{j=1}^N \|o_j - t_j\| = 0$ [9]. In most cases, the number of hidden nodes is significantly lower than the number of diverse training samples. i.e. $\tilde{N} \ll N$, so H will be a non-square matrix and there may not exist b_i, w_i, β_i , where $i=1, \dots, \tilde{N}$ such that $H\beta = T$. In this situation, we might have to identify particular set of $\hat{w}, \hat{\beta}, \hat{b}$, (where, $i=1, \dots, \tilde{N}$) in such a manner [10]-

$$\|H(\hat{w}_1, \dots, \hat{w}_{\tilde{N}}, \hat{b}_1, \dots, \hat{b}_{\tilde{N}})\beta - T\| = \min_{w_i, b_i, \beta} \|H(\hat{w}_1, \dots, \hat{w}_{\tilde{N}}, \hat{b}_1, \dots, \hat{b}_{\tilde{N}})\beta - T\| \dots \dots \dots (6)$$

Various approaches are available to evaluate Moore-Penrose widespread inverse matrix including iterative approaches, orthogonal projection technique, orthogonalization strategy and SVD (singular value decomposition), etc. The random generation of input weights and hidden biases is a distinguishing feature of ELM instead of tuning these network parameters to turn the non-linear system into a linear system, $H\beta = T$ [6].

In these two cases, a regularisation term is added to β in order to make the result stable and improve the network's generalisation capacity:

- a) In this situation β can be described as follows, when the amount of training data goes beyond the number of hidden neurons:

$$\beta = (1/\lambda + H^T H)^{-1} H^T T \dots \dots \dots (7)$$

- b) In this situation, β could be affirmed as follows when the number of samples for training is below the number of hidden nodes:

$$\beta = H^T (1/\lambda + HH^T)^{-1} T \dots \dots \dots (8)$$

2.1 Essence of ELM

- Hidden layer neurons tuning is not essential.
 - "Randomness" is one of the ELM implementation, but not every method.
 - Some of the typical approaches employed "semi-randomness".
- Universal approximation conditions met by hidden layer mapping ($h(x)$).
- Minimize:

$$\|H\beta - T\|_p \text{ and } \|\beta\|_q$$

Where, p and q = may have different values like 1, 1/2, 2, ...
- Both ridge regression theory and network generalization theory satisfied by the ELM.
- It act as a link between SVM, neural networks, Fourier series, random projection, linear systems, matrix theories etc.

2.2 Kernel and non-kernel learning

ELM is justifiable for both non-kernel and kernel learning.

- a) In case of non-kernel based learning, mathematically it is given as-

$$f(x) = h(x) H^T \left(\frac{1}{c} + H H^T \right)^{-1} T \dots\dots\dots(9)$$

And

$$f(x) = h(x) \left(\frac{1}{c} + H^T H \right)^{-1} H^T T \dots\dots\dots(10)$$

- b) In case of kernel based learning (if $h(x)$ is unknown), it is given as-

$$f(x) = \begin{bmatrix} K(x, x_1) \\ \dots\dots\dots \\ K(x, x_N) \end{bmatrix}^T \left(\frac{1}{c} + \Omega_{ELM} \right)^{-1} T \dots\dots\dots(11)$$

Where, $\Omega_{ELM_{i,j}} = h(x_i) \cdot h(x_j) = K(x_i, x_j)$

2.3 Pertinent features of ELM

- The machine typically has one layer of hidden nodes. The weights between the input and hidden nodes are never changed.
- In a single step, weight between hidden nodes and the output neurons can be deduced.
- As a consequence, there is no parameter interdependence (weights & biases).
- Hidden nodes are created at random and are not iteratively tuned.
- Need to be learn- free parameter between hidden & output layer.
- Efficient and reach global optimum.
- Even randomly produced hidden nodes, this maintains the universal approximation capacity of SLFNs.
- Generalization performance better.
- ELM's generalisation ability outperforms SVM and its variants such as LS-SVM.
- Least human intervention
- Fast training speed and learning accuracy.
- RKS (Random Kitchen Sinks) - A subset of ELM that necessitates the construction of the hidden layer on a Fourier basis.
- ELM trains SLFN in 2 stages:
 - Feature mapping is random
 - Solution provided by linear parameters
- Moore-Penrose generalised inverse of H yields the optimal solution.
- ELM employ random feature mapping whereas SVM use kernel function for feature mapping and Deep NN use RBM/ auto encoder/ auto decoder for feature mapping.
- Crux of ELM
 - Generalization performance (min. training error & smallest norm of wt.)
 - Universal approximation capability
 - Learning without the hidden nodes iteratively tuned
 - Uniform theory of learning
- Learning principle of ELM [52]-
 - I) Huang (2014) demonstrate that hidden neurons of SLFNs with almost any nonlinear piecewise continuous activation function or their linear combinations can be generated at random using any continuous sampling distribution probability, and such hidden neurons can be

independent of training samples as well as their learning environment. Learning stability is also taken into account by the ELM learning network and generalization performance which have been omitted by most conventional learning algorithms when they were first time proposed.

- II) (Huang, 2014) The output standards of generalised SLFNs must be smaller with certain optimization restraints in order to preserve system stability and generalisation performance.
- III) (Huang, 2014) As per perspective of optimisation, SLFN output nodes should be free of biases (or set bias zero). This principle is distinctive from the NN community's broad consensus of that hidden nodes of output require bias (Schmidt et al., 1992; White, 1989, 2006). Biases at the output nodes lead to inadequate (suboptimal) solutions, from the perspective of optimization.

- ELM has also been used in microarray data classification, where it outperformed SVM.
- Input data is non-linearly embedded in a higher dimensional space.
- Adapted without an increase in training time to multiclass classification and multi-output regression.
- Data separation improves as the dimensionality of the data increases.

3. ELM Theories

This section provides a brief description of ELM's theory of interpolation, universal approximation and generalisation bound.

3.1 Interpolation theory

The following theorem explains how ELM's learning capabilities can be explained from an interpolating standpoint:

Theorem 1 (Huang, Zhu, et al., 2006). *Given any small positive value $\epsilon > 0$, any activation function which is infinitely differentiable in any interval, and N arbitrary distinct samples $(\mathbf{x}_i, \mathbf{t}_i) \in \mathbf{R}^d \times \mathbf{R}^m$, there exists $L < N$ such that for any $\{\mathbf{a}_i, \mathbf{b}_i\}_{i=1}^L$ randomly generated from any interval of $\mathbf{R}^d \times \mathbf{R}$, according to any continuous probability distribution, with probability one, $\|\mathbf{H}\beta - \mathbf{T}\| < \epsilon$. Furthermore, if $L = N$, then with probability one, $\|\mathbf{H}\beta - \mathbf{T}\| = 0$.*

Description: Above assertion shows that an ELM network exists that provides reasonably tiny training errors in the sense of a squared error, and this applies to each particular training set, which means that the total number of hidden neurons does not surpass the total number of training sets. Simply, it convey the message that if there are both equal numbers of separate training samples and hidden neurons, the training error will decrease to zero with probability one. So, from an interpolation theory point of view, ELM fits perfectly to each sample, with sufficient quantities of hidden neurons.

3.2 Universal approximation capability

In SLFNs, during training phase, hidden neuron parameters must be adjusted. It is often said that the activation function of hidden neurons is continuous and differentiable. However, without being trained, the hidden layer parameters are produced at random using ELM, regardless of the universal learner.

Theorem 2 (Huang & Chen, 2007, 2008; Huang, Chen, et al., 2006). Given any nonconstant piecewise continuous function $G : \mathbf{R}^d \rightarrow \mathbf{R}$, if $\text{span} \{G(\mathbf{a}, b, \mathbf{x}) : (\mathbf{a}, b) \in \mathbf{R}^d \times \mathbf{R}\}$ is dense in L^2 , for any continuous target function f and any function sequence $\{G(\mathbf{a}_i, b_i, \mathbf{x})\}_{i=1}^L$ randomly generated according to any continuous sampling distribution, $\lim_{L \rightarrow \infty} \|f - f_L\| = 0$ holds with probability one if the output weights β_i are determined by ordinary least square to minimize $\|f(\mathbf{x}) - \sum_{i=1}^L \beta_i G(\mathbf{a}_i, b_i, \mathbf{x})\|$.

According to the above theorem, there is no requirement for a continuous and differentiable activation function in the case of ELM; instead, a threshold and certain additional activation functions can be used and these other functionalities, together with the most frequently utilised activation function, meet requirements for the universal approximation of ELM.

Theorem 3 (Huang, Zhou, et al., 2012). Given any feature mapping $\mathbf{h}(\mathbf{x})$, if $\mathbf{h}(\mathbf{x})$ is dense in $C(\mathbf{R}^d)$ or in $C(\Omega)$, where Ω is a compact set of $\mathbf{R}^d$, then a generalized SLFN with such a random hidden layer mapping $\mathbf{h}(\mathbf{x})$ can separate arbitrary disjoint regions of any shapes in $\mathbf{R}^d$ or Ω .

The theorem demonstrates that, for classification tasks, if the number of hidden layer nodes is large enough, ELM approximates any complex decision limit.

3.3 Generalization bound

Different techniques used to learn general machine ability are Bayesian framework, statistical learning theory (SLT) and cross validation. VC (Vapnik-Chervonenkis) dimensional theory in the SLT was one of the most frequently used generalisation bond analytical frameworks. Vladimir Vapnik and Alexey Chervonenkis originated this VC dimension. The algorithm is described as the measure of an algorithm's capacity for statistical classification and gives the cardinality of the greatest number of points.

The test error expectation can be specified as-

$$R(\alpha) = \int l(y, f(x, \theta)) d\mathbf{p}(x, y) \dots \dots \dots (12)$$

Where, f = Prediction function

θ = Its parameters

l = Loss function

$\mathbf{p}(x, y)$ = Cumulative distribution of probabilities creating training and testing samples.

In practice, actual risk ($R(\theta)$) is difficult to compute. So, we compute empirical risk as-

$$R_{\text{emp}}(\theta) = \frac{1}{N} \sum_{i=1}^N l(y_i, f(x_i, \theta)) \dots \dots \dots (13)$$

The following inequalities holds with probability $(1-\eta)$ given as-

$$R(\theta) \leq R_{\text{emp}}(\theta) + \frac{\sqrt{v \left(\log \left(\frac{2N}{v} \right) + 1 \right) - \log \left(\frac{\eta}{4} \right)}}{N} \dots \dots \dots (14)$$

Where, v = Learning machine's VC dimension

The inequality mentioned above shows “VC bound” or an upper bound. SRM i.e. structure risk minimization is the minimization of VC bound. To guarantee better generalization performance on test set as from SRM point of view, an algorithm should attain low training error on training set as well as get lower VC dimension. Some traditional techniques (SVM, BP etc.) achieve acceptable performance on training sets but, computation of VC dimension of these algorithms are very cumbersome problem, sometimes it reached to infinity.

In case of ELM, it is very favourable to calculate generalization bound because the VC dimension is equal to the number of hidden neurons with probability one in this case.

Theorem 4 (Liu, Gao, et al., 2012). *The VC dimension of ELM with L hidden nodes which are infinitely differentiable in any interval is equal to L with probability one.*

Under the SRM framework, ELM is considered as an ideal classification model from collaboration of **Theorem 1** (ELM can result in a low training set approximation error) and **Theorem 4** (The VC dimension of ELM is very small). It can be analysed in case of ELM that, some activation function like Gaussian type function may deform its generalization performance in regression tasks, but by employing regularization methods (Huang, Zhou, et al., 2012; Lin et al., 2014) this problem can be sorted. The generalization ability can be computed by using LGEM i.e. localized error model [11].

4. Application Area of ELM

ELM contribute its major role in the area especially in classification, regression, feature learning, compression and clustering (as shown in Fig 13). Some specific applications comes in these categories are shown in Fig 14.

Fig 13: Application area of ELM

Fig 14: Some specific applications of ELM

4.1 Classification

In the field of machine learning, it comes under supervised learning and it incorporates to classify given set of data into classes (as shown in Fig 15).

Fig 15: Classification in machine learning

Some of the applications of ELM in classification area are: Guang-Bin Huang et al. [12] gives an idea of multiclass classification using ELM which is advantageous over LS-SVM and PSVM because these traditional approaches are not directly applied to multiclass classification. In order to be classified in multi-label face recognition applications, Weiwei Zong et al. [13] introduced OAA (one against all) and OAO (one against one) ELM. Wu Jun et al. [14] used ELM in image classification to quickly train positive and negative fuzzy rule systems. Yanpeng Qu et al. [15] used the concept of ELM in mammographic risk analysis. Qi Yuan et al. [16] implemented Epileptic ECG classification based on ELM. Sergio Decherchi et al. [17] performed digital implementation with ELM for classification. Hongming Zhou et al. [18] carried out credit risk evaluation with ELM. Mahesh Pal et al. [19] implemented remote sensing image classification using ELM. For speaker recognition, Yuan Lan et al. [20] proposed an extreme learning machine approach. Deepa et al. [21] implemented classification for brain tumour in 3D MR image with ELM. Wenbin Zheng et al. [22] performed text recognition using ELM. Wenbin Zheng et al. [23] implemented spectroscopy based food classification with ELM. Xiaoxuan Lu et al. [24] applied robust ELM to indoor positioning. Xiaodong Li et al. [25] suggested multiple-kernel-learningbased ELM for classification design. Shuliang Xu et al. [26] implemented data stream classification with dynamic ELM. Zaher Mundher et al. [27] carried out implementation of ELM for river flow forecasting. Dibyasundar Das et al. [28] used ELM in handwritten character recognition. Honghao Zhu et al. [29] implemented ELM in credit card fraud detection. ELM concept has been applied by Mohammed Attique et al. [30] in the classification of COVID-19 infections caused by normal Chest CT scans.

4.2 Regression

It is also comes under supervised learning and basically adopted for predicting continuous values (as shown in Fig 16).

Fig 16: Types of regression in machine learning

A solution of optimization approximation was given by Yubo Yuan et al. [31] for ELM-based regression problems. F.L. Chen et al. [32] used the concept of gray extreme learning machine in sales forecasting system. Yimin Yang et al. [33] applied bidirectional ELM for regression problem. Guorui Feng et al. [34] used evolutionary selection ELM for regression. Jiuwen Cao et al. [35] implemented self-adaptive evolutionary ELM for these type of problems. Yoan Miche et al. [36] used regularised ELM to solve regression problems with missing data. Parallel ELM for regression based on MapReduce was proposed by Qing He et al. [37]. For regression, Guoqiang Li et al. [38] used the notion of enhanced ELM based on ridge regression. S. Balasundaram et al. [39] applied the Newton method notion to a 1-Norm ELM for regression and multiclass classification. Outlier-robust ELM for regression problems has been recommended by Kai Zhang et al. [40]. Jarley Palmeira et al. [41] developed an online sequential ELM which is Kalman filter-driven, especially for regression problems. For regression and multi-class classification, Li Ying et al. [42] suggested an orthogonal incremental extreme machine. Using an extreme learning machine, Xiong Luo et al. [43] recommended L1- and L2-norm regression and classification. Shen Yuong et al. [44] used an extreme learning machine with constrained optimization for noisy data regression. Tassadaq et al. [45] used ELM in speech enhancement. For regression, Jie Zhang et al. [46] suggested using a residual compensation extreme learning machine. Zaher et al. [47] implemented ELM in predicting compressive strength of lightweight foamed concrete. Yong Shi et al. [48] recommended fast learning machine for ordinal regression. Ahmed et al. [49] used ELM model in water network management. ELM was used by Zhenglei et al. [50] in Modeling colour fading ozonation of reactive-dyed cotton. Using ELM combinations and the support vector regression Turan et al. [51] have used numerous spray coating strategies to detect wear losses in magnesium-coated alloys.

5. ELM approaches and paper contemplate

5.1 ELM Variants or Proactive approaches of ELM

ELM implemented in various applications and the researchers give various algorithms associated with the concept of ELM and proved that extreme learning machine based techniques are better than conventional approaches. Some of the ELM variant are listed in Table 1. Some of the journal papers detail are listed in Table 2 which includes advantages of technique used in papers and gaps or future work associated with papers.

Table 1: Different proactive approaches of ELM

S.No.	Reference No.	ELM Variant Name	Proposed year	Description & Purpose
1.	[11]	I-ELM	2008	Incremental ELM (Improving compactness)
2.	[11]	EI-ELM	2008	Enhanced Incremental ELM (Improving compactness)
3.	[11]	EM-ELM	2009	Error Minimized ELM (Improving compactness)
4.	[11]	P-ELM	2008, 2010	Pruning ELM (Improving compactness)
5.	[11]	2-stage ELM	2011	Two-stage ELM (Improving compactness)

6.	[2]	B-ELM	2012	Bidirectional ELM (Improving compactness)
7.	[2]	A-ELM	2012	Adaptive ELM (Improving compactness)
8.	[2]	Extended 2-stage ELM	2013	Extended two-stage ELM with regularized framework (Improving compactness)
9.	[2]	ELM with RP	2013	ELM with Random Projection (Improving compactness)
10.	[2]	S-ELM	2014	Sparse ELM (Improving compactness)
11.	[2]	ELM with Sparse Bayesian Approach	2014	ELM with Sparse Bayesian approach (Improving compactness)
12.	[2]	ELM with Recursive Orthogonal Least Square	2014	ELM with recursive orthogonal least square (Improving compactness)
13.	[11]	OS-ELM	2006	Online sequential ELM (For online- sequential data)
14.	[11]	OS-Fuzzy- ELM	2009	Online sequential Fuzzy based ELM (For online- sequential data)
15.	[11]	FOS-ELM	2012	Forgetting mechanism to online sequential ELM (For online- sequential data)
16.	[11]	Extended OS-ELM	2013	Extended online sequential ELM to train time-varying data (For online- sequential data)
17.	[2]	W-ELM	2013	Weighted ELM (For imbalance data)
18.	[2]	Iteratively reweighted ELM	2013	Iteratively re-weighted ELM (For imbalance data)
19.	[2]	Extended W-ELM	2014	Extended Weighted ELM to semi-supervised learning problems (For imbalance data)
20.	[2]	FIR-ELM	2011	Finite Impulse Response Filter ELM (For noisy/missing data)
21.	[2]	DFT-ELM	2012	Discrete Fourier Transform ELM (For noisy/missing data)
22.	[2]	IRWLS-ELM	2013	Iteratively Reweighted Least Squares ELM (For noisy/missing data)
23.	[2]	MLTS-ELM	2013	Multivariate Least Trimmed Squares ELM (For noisy/missing data)
24.	[2]	RMLTS-ELM	2013	Reweighted Multivariate Least Trimmed Squares ELM (For noisy/missing data)

25.	[2]	TROP-ELM	2013	Tikhonov Regularized Optimally Pruned ELM (For noisy/missing data)
26.	[11]	EN-ELM	2009	Ensemble ELM (For overfitting problem)
27.	[11]	EOS-ELM	2009	Ensemble of Online Sequential ELM (For overfitting problem)
28.	[11]	V-ELM	2009	Voting based ELM (For overfitting problem)
29.	[2]	Pointwise-RankELM	2014	Pointwise Rank ELM (For ranking)
30.	[2]	Pairwise-RankELM	2014	Pairwise Rank ELM (For ranking)
31.	[2]	SS-ELM	2014	Semi Supervised ELM (For semi-supervised learning)
32.	[2]	Bayesian- ELM	2011, 2014	Bayesian ELM (For semi-supervised learning)
33.	[2]	Fuzzy- ELM	2011, 2012, 2013	Fuzzy-ELM (For semi-supervised learning)
34.	[2]	Wavelet- ELM	2010, 2011, 2012	Wavelet ELM (For semi-supervised learning)
35.	[11]	Complex- ELM	2008, 2012	Complex ELM (For semi-supervised learning)
36.	[2]	Genetic/ Evolutionary ELM	2011, 2012, 2013	Genetic or Evolutionary ELM (For semi-supervised learning)
37.	[2]	US-ELM	2014	Unsupervised Extreme Learning Machine (For embedding and clustering)
38.	[2]	SAE & SAD	2013	Stacked Auto-encoder and Stacked Auto-decoder (For representational Learning)
39.	[2]	FS-ELM	2013	Feature Selection Extreme Learning Machine (For feature selection)
40.	[2]	Parallel implementation of ELM	2011, 2013	Parallel implementation of ELM with MapReduce & CPU implementation, cloud computing.
41.	[2]	Hardware implementation of ELM	2012,2013	Hardware implementation of ELM include silicon implementation
42.	[7]	T-ELM	2016	Two Hidden layer ELM
43.	[7]	M-ELM	2017	Multiple Hidden layer ELM (For large datasets)

44.	[55]	HELM	2016	Hierarchical Extreme Learning Machine (For big data process)
45.	[55]	dp-ELM	2019	Distributed and parallel ELM (For big data process)
46.	[55]	dp-HELM	2019	Distributed and parallel Hierarchical ELM (For big data process)
47.	[59]	ELM based Modified AdaBoost.RT	2010	Ensemble ELM based on modified AdaBoost.RT (For regression problem)
48.	[61]	eF-OP-ELM	2010	Evolving Fuzzy Optimally Pruned ELM (For Compactness of network)
49.	[65]	EOSELM	2009	Ensemble-Online sequential ELM
50.	[65]	MAS-OSELM-BF	2011	Multi Agent System based on Online Sequential ELM and Bayesian Formalism
51.	[6]	GSO-ELM	2011	Group Search Optimizer based ELM (For continuous optimization problem)
52.	[67]	LP-ELM	2011	Linear Programming ELM (For optimization problem)
53.	[68]	HS-ELM	2012	Harmony Search ELM (For classification problems)
54.	[53]	R-ELM	2011	Regularized ELM (For remove irrelevant hidden nodes and compactness of network)
55.	[11]	EELM	2011	Effective ELM (For robustness of network and improves learning rates)
56.	[26]	DELM	2017	Dynamic ELM (For data stream classification)
57.	[38]	RRELM	2011	Ridge-regression ELM (For regression)
58.	[42]	OI-ELM	2013	Orthogonal Incremental ELM (For regression and multi-class classification)
59.	[44]	CO-ELM-R	2015	Constrained Optimization based ELM for Regression (For noisy data regression)
60.	[46]	RC-ELM	2018	Residual Compensation ELM (For regression)

Table 2: Details of some papers included in survey

S.No.	Reference No.	Year	Technique used	Advantages	Gap/ Future work
-------	---------------	------	----------------	------------	------------------

1.	[57]	2010	OP-ELM, ELM	<ul style="list-style-type: none"> -ELM-1) Hidden neurons weight selection fast. 2) Short training and computational time 3) Simple algo. / easy implementation 4) Minimum training error -OP-ELM-1) More robust 2) Manage multi-output multiclass regression and classification problems 3) Faster Convergence 4) Faster than conventional methods 5) Same accuracy -SVM, MLP or GP 6) Small response time 	<ul style="list-style-type: none"> -ELM-1) Issues with irrelevant or correlated data -OP-ELM-1) Slower than ELM 2) impact on data set size or problem type
2.	[60]	2010	Location Awareness, ANN, ELM	<ul style="list-style-type: none"> -ELM-1) Fast learning speed 2) Better generalization performance 3) Simple algo 4) Overcome - local minima, improper learning rate, overfitting etc. 5) Non-differential activation function 6) Efficient & accurate predictive accuracy 	<ul style="list-style-type: none"> 1) Accuracy comparison with only one algo. 2) Comparison with other NN & accuracy with Kalman filter
3.	[65]	2011	OSELM, EOSELM, MAS-OSELM-BF	<ul style="list-style-type: none"> -OSELM-1) Get Feedforward incremental learning network 2) Extended version of ELM 3) Works with additive-sigmoid and RBF hidden neurons -EOSELM-1) Better than OSELM -MAS-OSELM-BF-1) overcome demerit of EOSELM 2) Achieve better accuracy rates 	<ul style="list-style-type: none"> -EOSELM-Treat all OSELM equally without considering error. -MAS-OSELM-BF-1) Studies needed from various application domain 2) Hybrid system using GA
4.	[71]	2015	Bing + ELM	<ul style="list-style-type: none"> -ELM- 1) Shorter training time Proposed scheme-1) Better performance than SVM 2) Overcome slow learning speed, human intervening, and poor computational stability. 	<ul style="list-style-type: none"> Proposed scheme-1) Cannot distinguish close together pedestrians 2) Certain candidates, such as trees, cars, lamps and streets, are rendered pedestrians
5.	[45]	2017	ELM, HELM	<ul style="list-style-type: none"> -ELM- 1) Fast training 2) Good generalization performance 3) Universal approximation capability 4) Not required parameters massive tuning -H-ELM- 1) Effective 2) Faster training speed 	<ul style="list-style-type: none"> HELM- 1) Only used on a small set of training data. 2) Include multi-task training & transfer learning 3) Formulate training criteria that are noise and SNR-aware.

6.	[55]	2019	dp-ELM, dp-HELM	-dp-ELM-1) Quality prediction model 2) Efficient in model training -dp-HELM-1) Large scale of data 2) Accurate prediction performance 3) Multimode process	-dp-HELM- 1) Modelling efficiency lower than dp-ELM 2) Slower 3) Extended to Plant wise processes
7.	[28]	2019	ELM	1) Faster 2) Smallest training error 3) Higher accuracy 4) Best ELM configuration used in classification and regression datasets	1) Investigation with deep ELM architecture 2) Employment of ELM variants in CNN
8.	[48]	2019	FKELMOR	1)Efficient in both small scale and big data scenario 2)Fast learning 3)Better generalization and stable 4)Reduced time complexity 5)Unified framework	1) Used model in managerial applications
9.	[29]	2020	WELM	1) Better classification performance	1) Comparison with other intelligent algo. 2) Applied in distributive computing platform.
10.	[51]	2020	SVR, ELM	-ELM- Fast and better prediction result MAE and MSE values better	1)Model implementation with other coating technology 2) Benefit to design engineers and design flexibility

6. Comparison between learning methods

After discussed the various aspects of extreme learning machine, there is need to know about some basic difference between biological learning, conventional learning and ELM. The performance of ELM is higher than regular learning (ex: BP) and very much like biological learning when compared to these three studying methods. As already before discussed the prominent features of ELM which makes it highlighted now a days. Here, in Fig 17, comparison between learning algorithms considering features like stability, parallelization, hardware implementation, involvement of humans, complexity, online sequential learning, speed, accuracy and its model.

Also, Fig 18 depicts the distinction between deep learning, from which everyone is familiar and used in most of the applications and the new learning technique, ELM. The comparison is done between these two techniques on the basis of features like amount of involvement of human, sensitivity to the network, suitability for which scale applications, resources requirement, its parallel and hardware implementation, suitability in micro level learning and online incremental learning and complexity related to applications. From this Fig 18, we conclude that ELM is better than deep learning in most of the circumstances and also, ELM gives faster results.

Fig 17: Comparison between biological learning, conventional learning and ELM

Fig 18: Distinction between Deep learning and ELM

7. Comparative Analysis and Suggested models

7.1 Comparative analysis of data

Table [3] provides a comparison of various learning algorithms with ELM. In [6], researcher compared ELM approaches and finds that the proposed algorithm (GSO-ELM) accuracy better than traditional ELM (as shown in Fig 19). Researchers in [9] find that the highest detection rate is achieved by 1200 hidden nodes, random block versions of 200 to 400 and sine activation by means of a faster OS-ELM than all the works in Malayalam's handwritten character recognition (as shown in Fig 20). The investigator in [10] finds that ELM performs better than others on the same dataset implemented in real time to detect blackout risk in comparison with different methods (as shown in Fig 21). In [11], researchers discovered that performance of ELM- AE is better or similar to other comparative methods but ELM-AE takes less training time than others (as shown in Fig 22). The researchers found in [12] that ELM can always deliver a similar performance with much more rapid learning speed than SVM and LS-SVM (as shown in Fig 23). Experiments in [16] revealed that as compared to BP network and SVM classification, ELM is more accurate and consumes much less time (as shown in Fig 24). In [18], by the use of two different dataset researcher conclude that the accuracy of ELM is more as compared to SVM (as shown in Fig 25). The findings in [19] show that the KELM encourages the classifying precision performance of the various data sets used in this study in relation to the SVM. As with the SVM, a probabilistic result is not produced by the KELM. On the other hand, KELM used all pixels compared to the SVM, producing a patchy solution requiring less classification pixel training (as shown in Fig 26). In [21], tumor segmentation results when compared with literature shows promising results. In contrast to Vector support, the most widely used ANN for medical classification images, the extreme learning machine has improved grading (as shown in Fig 27). In [23], researchers conclude that on all four datasets, ELM has a relatively large precision (as shown in Fig 28). Comparisons with the three current LM, CFWNN-ELM and SVM approaches demonstrated in [35] that SaE-ELM can improve network generalization performance effectively (as shown in Fig 29). In [38], scientists demonstrate that RR-ELM has good performance and stability, reducing adverse effects of interruptions and multicollinearities in the linear model in an efficient manner. RR-ELM mainly addresses problems of regression. The experiment shows in [46] that RC-ELM shows good overall performance and robustness, and a generalized framework to address the regression problems can also be suggested by the RC-ELM. The results of the experiment in [56] show that in terms of accuracy, in comparison of original ELM as well as face recognition classifiers and several classification tasks, EN-ELM performance is remarkable (as shown in Fig 30). In [65-106], EOSELM efficiency is higher with regards to accuracy, as the results show. (as shown in Fig 31).

Table 3: Relative survey analysis of papers

S. No.	Paper reference	Dataset	Model used	No. of hidden nodes	Testing accuracy (%)	Training time (s)	Activation function	Method complexity
1.	[6]	Cancer	ELM	10	96.09±1.48	0.0014	Sigmoid function	Little
			PSO-ELM	5	96.70±1.05	32.31	Sigmoid function	Intermediate
			GSO-ELM	5	96.90±1.09	60.11	Sigmoid function	Intermediate
		Diabetics	ELM	10	75.89±2.47	0.0014	Sigmoid function	Little
			PSO-	10	76.50±2.22	81.52	Sigmoid	Intermediate

			ELM				function	
			GSO-ELM	10	77.16±2.19	64.00	Sigmoid function	Intermediate
2.	[9]	Malayalam Characters	OS-ELM	1200	94.68	92.90	Sigmoid	Little
			OS-ELM	1200	79.37	78.17	Hardlim	Little
			OS-ELM	1200	72.01	137.24	RBF	Little
			OS-ELM	1200	95.02	95.64	Sine	Little
3.	[10]	Dynamic security assessment	ELM	85	98.7	0.0942	Sine	Little
			DT	85	98.9	139.694	-----	Intermediate
			SVM	85	98.1	1592.89	-----	Huge
			RBFNN	85	97.4	321.687	-----	Huge
4.	[11]	MNIST	ELM-AE	10000	99.03	444.655	Sigmoid	Little
			DBN	10000	98.87	20,580	-----	Intermediate
			DBM	10000	99.05	68,246	-----	Huge
			SAE	10000	98.6	>17h	-----	Intermediate
			SDAE	10000	98.72	>17h	-----	Intermediate
5.	[12]	Diabetes, Australian, colon, credit, colon etc...	SVM	1000	84.51	29.382	-----	Intermediate
			LS-SVM	1000	84.71	5.5703	-----	Intermediate
			ELM (Gaussian Kernel)	1000	84.58	3.3116	Gaussian	Little
			ELM (Random sigmoid hidden nodes)	1000	84.59	0.4246	Sigmoid	Little
6.	[16]	EEG images	ELM	10	96.00±0.50	0.0803	Sigmoid	Little
			BP	10	95.50±0.50	1.6363	Sigmoid	Huge
			SVM	-----	95.25±0.25	12.6410	-----	Intermediate
7.	[18]	Australian credit, German credit	ELM (Gaussian kernel)	-----	92.53	-----	-----	Little
			SVM (Gaussian kernel)	-----	90.68	-----	-----	Intermediate
8.	[19]	ATM, DIAS	SVM	-----	92.5, 92.0	1.30, 40.78	-----	Intermediate
			KELM	-----	94.1, 92.2	0.17, 1.30	-----	Little
9.	[21]	Harvard benchmark	BPN	34	86.25	65	-----	Huge
			SVM	8	91.2	75	-----	Intermediate
			ELM	60	93.2	45	Sigmoid	Little
10.	[23]	Coffee, Meat, Oil, Fruit	KNN	-----	98.21 ± 0.00, 92.31 ± 2.20, 83.25 ± 3.23, 67.01 ± 1.33	-----	-----	Intermediate
			LS-SVM	-----	97.50 ± 1.73, 95.83 ± 2.98, 95.08 ± 3.92, 96.01 ± 1.88	-----	-----	Intermediate
			BP-ANN	-----	94.82 ± 4.49, 95.65 ± 2.76,	-----	-----	Huge

					90.50 ± 3.45, 95.33 ± 1.79			
			ELM	-----	100.00 ± 0.00, 97.78 ± 3.41, 97.35 ± 3.65, 95.05 ± 3.15	-----	Sigmoid	Little
11.	[35]	Disease, Iris, Wine, Protein sequence	SaE-ELM	18,18,22, 80	82.53, 97.20, 97.95, 97.53	0.7086, 0.2289, 0.3050, 43.6344	-----	Little
			LM	20,10,20, 60	71.75, 95.40, 92.97, 87.45	0.3066, 0.3641, 0.4988, 13.26	-----	Intermediate
			CFWNN-ELM	30,20,30, 130	76.85, 95.92, 95.15, 87.48	0.0016, 0.0009, 0.0028, 0.0859	-----	Little
			SVM	81,23,47, 261	76.10, 94.36, 97.48, 90.63	6.7094, 4.5488, 5.8941, 274.9672	-----	Intermediate
12.	[38]	Servo, Abalone, Boston housing	ELM	30,30,50	-----	0.0001, 0.0102, 0.0078	Sigmoidal	Little
			I-ELM	2000, 1840, 2000	-----	0.1398, 0.4820, 0.1844	Sigmoidal	Little
			EM-ELM	28,11,28	-----	0.0125, 0.0117, 0.0219	Sigmoidal	Little
			RR-ELM	30,30,50	-----	0.0091, 0.0328, 0.0165	Sigmoidal	Little
13.	[46]	Blast volume, blast temperature, blast pressure, top pressure, top temperature, permeability index, and oxygen enrichment percentage.	RC-ELM	500	-----	0.0791	Sigmoidal	Little
			ELM	500	-----	0.0125	Sigmoidal	Little
			SVM	-----	-----	9.3166	-----	Intermediate
			BPNN	700	-----	5.3981	-----	Huge
14.	[56]	Combo	ELM	100	81.82±1.35	0.0672	Sigmoidal	Little
			E-ELM	100	81.90±1.23	88.641	Sigmoidal	Little
			k-NN	-----	79.65±1.55	1.0133	k = 1	Intermediate
			BP	100	78.85±1.63	570.35	Log-sigmoid	Huge
			EN-ELM	100	84.67±1.55	32.488	Sigmoidal	Little
		Satellite	ELM	200	88.10±0.32	0.8297	Sigmoidal	Little
			E-ELM	200	88.44±0.62	4324.9	Sigmoidal	Little
			k-NN	-----	85.82±0.79	15.013	k = 1	Intermediate
			BP	200	83.08±0.97	30538	Log-sigmoid	Huge
			EN-ELM	200	88.86±0.21	443.27	Sigmoidal	Little
		Segmentation	ELM	200	94.83±0.80	0.4180	Sigmoidal	Little
			E-ELM	200	95.46±1.24	1440.8	Sigmoidal	Little

15.	[65]		k-NN	-----	94.32±0.92	0.9859	k = 1	Intermediate
			BP	200	85.59±1.06	15316	Log-sigmoid	Huge
			EN-ELM	200	96.19±0.28	210.83	Sigmoidal	Little
		Satellite Image segmentation Monik-1	OSELM	400	88.90	-----	Sigmoidal	Little
			EOSELM	400	89.01	-----	Sigmoidal	Little
			OSELM	180	94.82	-----	Sigmoidal	Little
			EOSELM	180	94.79	-----	Sigmoidal	Little
			OSELM	70	78.74	-----	Sigmoidal	Little
			EOSELM	70	78.77		Sigmoidal	Little

Fig 19: Variation of testing accuracy vs number of hidden nodes of cancer dataset

Fig 20: OSELM testing accuracy (%) performance on basis of activation function

Fig 21: Performance of different algorithms based on training time

Fig 22: Performance of various algorithms based on training time and testing accuracy

Fig 23: Performance of different algorithm on same datasets based on training time

Fig 24: Performance evaluation of different approaches based on testing accuracy and training time

Fig 25: Testing accuracy performance of ELM and SVM

Fig 26: Performance of SVM and KELM on ATM dataset

Fig 27: Performance evaluation based on training time and testing accuracy

Fig 28: Testing accuracy performance of various algorithms on coffee dataset

Fig 29: Performance evaluation on Iris dataset of various approaches

Fig 30: EN-ELM performance comparison with other algorithms

Fig 31: Performance evaluation on satellite image dataset of OSELM AND EOSELM

7.2 Suggested models

After examining Table 3, ELM will be considered as a nice option for machine learning applications like classification, regression, feature learning and act as a good alternative of conventional learning techniques especially in case of accuracy, training time, complexity and tuning of hidden node parameters. There are various variants of ELM available which can be used in various applications. As we have saw in above table [3] and its related figures that ELM performance in most of the applications is outstanding or remarkable, also takes lesser training time as compared to existing approaches, these two main features makes ELM attractive now a days. Besides these, its parallel and hardware implementation is easy because in this we do not need huge amount of resources, easy in micro level learning and online incremental learning, not sensitive to network size, almost free of human intervention and suitable for all the type of applications.

There are four suggested model which may useful in most of the applications and may solved problems associated with it-

- *Fuzzy with HELM*- This implementation may handle imprecise and uncertain situation. In brief, this model can be manage uncertainty.
- *Right evaluation metrics with WELM*- This model may effectively handle imbalance data. Evaluation metrics can be confusion metrics, F₁ score, Gain & Lift charts, Log loss, Gini coefficient, cross validation etc.
- *Polishing with HELM*- Polishing or data scrubbing or relabelling with HELM may useful in removing both class & attribute noise.
- *Gamma classifier with ELM*- This model may be very useful in data stream classification.

8. Conclusion

In machine learning and data analysis, SVM and NN (especially classification) play a significant role, however they do face some challenge, namely rigorous human involvement, reluctant learning speed and inadequate scalability of learning. These problems are solved by ELM and it does not work for all the kernels i.e. if the kernel satisfies universal approximation then ELM works; also in this learning technique, we do not have to determine the number of hidden layers, learning rate and other hyper-parameters in advance. Some of the most significant benefits of ELM are as follows: Learning possible without iterative tuning for generalised SLFNs; it offers a polarized learning paradigm for regression, classification, feature learning, clustering and compression applications; also incorporates small training error and obtain the smallest norm of weights which results in good generalization performance. ELM works with various hidden nodes, including kernels. This paper covered nearly all related aspects of the new eminent learning technology, the extreme learning machine. More research is needed in the field of ELM so that its pros and cons can be discussed in further papers and use the concept of ELM in most of the applications so that it can take its advantages.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent to publish

Not applicable.

Availability of data and materials

The datasets generated during and/or analyzed during the current study are not publicly available, but are available from the corresponding author who was an organizer of the study.

Competing interests

The authors declare that they have no competing interests.

Funding

No Funding till date but we applied for in a government organization.

Authors' contributions

All of the authors contributed to the conception of the study, drafting and critical revision of the manuscript, and provided final approval of the manuscript.

References

- [1] <https://brilliant.org/wiki/feedforward-neural-networks/>
- [2] Gao Huang, Guang-Bin Huang, Shiji Song, Keyou You, "Trends in extreme learning machine: a review", *Elsevier*, vol. 61, pp. 32-48, Jan. 2015.
- [3] <https://www3.ntu.edu.sg/home/egbhuang/ELM2016/index.html>
- [4] Oyekale Abel Alade, Ali Selamat, Roselina Sallehuddin, "A Review of Advances in Extreme Learning Machine Techniques and Its Applications", *Recent Trends in Information and Communication Technology. IRICT 2017, Lecture Notes on Data Engineering and Communications Technologies*, vol 5. SpringerCham.
- [5] <https://www3.ntu.edu.sg/home/egbhuang/pdf/ELM-Tutorial.pdf>
- [6] D. N. G. Silva, L. D. S. Pacifico, T. B. Ludermir, "An Evolutionary Extreme Learning Machine Based on Group Search Optimization", *2011 IEEE Congress of Evolutionary Computation (CEC)*, pp. 574-580, Jun. 2011.
- [7] Dong Xiao, Beijing Li, Yachun Mao, "A Multiple Hidden Layers Extreme Learning Machine Method and Its Application", *Mathematical Problems in Engineering*, vol. 2017, Article ID 4670187, pp. 1-10, Dec. 2017.
- [8] Huang Shu-guang, Zhang Liang, Wang peng-po, Han Hong-wei, "A Captcha Recognition Algorithm Based on Holistic Verification", *2011 First International Conference on Instrumentation, Measurement, Computer, Communication and Control*, pp. 525-528, Oct. 2011.
- [9] Binu P Chacko, Babu Anto P, "Online Sequential Extreme Learning Machine Based Handwritten Character Recognition", *IEEE Technology Students' Symposium*, pp. 142-147, Jan. 2011.
- [10] Y. Xu, Z.Y. Dong, K. Meng, Z. Xu, R. Zhang, Andrew Y. Wu, K. P. Wong, "Earlier Detection of Risk of Blackout by Real-Time Dynamic Security Assessment Based on Extreme Learning Machine", *2010 International Conference on Power System Technology*, pp. 1-7, Oct. 2010.
- [11] Musatafa Abbas Abbood Albadr, Sabrina Tiun, "Extreme Learning Machine: A Review", *International Journal of Applied Engineering Research (IJAER)*, vol. 12, pp. 4610-4623, Nov. 2017.
- [12] Guang-Bin Huang, Hongming Zhou, Xiaojian Ding, Rui Zhang, "Extreme Learning Machine for Regression and Multiclass Classification", in *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 42, no. 2, pp. 513-529, April 2012 .
- [13] Weiwei Zong, Guang-BinHuang, "Face recognition based on extreme learning machine", *Neurocomputing*, vol. 74, no. 16, pp. 2541-2551, Sept. 2011.
- [14] Wu Jun, Wang Shitong, "Positive and negative fuzzy rule system, extreme learning machine and image classification", Springer (2011).
- [15] Yanpeng Qu, Changjing Shang, Wei Wu, and Qiang Shen, "Evolutionary Fuzzy Extreme Learning Machine for Mammographic Risk Analysis", *International Journal of Fuzzy Systems*, vol. 13, no. 4, Dec. 2011.
- [16] Qi Yuan, Weidong Zhou, ShufangLi, DongmeiCai, "Epileptic EEG classification based on extreme learning machine and nonlinear features", *Epilepsy Research*, vo. 96, no. 1-2, pp. 29-38, Sept. 2011.
- [17] Sergio Decherchi, Paolo Gastaldo, Alessio Leoncini, Rodolfo Zunino, "Efficient Digital Implementation of Extreme Learning Machines for Classification", in *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 59, no. 8, pp. 496-500, Aug. 2012.
- [18] Hongming Zhou, Yuan Lan, Yeng Chai Soh, Guang-Bin Huang, Rui Zhang, "Credit risk evaluation with extreme learning machine", *2012 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 1064-1069, Oct. 2012.
- [19] Mahesh Pal, Aaron E. Maxwell, Timothy A. Warner, "Kernel Based Extreme Learning Machine for remote sensing image classification", *Remote Sensing Letters*, vol.4, no. 9, pp. 853-862, Jun. 2013.
- [20] Yuan Lan, Zongjiang Hu, Yeng Chai Soh , Guang-Bin Huang, "An extreme learning machine approach for speaker recognition", *Neural Computing and Applications*, vol.22, pp. 417-425, May 2012.
- [21] Deepa, S N, Arunadevi, B.Informatologia, "Extreme Learning Machine For Classification of Brain Tumour in 3D MR IMAGES", *Informatol*, vol. 46, no. 2, pp. 111-121, 2013.
- [22] Wenbin Zheng, Yuntao Qian, Huijuan Lu, "Text categorization based on regularization extreme learning machine", *Neural Computing and Applications*, vol. 22, pp. 447-456, Mar. 2013.
- [23] Wenbin Zheng, Xiaping Fu, Yibin Ying, "Spectroscopy-based food classification with extreme learning machine", *Chemometrics and Intelligent Laboratory Systems*, vol. 139, pp. 44-47, Dec. 2014.

- [24] Xiaoxuan Lu, Han Zou, Hongming Zhou, Lihua Xie, Guang-Bin Huang, “Robust Extreme Learning Machine With its Application to Indoor Positioning”, in *IEEE Transactions on Cybernetics*, vol. 46, no. 1, pp. 194-205, Jan. 2016.
- [25] Xiaodong Li, Weijie Mao, Wei Jiang, “Multiple-kernel-learning-based extreme learning machine for classification design”, *Neural Computing and Applications*, vol. 27, pp. 175-184, Jan. 2016.
- [26] Shuliang Xua, Junhong Wang, “Dynamic extreme learning machine for data stream classification”, *Neurocomputing*, vol. 238, pp. 433-449, May 2017.
- [27] Zaher Mundher Yaseen, Sadeq Oleiwi Sulaiman, Ravinesh C.Deo, Kwok-WingChau, “An enhanced extreme learning machine model for river flow forecasting: State-of-the-art, practical applications in water resource engineering area and future research direction”, *Journal of Hydrology*, vol. 569, pp. 387-408, Feb. 2019.
- [28] Dibyasundar Das, Deepak Ranjan Nayak, Ratnakar Dash, Banshidahar Majhi, “An empirical evaluation of extreme learning machine: application to handwritten character recognition”, *Multimedia Tools and Applications*, vol. 78, pp. 19495–19523, Feb. 2019.
- [29] Honghao Zhu, Guanjin Liu, Mengchu Zhou, Yu Xie, Abdullah Abusorrah, QiKang, “Optimizing Weighted Extreme Learning Machines for imbalanced classification and application to credit card fraud detection”, *Neurocomputing*, vol. 407, pp. 50-62, Sept. 2020.
- [30] Muhammad Attique Khan, Seifedine Kadry, Yu-Dong Zhang, Tallha Akram, Muhammad Sharif, Amjad Rehman, TanzilaSaba, “Prediction of COVID-19 - Pneumonia based on Selected Deep Features and One Class Kernel Extreme Learning Machine”, *Computers & Electrical Engineering*, vol. 90, Article ID 106960, Mar. 2021.
- [31] Yubo Yuan, Yuguang Wang, Feilong Cao, “Optimization approximation solution for regression problem based on extreme learning machine”, *Neurocomputing*, vol. 74, no. 16, pp. 2475-2482, Sept. 2011.
- [32] F.L. Chen, T.Y.Ou, “Sales forecasting system based on Gray extreme learning machine with Taguchi method in retail industry”, *Expert Systems with Applications*, vol. 38, no. 3, pp. 1336-1345, Mar. 2011.
- [33] Yimin Yang, Yaonan Wang, Xiaofang Yuan, “Bidirectional Extreme Learning Machine for Regression Problem and Its Learning Effectiveness”, in *IEEE Transactions on Neural Networks and Learning Systems*, vol. 23, no. 9, pp. 1498-1505, Sept. 2012.
- [34] Guorui Feng, Zhenxing Qian, Xinpeng Zhang, “Evolutionary selection extreme learning machine optimization for regression”, *Soft Computing*, vol. 16, pp. 485–1491, Feb. 2012.
- [35] Jiuwen Cao, Zhiping Lin, Guang-Bin Huang, “Self-Adaptive Evolutionary Extreme Learning Machine”, *Neural Processing Letters*, vol. 36, pp. 285–305, Jul. 2012.
- [36] Yoan Miche, Emil Eirola, Markvan Heeswijk, Eric Séverin, Amaury Lendasse, “Regularized extreme learning machine for regression with missing data”, *Neurocomputing*, vol. 102, pp. 45-51, Feb 2013.
- [37] Qing He, Tianfeng Shang, Fuzhen Zhuang, Zhongzh iShi, “Parallel extreme learning machine for regression based on MapReduce”, *Neurocomputing*, vol. 102, pp. 52-58, Feb. 2013.
- [38] Guoqiang Li, Peifeng Niu, “An enhanced extreme learning machine based on ridge regression for regression”, *Neural Computing and Applications*, vol. 22, pp. 803–810, 2013.
- [39] S. Balasundaram, Deepak Gupta, Kapil, “1-Norm extreme learning machine for regression and multiclass classification using Newton method”, *Neurocomputing*, vol. 128, pp. 4-14, Mar. 2014.
- [40] Kai Zhang, Minxia Luo, “Outlier-robust extreme learning machine for regression problems”, *Neurocomputing*, vol. 151, pp. 1519-1527, Mar 2015.
- [41] Jarley Palmeira Nobrega, Adriano L. I. Oliveira, “Kalman filter-based method for Online Sequential Extreme Learning Machine for regression problems”, *Engineering Applications of Artificial Intelligence*, vol. 44, pp. 101-110, Sept. 2015.
- [42] Li Ying, “Orthogonal incremental extreme learning machine for regression and multiclass classification”, *Neural Computing and Applications*, vol. 27, pp. 111–120, 2016.
- [43] Xiong Luo, Xiaohui Chang, Xiaojuan Ban, “Regression and classification using extreme learning machine based on L1-norm and L2-norm”, *Neurocomputing*, vol. 174, pp. 179-186, Jan. 2016.
- [44] Shen Yuong Wong, Keem Siah Yap, Hwa Jen Yap, “A Constrained Optimization based Extreme Learning Machine for noisy data regression”, *Neurocomputing*, vol. 171, pp. 1431-1443, Jan. 2016.
- [45] Tassadaq Hussain, Sabato Marco Siniscalchi, Chi-Chun Lee, Syu-Siang Wang, Yu Tsao, Wen-Hung Liao, “Experimental Study on Extreme Learning Machine Applications for Speech Enhancement”, in *IEEE Access*, vol. 5, pp. 25542-25554, 2017.
- [46] Jie Zhang, Wendong Xiao, Yanjiao Li, Sen Zhang, “Residual compensation extreme learning machine for regression”, *Neurocomputing*, vol. 311, pp. 126-136, Oct. 2018.

- [47] Zaher Mundher Yaseen, Ravinesh C. Deo, Ameer Hilal, Abbas M. Abd, Laura Cornejo Bueno, Sancho Salcedo-Sanz, Moncef L. Nehdi, "Predicting compressive strength of lightweight foamed concrete using extreme learning machine model", *Advances in Engineering Software*, vol. 115, pp. 112-125, Jan. 2018.
- [48] Yong Shi, Peijia Li, Hao Yuan, Jianyu Miao, Lingfeng Niu, "Fast kernel extreme learning machine for ordinal regression", *Knowledge-Based Systems*, vol. 177, pp. 44-54, Aug. 2019.
- [49] Ahmed M. A. Sattar, Ömer Faruk Ertuğrul, B. Gharabaghi, E. A. McBean, J. Cao, "Extreme learning machine model for water network management", *Neural Computing and Applications*, vol. 31, pp. 157–169, Jan. 2019.
- [50] Zhenglei He, Kim-Phuc Tran, Sébastien Thomassey, Xianyi Zeng, Jie Xu, Yi Changhai, "Modeling color fading ozonation of reactive-dyed cotton using the Extreme Learning Machine, Support Vector Regression and Random Forest", *Textile Research Journal*, vol. 90, no. 7-8, pp. 896-908, Apr. 2020.
- [51] Turan Gurgenc, Osman Altay, Mustafa Ulas, Cihan Ozel, "Extreme learning machine and support vector regression wear loss predictions for magnesium alloys coated using various spray coating methods", *Journal of Applied Physics*, vol. 127, no. 18, Article ID 185103, May 2020.
- [52] N Jain, N Rathore, and A Mishra "A Hybrid Extreme Learning Machine With Fuzzy Sigmoid Kernel-Based Efficient Image Forgery Detection For Both Copy Move And Spliced Image ", *Cluster Computing Journal, Springer Publication-US*, DOI: 10.1007/s10586-018-2651-8, IF=2.620, 2020. (Accepted)
- [53] Jose M. Martinez,Pablo Escandell-Montero,Emilio Soria-Olivas,Jose D. Martin-Guerrero,Rafael Magdalena Bened-ito,Juan Gomez-Sanchis, "Regularized extreme learning machine for regression problem", *Neurocomputing*, vol. 74, no. 17, pp. 3716-3721, Oct. 2011.
- [54] Yuguang Wang,Feilong Cao,Yubo Yuan, "A study on effectiveness of extreme learning machine", *Neurocomputing*, vol. 74, no. 16, pp. 2483-2490, Sept. 2011.
- [55] Le Yao, Zhiqiang Ge, "Distributed parallel deep learning of hierarchical extreme learning machine for multimode quality prediction with big process data", *Engineering Applications of Artificial Intelligence*, vol. 81, pp. 450-465, May 2019.
- [56] Nan Liu and Han Wang, "Ensemble based extreme learning machine", in *IEEE Signal Processing Letters*, vol. 17, no. 8, pp. 754-757, Aug. 2010.
- [57] Yoan Miche,Antti Sorjamma,Patrick Bas,Olli Simula, christian Jutten and Amaury Laendasse, "OP-ELM: Optimally pruned extreme learning machine", in *IEEE Transactions on Neural Networks*, vol. 21, no. 1, pp. 158-162, Jan. 2010
- [58] Ganglong Duan,Zhiwen Hauang,Jianren Wang, "Extreme learning machine for financial distress prediction for listed company", *2010 International Conference on Logistics Systems and Intelligent Management (ICLSIM)*, pp. 1961-1965, May 2010.
- [59] Hui-Xin Tian,Zhi Zhong Mao, "An ensemble ELM based on modified Adaboost.RT algorithm for predicting the temperature of molten steel in ladle furnace", in *IEEE Transactions on Automation Science and Engineering*, vol. 7, no. 1, pp. 73-80, Jan. 2010.
- [60] Teddy Mantoro,Akeem Olowolayemo,Sunday Olusanya,Olatunji, "Mobile user location determination using extreme learning machine", in *IEEE Transactions on Automation Science and Engineering*, vol. 7, no. 1, pp. 73-80, Jan. 2010.
- [61] Federico Montesino Pouzols, "Evolving fuzzy optimally pruned extreme learning machine: A comparative analysis", *International Conference on Fuzzy Systems*, pp. 1-8, Jul. 2010.
- [62] Guangyong Gao, Guoping Jiang, "A zero watermarking image authentication scheme using Zernike moment and extreme learning machine", *2011 International Conference on Wireless Communications and Signal Processing (WCSP)*, pp. 1-5, Nov. 2011.
- [63] Emilio Soria-Olivas,Juan Gomez-Sanchis, Jose D. Martin,Joan Vila-Frances,Marcelino Martinez,Jose R. Magdalena,Antonio J. Serrano, "BELM: Bayesian extreme learning machine", in *IEEE Transactions on Neural Networks*, vol. 22, no. 3, pp. 505-509, Mar. 2011.
- [64] Mosiuoa M. Sole,M. S. Tsoeu, "Sign language recognition using extreme learning machine", *IEEE Africon '11*, pp. 1-6, Sept. 2011.
- [65] Keem Siah Yap, "A new multi agent system based on online sequential extreme learning machines and Bayesian formalism", *2011 International Conference on Networking, Sensing and Control*, pp. 74-79, Apr 2011.
- [66] Zhan-Li Sun, Kien Ming Ng, Joanna Soszynska-Budny,Mohamed Salahuddin Habibullah, "Application of the LP-ELM model on transportation system lifetime optimization", in *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 4, pp. 1484-1494, Dec. 2011.

- [67] Itziar Landa Torres, Emilio G. Ortiz-Garcia, Sancho Salcedo-Sanz, Maria J. Segovia-Vargas, Sergio Gil-Lopez, Marta Miranda, Jose M. Leiva-Murillo & Javier Del Ser, "Evaluating the internationalization success of companies through a hybrid grouping harmony search- extreme learning machine approach", in *IEEE Journal of Selected Topics in Signal Processing*, vol. 6, no. 4, pp. 388-398, Aug. 2012.
- [68] Hui-Jun Cheng, Jyh-Da Wei, Chun-Yuan Lin and Jin Ye, "Detecting sea -curved image by extreme learning machine using patch analysis method, jury voting, and combinatorial fusion", in *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 50, no. 5, pp. 1850-1864, May 2020.
- [69] N Rathore, N Jain and A Mishra "Image Forgery Detection Using Singular Value Decomposition with Some Attacks ", *National Academy Science Letters, Springer Publication*, July 2020.
- [70] N Jain, N Rathore, and A Mishra "An Efficient Image Forgery Detection Using Biorthogonal Wavelet Transform and improved Relevance Vector Machine ", *Wireless Personal Communication, Springer Publication-New-York (USA)*, vol: 101 (4), pp. 1983-2008, 10-May-2018.
- [71] Chuwei Yang, Huaping Lin, Shouyi Liao & Shicheng Weng, "Pedestrian detection in thermal infrared image using extreme learning machine", *Proceedings of ELM-2014 Volume 2. Proceedings in Adaptation, Learning and Optimization*, vol. 4. Springer, Cham, 2014.
- [72] P. R. Asha and S Karpagavalli, "Diabetic retinal exudates detection using extreme learning machine", *Emerging ICT for Bridging the Future - Proceedings of the 49th Annual Convention of the Computer Society of India CSI Volume 2. Advances in Intelligent Systems and Computing*, vol. 338, pp. 573-578, Springer, Cham, 2015.
- [73] Subhash Chandra, "Verification of dynamic signature using machine learning approach", *Neural Computing and Applications*, vol. 32, pp. 11875-11895, Aug. 2020.
- [74] N Jain, N Rathore and A Mishra "An Efficient Image Forgery Detection Using Biorthogonal Wavelet Transform and Improved Relevance Vector Machine with Some Attacks ", *Interciencia Journal*, volume: 42 (11), pp. 95-120, 2017.
- [75] Erik Cambria, Guang-Bin Huang, "Trends and controversies: Extreme learning machine", in *IEEE Intelligent Systems*, vol. 28, no. 6, pp. 30-59, Nov.-Dec. 2013.
- [76] Neeraj Kumar Rathore and Farah Khan, "Survey of IoT", *Journal of Cloud Computing (JCC), ManTech Publication*, vol. 1, no-1, pp. 1-13, 2018.
- [77] Chris McIntosh, Igor Svistoun & Thomas G Purdie, "Groupwise conditional random forests for automatic shape classification and contour quality assessment in radiotherapy planning", in *IEEE Transactions on Medical Imaging*, vol. 32, no. 6, pp. 1043-1057, Jun. 2013.
- [78] Neeraj Kumar Rathore, "Ethical Hacking & Security against Cyber Crime", *Journal on Information Technology (JIT)*, vol-5, no. 1, pp. 7-11, Feb. 2016.
- [79] Guang-Bin Huang, Xiaojian Ding, Hongming Zhou, "Optimization method based extreme learning machine for classification", *Neurocomputing*, vol. 74, no. 1-3, pp. 155-163, Dec. 2010.
- [80] Gaitang Wang, Ping Li, "Dynamic adaboost ensemble extreme learning machine", *2010 3rd International Conference on Advanced Computer Theory and Engineering (ICACTE)*, pp. V3-54-V3-58, Aug. 2010.
- [81] Pradeep Laxkar and N Rathore "Load Balancing Algorithm in Distributed Network", *Solid State Technology*, ISSN: 0038-111X, Vol. 63, Issue 2s, SCOPUS Indexed, IF=0.05, 2020.
- [82] Dhiraj Pandey, Umashankar Rawat, Neeraj Kumar Rathore, Kavita Pandey, and Piyush Kumar Shukla "Distributed Biomedical Scheme for Controlled Recovery of Medical Encrypted Images", *Innovation and Research in BioMedical engineering*, Elsevier IF=1.022, July 2020.
- [83] N Rathore, Umashankar Rawat, Satish Chandra Kulhari, "Efficient Hybrid Load Balancing Algorithm", *National Academy Science Letters, Springer Publication*, Vol-43 (2), 10.1007/s40009-019-00834-w, ISSN 0250-541X, IF=0.604, 26-October-2019.
- [84] N Jain, N Rathore, and A Mishra "An Efficient Image Forgery Detection Using Biorthogonal Wavelet Transform and improved Relevance Vector Machine ", *Wireless Personal Communication, Springer Publication-New-York (USA)*, vol: 101 (4), pp. 1983-2008, IF-1.200, 10-May-2018.
- [85] N Rathore, "Performance of Hybrid Load Balancing Algorithm in Distributed Web Server System", *Wireless Personal Communication, Springer Publication-New-York (USA)*, pp. 1233-1246, vol:101 (4), IF -1.200, 2018.
- [86] N Jain, N Rathore and A Mishra "An Efficient Image Forgery Detection Using Biorthogonal Wavelet Transform and Improved Relevance Vector Machine with Some Attacks ", *Interciencia Journal*, volume: 42 (11), pp. 95-120, 2017.

- [87] N Rathore, "Dynamic Threshold Based Load Balancing Algorithms", Wireless Personal Communication, Springer Publication-New-York (USA), vol: 91 (1), pp. 151-185, Nov 2016.
- [88] N Rathore and I Chana, "Job Migration Policies for Grid Environment", Wireless Personal Communication, Springer Publication-New-York (USA), volume: 89 (1), pp. 241-269, IF -0.979, July-2016.
- [89] Rathore, Neeraj, and Inderveer Chana. "Variable threshold-based hierarchical load balancing technique in Grid." Engineering with computers 31.3 (2015): 597-615.
- [90] Vishal Sharma, Rajesh Kumar & Neeraj Kumar Rathore, "Topological Broadcasting Using Parameter Sensitivity Based Logical Proximity Graphs in Coordinated Ground-Flying Ad Hoc Networks", Journal of Wireless Mobile Networks Ubiquitous Computing and Dependable Applications (JoWUA), SCOPUS indexed, volume: 6, number: 3, pp. 54-72, September 2015.
- [91] Rathore, Neeraj, and Inderveer Chana. "Load balancing and job migration techniques in grid: a survey of recent trends." Wireless Personal Communications 79.3 (2014): 2089-2125.
- [92] Rathore, Neeraj, and Inderveer Chana. "Job migration with fault tolerance based QoS scheduling using hash table functionality in social Grid computing." Journal of Intelligent & Fuzzy Systems 27.6 (2014): 2821-2833.
- [93] Neeraj Kumar Rathore and Jyoti Rathore, "Efficient Checkpoint Algorithm for Distributed System", International Journal of Engineering and Computer Science (IJECS), E-ISSN: 2663-3590, P-ISSN: 2663-3582, vol. 1, no-2, pp.-59-66, 2019.
- [94] Neeraj Kumar Rathore and Farah Khan, "INTERNET OF THINGS: A Review Article", Journal of Cloud Computing (JCC), ISSN Print: 2349-6835 ISSN Online: 2350-1308, vol. 5, no-1, pp.-20-25, IF=0.333, Jan-Jun 2018.
- [95] Neeraj Kumar Rathore and Pramod Kumar Singh, "A Comparative Analysis of Fuzzy based Load Balancing Algorithm", Journal of Computer Science (JCS), vol-5, No-2, pp. 23-33 , June-August 2017.
- [96] Neeraj Kumar Rathore and Harikesh Singh, "Analysis of Grid Simulators Architecture", Journal on Mobile Applications and Technologies (JMT), vol-4, No-2, pp.32-41, July-December 2017.
- [97] Neeraj Kumar Rathore, "Checkpointing: Fault Tolerance Mechanism", Journal on Cloud Computing (JCC), vol-3, number-4, pp.27-34 August- October 2016.
- [98] Neeraj Kumar Rathore, "A Review towards: Load balancing Techniques", Journal on Power Systems Engineering (JPS), vol-4, No-4, pp. 47-60, November 16-January 2017.
- [99] Neeraj Kumar Rathore, "Faults in Grid", International Journal of Software and Computer Science Engineering, MANTECH PUBLICATIONS, vol-1, number-1, pp. 1-19, 2016.
- [100] Neeraj Kumar Rathore, "Installation of Alchemi.NET in Computational Grid", Journal on Computer Science (JCOM), vol-4, number-2, pp.1-5, June-August 2016.
- [101] Neeraj Kumar Rathore, "Ethical Hacking & Security Against Cyber Crime", Journal on Information Technology (JIT), vol-5, number-1, pp. 7-11 December 2015-February 2016.
- [102] Neeraj Kumar Rathore, "Efficient Agent Based Priority Scheduling and Load Balancing Using Fuzzy Logic In Grid Computing", Journal on Computer Science (JCOM), vol-3, number-3, pp. 11-22, September-November 2015.
- [103] Neeraj Kumar Rathore, "Map Reduce Architecture for Grid", Journal on Software Engineering (JSE), vol-10, issue-1, pp 21-30, July-September, 2015.
- [104] Neeraj Kumar Rathore, "GridSim Installation and Implementation Process", Journal on Cloud Computing (JCC), vol-2, number-4, pp.29-40 August- October 2015.
- [105] Neeraj Kumar Rathore and Inderveer Chana, "Report on Hierarchal Load Balancing Technique in Grid Environment", Journal on Information Technology (JIT), vol. 2, No. 4, ISSN Print: 2277-5110, pp-21-35, Sep - Nov 2013.
- [106] Neeraj Kumar Rathore and Inderveer Chana, "Checkpointing Algorithm in Alchemi.NET", Pragyaa: Journal of Information Technology, IMS Dehradun. ISSN No.: 0974-5513, IEEE, CSI and MPCET, Vol 8: Issue 1, pp no. 32-38, Dehradun, June 2010.