Neurocomputing 556 (2023) 126618

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Survey Paper

A comprehensive review of extreme learning machine on medical imaging

Yoleidy Huérfano-Maldonado [a,b], Marco Mora [b,c,*], Karina Vilches [b,d], Ruber Hernández-García [b,e], Rodrigo Gutiérrez [d], Miguel Vera [f]

[a] Doctorado en Modelamiento Matemático Aplicado, Universidad Católica del Maule, Chile
[b] Laboratory of Technological Research in Pattern Recognition, Universidad Católica del Maule, Chile
[c] Facultad de Ciencias de la Ingeniería, Universidad Católica del Maule, Chile
[d] Facultad de Ciencias Básicas, Universidad Católica del Maule, Chile
[e] Centro de Investigación de Estudios Avanzados del Maule, Universidad Católica del Maule, Chile
[f] Facultad de Ciencias Básicas y Biomédicas, Universidad Simón Bolivar, Colombia

ARTICLE INFO

Communicated by A. Iosifidis

Keywords:
Extreme learning machine
Medical imaging
Supervised training
Unsupervised training
Semi-supervised training

ABSTRACT

The feedforward neural network based on randomization has been of great interest in the scientific community, particularly extreme learning machines, due to its simplicity, training speed, and levels of accuracy comparable to traditional learning algorithms. Extreme learning machines (ELMs) are a type of artificial neural network (ANN) with one or more hidden layers that are trained under supervised, unsupervised, or semi-supervised learning approaches. These networks are widely used in various research areas, such as medical image processing (MI). This research work presents an exhaustive review of extreme learning machines (ELM) and medical image processing (MI), due to the high impact that these networks have had on the scientific community and the importance of MI for physicians who use them to diagnose different injuries and diseases. First, the theoretical construct of ELMs is developed based on the types of supervised, unsupervised, and semi-supervised learning. Then, the importance of MI for the diagnosis of a disease or classification of the most commonly used imaging modalities is analyzed for articles concerning radiography, computed tomography (CT), magnetic resonance (MR), ultrasound (US), and mammography (MG). Next, the reference data sets linked to various human body organs, such as the brain, lungs, skin, eyes, breasts, and cervix are described. Then, a review, analysis, and classification of the development of the last 6 years (2017–2022) of ELMs, based on learning types and MI, is performed. With the information obtained above, a construction of summary tables of the articles, classified according to the type of learning, is performed, highlighting the organ, reference, year, methodology, database, modality, and results. Finally, the discussion, conclusions and challenges related to this topic are presented. The findings indicate that the review articles reported in the literature have not addressed the relationship between ELMs and medical imaging in depth and have excluded key aspects, which are developed in this article. These aspects include a comprehensive analysis of the most popular imaging modalities, a detailed description of both the most popular databases and the most relevant databases for the machine learning community and, finally, the incorporation of schemes that explain the fundamentals of the main learnings considered when generating ELM-based trained smart models, which can be useful for medical image processing.

1. Introduction

Artificial neural networks (ANNs) are computational models that are inspired by the functioning of the human brain. They are fundamental elements of artificial intelligence and are applied to solve a wide range of problems, including pattern recognition, classification, regression, and natural language processing [1].

In general, the process of supervised training of an ANN involves minimizing a loss function that compares the desired outcomes of the data with the ANN results by updating the parameters. Traditional methods often adjust the parameters using the derivatives of the loss function. However, this approach presents significant disadvantages such as slow convergence, issues with local minima, and difficulty in selecting the appropriate model [1,2].

The approach of Random Weight Neural Network (NNRW) provides a solution to the problems present in conventional ANNs and deep learning approaches based on backpropagation (BP), as it allows for

* Corresponding author at: Laboratory of Technological Research in Pattern Recognition, Universidad Católica del Maule, Chile.
E-mail addresses: yoleidy.huerfano@gmail.com (Y. Huérfano-Maldonado), mmora@ucm.cl (M. Mora), kvilches@ucm.cl (K. Vilches), rhernandez@ucm.cl (R. Hernández-García), rgutierreza@ucm.cl (R. Gutiérrez), miguel.vera@unisimon.edu.co (M. Vera).

https://doi.org/10.1016/j.neucom.2023.126618
Received 20 January 2023; Received in revised form 8 June 2023; Accepted 24 July 2023
Available online 29 July 2023

much faster training speed without compromising accuracy [1]. NNRW is defined as a non-iterative training algorithm in which hidden weights and biases are randomly chosen within a specified range and kept constant throughout the training process, while weights between the hidden layer and the output layer are determined analytically [1,2].

Methods based on randomization have achieved notable results in various network structures, such as single- hidden-layer feedforward neural network (NN) [3], radial basis function (RBF) neural networks [4], deep neural networks with multiple hidden layers [5], convolutional neural networks (CNN) [6], recurrent neural networks (RNN) [7], auto-encoder neural networks [8], among others.

The Random Vector Functional Link (RVFL) neural net-work, proposed by Pao in 1992 [9], can be considered a random version of functional link neural networks. RVFL is a special single-layer feedforward neural network (SLFN) where the input layer is directly connected to both the hidden layer and the output layer. RVFL can be of two types: iterative RVFL iteratively obtains the output weights based on the gradient of the error function, while closed-form RVFL obtains the output weights in a single step using a closed-form solution. This algorithm demonstrates that real-valued weights, from the input layer to the hidden layer, can be randomly generated within an appropriate range and remain fixed during training [10].

Here are some variants of RVFL that explore new approaches. For instance, Vukovic et al. [11] proposed an Orthogonal Polynomial Expansion RVFL Neural Network (OPE-RVFLNN) that exploits the nonlinear expansion of the input vector into a set of orthogonal functions, particularly using Chebyshev orthogonal polynomials. The research confirms the findings of Zhang et al. [10], who claimed that the direct connection between network layers improves network performance, and that ridge regularization produces more accurate results than pseudo-inverse in regression experiments for the considered databases.

In the same decade, Schmidt et al. proposed another standard feedforward neural network with random weights (SNNRW) [12]. Unlike RVFL, in SNNRW there is no direct connection between the input layer and the output layer. The weights between the input layer and the hidden layer are randomly selected and kept fixed throughout the training process. The weights between the hidden layer and the output layer are determined using the Fisher method [1]. Since then, many researchers have devoted significant efforts to the development of theories, approaches, and applications for neural networks with random weights. One of the most attractive approaches is the one proposed by Huang et al. [13] called Extreme Learning Machine (ELM). It is important to note that both SNNRW and RVFL are considered to be precursors of ELM in the literature and, in that sense, it is evident that ELMs constitute a particular case of SNNRW.

ELMs are artificial neural networks that include one or more hidden layers and need training, which is either supervised [14], semi-supervised [15] or unsupervised [16]. The scientific community is very interested in this approach since it performs well, in terms of the computational time required for training, and is computationally simpler than other machine learning algorithms, such as backpropagation ANN (BP-ANN) [17] and Support Vector Machines (SVM) [13].

The core of ELM is that the hidden nodes' learning parameters, such as their input weights and biases, are allocated pseudo-randomly without the requirement for adjustment; whereas, the output weights may be calculated analytically by solving an overdetermined linear system using pseudo-inverse matrices. The model is easy to use, and ELMs are useful for classification, regression, and clustering issues due to the model's simplicity, low convergence speed, limited number of training parameters, and strong generalization capabilities.

In Table 1, a comparison is presented between some randomly weighted architectures. There, some differences that occur in these algorithms can be noted.

Along with ELMs, the literature reports other training paradigms such as support vector machines (SVMs), randonomic machines (e.g., RVLF) and those based on deep learning (e.g., CNN). In this context, there is strong evidence that ELMs openly outperform SVM-based approaches in both efficiency and accuracy, considering the most popular existing databases in the machine learning world [23–25]. Additionally, while ELMs are used with respect to any of the supervised, unsupervised or semi-supervised learning types, the fundamentals of SVMs mean that they are mostly used in supervised and/or reinforcement learning, leaving aside unsupervised learning, which is vital when the available data is not labeled [26].

It is known that CNNs outperform ELMs, in terms of accuracy, in any comparative scenario [27,28], but it is also true that ELMs exhibit considerably better efficiency than that obtained by the CNN approach. The latter situation naturally derives from the need for CNNs to iteratively adjust the weights during learning, which negatively impacts the training time employed by this paradigm [29,30]; whereas, ELMs keep these weights fixed throughout the learning process, once they have been randomly assigned.

1.1. Review of ELM and ML-ELM

Some review articles with a sole focus on ELMs have been published in the past ten years to demonstrate their significance and development [31–34]. These studies presented a more comprehensive review of the algorithms at the theoretical and applied levels, highlighting variants that improve the performance of the ELM algorithm, such as incremental ELM (I-ELM), pruning ELM, error-minimized ELM (EM-ELM), two-stage ELM (TS-ELM), online sequential ELM (OS-ELM), auto-encoder-based ELM (AE-ELM), multilayer ELM (ML-ELM), hierarchical ELM (H-ELM) and ELM, considering conjugate gradient (CG-ELM).

In the review articles presented by [34,35], the authors introduced and described several applications of classic ELM. In addition, they examined additional variants of ELM that were used to solve the classification, regression, clustering, compression, and function learning problems. In addition, metaheuristic algorithms, which improve the fit of ELMs, can be performed [36]. Table 2 provides a summary of the ELM variants where the contribution of these variants is highlighted accompanied by the respective references.

It is important to mention that within the search for ELM reviews, some ML-ELM-based review articles were found, which present a comprehensive review of the development of the ML-ELM architecture along with some variants and applications. For example, in [43] they made a detailed study and applications of the ML-ELM approach, including stacked ELM automatic encoder (ELM-AE), residual ELM and local receptive field based ELM (ELM-LRF). An interesting aspect is the fact that these authors also reviewed the connection between random neural networks and conventional deep learning, which is quite interesting.

Additionally, in [44] they made a current description of the evolution of multilayer models, which are grouped into random mappings, kernel correntropy strategies and conditional probability techniques. They also present a class of fast iterative algorithms called Shrinkage-Thresholding, which solve the minimization problem associated with an Autoencoder and conditional probability techniques. On the other hand, [45] performed a comprehensive review of both the development of the ML-ELM architecture and its variants and applications.

When looking for review articles on ELMs, there are not many studies that discuss the relationship between ELMs and medical image (MI) processing. These reputable publications only briefly mention papers generated between 2011 and 2013 on topics like epileptic electroencephalogram (EEG) pattern identification and thyroid ailment detection [34]. For example, in their section on medical/biomedical applications. Following the presentation of various investigations using MR, CT, and MG images in addition to ELMs [48] the ELM model based on meta-heuristics is also presented for use in medical applications [49]. Medical imaging is not really covered in depth in these papers.

Table 1
Comparison of some random weight architectures.

Algorithm	Input–output connection	No tuning input-hidden's weights	Hidden-output weights method	References
RVFL	✓	x	MPp	[18–20]
SNNRW	x	✓	Fisher	[2,21]
ELM	x	x	MPp	[14,22]

Table 2
A brief summary of ELM variants.

Algorithm	Contribution	Reference
I-ELM	Hidden nodes are randomly assigned to the hidden layer one after another and the output weights of existing hidden neurons are modified.	[37,38]
Pruning ELM	Starts with a large number of hidden nodes and then eliminates irrelevant hidden nodes by considering their relevance to class labels during learning. This algorithm is mainly adapted to pattern classification problems.	[37–39]
EM-ELM	This algorithm is similar to I-ELM. Nodes in hidden layers are assigned one after the other or group by group. Output weights are consistently updated at each step, which decreases computational complexity. Better generalization performance compared to I-ELM.	[37,38,40]
TS-ELM	In the first stage, an iterative feedforward algorithm is involved at each step, to select the hidden nodes from the randomly generated candidates which are added to the network until the established stopping criteria are met. In the second stage, selected hidden nodes are reviewed to eliminate irrelevant nodes, which drastically reduce the complexity of the network.	[37,38,41]
OS-ELM	Allows correcting or varying the size of the data set and then learning the data one after the other or set by set. It can handle additive nodes and radial basis functions in a unified framework.	[37,38,42]
AE-ELM	It plays a key role in unsupervised learning and the formation of deep architectures. It is non-iterative and reconstructs input information from random and orthogonal transformations. In the hidden layers, it has orthogonal weights and biases, which raises the quality of the ELM performance.	[43–45]
ML-ELM	Their unsupervised training is predominantly conducted using the auto-encoder approach. Eliminates the need to adjust parameters, allowing a better representation of the learning process by using the hierarchy of the different layers that make up the learning process.	[43–45]
H-ELM	This algorithm is usually used in the context of supervised classification based on feature extraction and selection.	[45–47]

The introduction of ELMs in various research areas such as chemistry, medicine, transportation, economics, and robotics, among others, has materialized thanks to their excellent characteristics and for the purpose of this research, the applications of ELMs in medicine, specifically in MI, are of great interest. The latter are of great utility for clinical experts because, through them, evaluations, diagnoses, monitoring and treatments of different injuries and diseases can be performed [50].

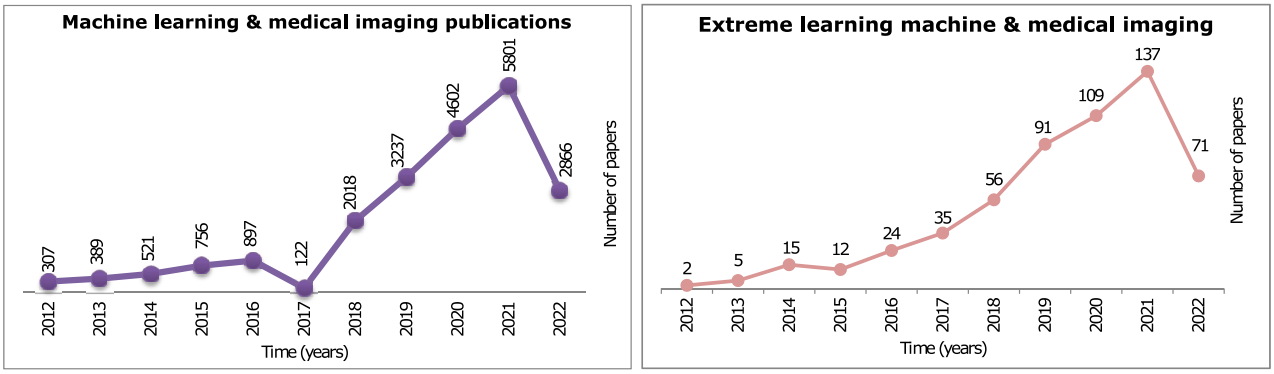
Many works have been carried out in the field of ELM and medical imaging with the aim of providing support to clinicians by generating more accurate and efficient diagnoses. To determine the trends in machine learning considering the MI-ELM binomial, information from the PubMed repository available on web page,[1] which is a free resource that supports the search and retrieval of biomedical and life sciences literature, was analyzed to search for information about the number of publications that have been generated from 2012 to 2022 by placing the keywords "machine learning and medical imaging" and "extreme learning machines and medical imaging". The information obtained from this repository is shown in Fig. 1 in which one can notice the sustained growth, starting in 2017, for machine learning with medical imaging Fig. 1(a) and the steady growth (although with only a few articles) of ELMs with medical imaging Fig. 1(b).

1.2. Research motivation

One of the problems identified during this research is the lack of available review articles, specifically on ELM and MI. The specialized literature only reports one review whose focus is on computer-aided diagnosis (CAD) systems based on ELMs [51]. When reviewing this article in depth, it was noted that the authors described both the classical ELM model and the CAD model. They primarily focused on the implementation of ELM and its enhanced algorithm in CAD systems, the feature extraction method, the feature selection method, and their interactions. This final argument was generated using several papers that were centered on various human organs.

While it is true that this review article presents relevant information on ELMs in a medical setting, the following aspects were not considered: a detailed and exhaustive study of the types of learning associated with ELM, in this case, supervised, unsupervised and semi-supervised learning. They also did not describe the imaging modalities most commonly used in the investigations, nor did they highlight the advantages and disadvantages of these modalities. This point is important, since

[1] https://pubmed.ncbi.nlm.nih.gov/



(a) (b)

Fig. 1. Number of publications in the last 10 years: (a) Machine learning in the medical context and (b) ELM vs MI.

it is necessary to know what the advantages or difficulties for any algorithm applied in MI would be (e.g., the types of noise present in each modality) and how to deal with them when working with these images. Additionally, there is no description of the public datasets most commonly used in research.

Therefore, the main motivation behind this research is to study the MI-ELM binomial in depth, allowing, on the one hand, identification of the challenges and limitations of ELMs when used in medical image analysis and, on the other hand, establishing future research directions, with an emphasis on the interpretability of ELM solutions and their generalizability, which is strongly related to the characteristics of the data.

1.3. Contributions

Considering the problems outlined above, a review of extreme learning machines and medical imaging is presented overcoming aspects of previous works that could be improved. In this sense, the contributions of the present work are the following:

- Comprehensive study of ELM architecture under supervised, unsupervised and semi-supervised learning approaches.
- Description of the imaging modalities most commonly used in research, highlighting their advantages and disadvantages.
- Classification of the most required databases in the experiments performed with ELM.
- Comparative analysis of ELM with other machine learning algorithms used in medical imaging.
- Discussion of future research directions for ELM in medical imaging, including potential improvements and challenges to be overcome.

A literature search, related to the last 6 years (2017–2022), was carried out using the following list of keywords: ‘extreme learning machine’, ‘medical imaging’, ‘supervised learning’, ‘unsupervised learning’, and ‘semi-supervised learning’. In addition, in order to obtain representative literature related to the research topic, several databases from reputable publishers in the scientific community were used, such as Scopus, IEEE Xplore, MDPI, ScienceDirect, Web of Science, SpringerLink, Taylor & Francis, UpToDate, PubMed, and other medical journal publishers.

Also, it is important to mention that the analysis of recent publications linking MI and ELM allowed us to affirm that the trend, in this context, is the development of hybrid techniques based on deep learning algorithms, particularly, convolutional neural networks (CNN) and ELM. It is interesting to note that this hybrid approach has generated effective and efficient computational strategies to address the detection of pathologies in the human body.

The structure of this paper is subdivided into eight main sections. Section 2 describes the theoretical foundations of ELM. Section 3 provides an analysis of the medical imaging modalities reported in the reviewed articles and the referenced public datasets. Sections 4, 5 and 6 present reviews of recent theoretical–practical advances related to medical image processing using supervised, unsupervised, and semi-supervised ELMs, respectively. Section 7 presents the discussions and Section 8 develops the conclusions and future challenges.

2. Fundamentals of extreme learning machines

Extreme learning machines are training paradigms for neural networks that address supervised [52], semi-supervised [15] and unsupervised [16] learning. Each of these learning methods has its own challenges, as they require the application of rigorous protocols to validate their performance. A crucial question in this context is determining which one of these approaches is the most suitable. Generally, answering this question is not straightforward, due to the complexity of the problem to be solved, the characteristics of the data, and the constraints imposed on these types of networks. Therefore, it is necessary to comprehend the fundamentals of each of these approaches, which are outlined below.

2.1. Supervised algorithm

ELMs were proposed in [14] like a single-hidden-layer feed-forward neural network model (SLF-ELM). Due a smaller number of training parameters, low convergence speed and high generalization capability [14,22,53–55], ELM networks present advantages in classification problems [56–63], regression [64–73] and clustering [74–79].

Generally speaking, the basic ELM setup includes input, hidden, and output layers. Unlike traditional neural network learning algorithms, only the tuning of the number of hidden neurons in the network is preliminarily required during the execution process of the ELM algorithm; while the hidden layer biases and the weights between the input and hidden layers are pseudo-randomly assigned. The ELM algorithm also computes the output matrix of the hidden layer and generates the weights between the hidden and output layers by solving systems of linear equations [55].

In general, obtaining this solution involves pseudo-inverse matrices. A brief retrospective view of the conceptualization and use of this type of matrix, shows that several authors have developed works allowing the inclusion of the pseudo-inverse in multiple areas of human knowledge. In this sense, Fig. 2 presents a timeline that highlights both pioneers and more recent authors. The most relevant theoretical and practical contributions related to this important concept are developed below.

In 1903, Fredholm first introduced the idea of a generalized inverse when he developed a pseudo-inverse for a linear integral operator that

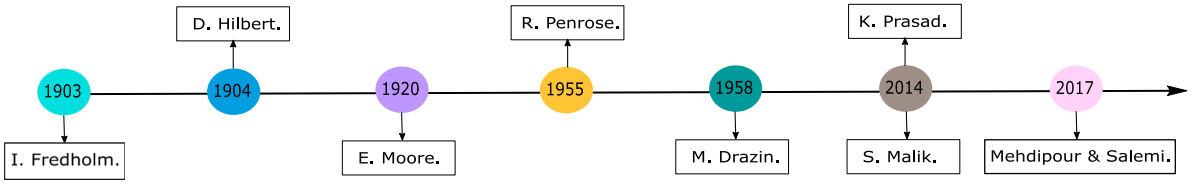


Fig. 2. Timeline about the generalized inverse or pseudo-inverse matrix.

was not invertible, in the traditional sense, which required a new class of determinant [80]. Furthermore, this author proposes and proves the so-called Fredholm Theorem providing the basis for the Hilbert-spaces theory, which can be interpreted in terms of the finite dimension of the kernel and the compactness of the above-mentioned integral operator.

A year later, Hilbert tackled the study of generalized inverse of differential operators [81]. In 1920, Moore pointed out the existence of a unique matrix, called the generalized inverse matrix, that satisfies four special properties, including the property that the original matrix multiplied by the generalized inverse is equal to its orthogonal projection [82]. This matrix is key to solve arbitrary systems of linear equations, and it is useful to find optimal solutions to least-squares problems.

In 1955, Penrose [83] published his results, starting from an algebraic definition equivalent to the one given by Moore. The generalized inverse introduced by Moore is an extension of the inverse matrix idea that fulfills some, but not all, of the properties of an analytic inverse matrix. A new generalized inverse for square matrices, in the context of abstract rings, is presented by Drazin [84]. An analysis of the Ref. [84] allows us to state a relationship between pseudo-polar rings and Drazin's pseudo-inverses, in associative rings and Banach algebras.

In 2014, Prasad and Mohana considered another alternative definition, based on inverse matrices with an equal projection kernel (core-EP inverse or CEP inverse matrices) [85]. This type of matrix is actually a new type of pseudo-inverse matrix and is extremely important in the construction of new characterizations of the Drazin and Moore–Penrose pseudo-inverse matrices, as can be seen in the Refs. [86–88]. Furthermore, EP matrices belong to the complex numbers and are defined according to the fulfillment of the equality of the algebraic ranks of both the original matrix and its conjugate transpose.

At the same time, Malik and Thome defined the so-called Drazin–Moore–Penrose (DMP) inverse [89], which is also a generalization of the Moore–Penrose and Drazin inverses, and it is valid for arbitrary index matrices. The authors derived some properties of DMP, useful in the resolution of linear systems and in the matrix ranks calculations. Subsequently, in 2017, the complex Moore–Penrose (CMP) inverse was defined by Mehdipour and Salemi for square matrices of arbitrary index [90]. These authors showed that CEP matrices have equal properties that the generalized matrices with kernels of equal projection [91]. This evolution has allowed the widening of the field of application of pseudo-inverse matrices.

In spite of the various ways of approaching pseudo-inverse matrices, in the context of ELM, their calculation has been biased towards the consideration of the ideas embodied by the international scientific community in the geometric algebraic approach called generalized Moore–Penrose inverse [92]. There are many methods for the effective calculation of inverse matrices, including QR factorization, tensor product matrix, Cholesky singular matrix factorization, the conjugate Gram–Schmidt process, and the Nyström method.

Additionally, considering ELM approaches, the supervised learning problem can be posed as described below. For N arbitrary and distinct samples $(\boldsymbol{x}_j, \boldsymbol{t}_j)$ then $\boldsymbol{x}_j = [x_{j1}, x_{j2}, \ldots, x_{jn}]^T \in R^n$, where n is the number of data attributes and $\boldsymbol{t}_j = [t_{j1}, t_{j2}, \ldots, t_{jm}]^T \in R^m$, where m is the number of classes characterizing the database taken into consideration. In addition, L is the number of hidden nodes and g is

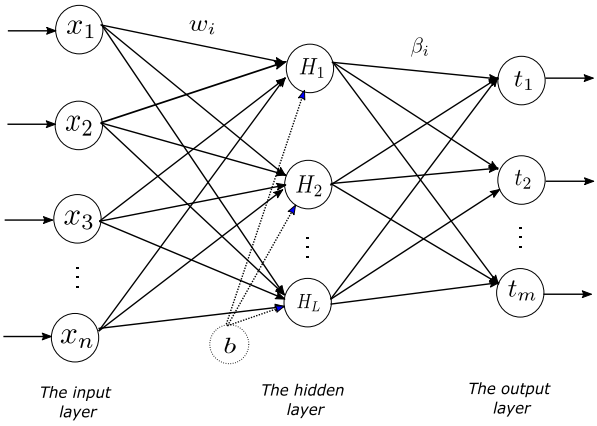


Fig. 3. Representation of the learning scheme of the supervised ELM algorithm.

an activation function; the output of the ELM in supervised learning problems is given by Eq. (1)

$$\sum_{i=1}^{L} \boldsymbol{\beta}_i g_i(\boldsymbol{x}_j) = \sum_{i=1}^{L} \boldsymbol{\beta}_i g(\boldsymbol{w}_i \boldsymbol{x}_j + b_i) = \boldsymbol{t}_j, \quad j = 1, \ldots, N \tag{1}$$

where $\boldsymbol{w}_i = [w_{i1}, w_{i2}, \ldots, w_{in}]^T$ is the matrix of weights connecting the ith hidden node to the jth training example $\boldsymbol{x}_j$, b_i contains the biases of the hidden layer, $\boldsymbol{\beta}_i = [\beta_{i1}, \beta_{i2}, \ldots, \beta_{im}]^T$ is the weights matrix of the output layer, and $\boldsymbol{t}_j$ represents the labels of the $\boldsymbol{x}_j$.

Eq. (1) can be written in a compact form, as shown in Eq. (2)

$$\boldsymbol{H}\boldsymbol{\beta} = \boldsymbol{T} \tag{2}$$

where

$$\boldsymbol{H} = \begin{bmatrix} g(\boldsymbol{w}_1 \boldsymbol{x}_1 + b_1) & \cdots & g(\boldsymbol{w}_L \boldsymbol{x}_1 + b_L) \\ \vdots & \ddots & \vdots \\ g(\boldsymbol{w}_1 \boldsymbol{x}_N + b_1) & \cdots & g(\boldsymbol{w}_L \boldsymbol{x}_N + b_L) \end{bmatrix}, \tag{3}$$

$$\boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta}_1^T \\ \vdots \\ \boldsymbol{\beta}_L^T \end{bmatrix}, \boldsymbol{T} = \begin{bmatrix} \boldsymbol{t}_1^T \\ \vdots \\ \boldsymbol{t}_N^T \end{bmatrix}$$

$\boldsymbol{H}$ is called the output matrix of the hidden layer of the neural network [14,93,94], $\boldsymbol{T}$ is the transposed label matrix of the data set, and $\boldsymbol{\beta}$ is the matrix of weights of the output layer which is calculated using Eq. (4)

$$\boldsymbol{\beta} = \boldsymbol{H}^{\dagger}\boldsymbol{T} \tag{4}$$

where $\boldsymbol{H}^{\dagger}$ is the generalized Moore–Penrose inverse of the $\boldsymbol{H}$ matrix [92]. Fig. 3 shows the general structure of the supervised ELM method.

In a complementary way, the solution of Eq. (2) can also be obtained by using optimization methods [55], where the objective is to make ELMs acquire a higher generalization capacity, minimizing both the norm of the output weights and the training errors [95], given by

$$\begin{aligned} &\min_{\beta,\xi} \left\{ \tfrac{1}{2}\|\boldsymbol{\beta}\|^2 + \tfrac{C}{2}\sum_{i=1}^{N} \|\boldsymbol{\xi}_i\|^2 \right\} \\ &\text{s. t.: } h(\boldsymbol{x}_i)\boldsymbol{\beta} = \boldsymbol{t}_i^T - \boldsymbol{\xi}_i^T, i = 1, \ldots, N \end{aligned} \tag{5}$$

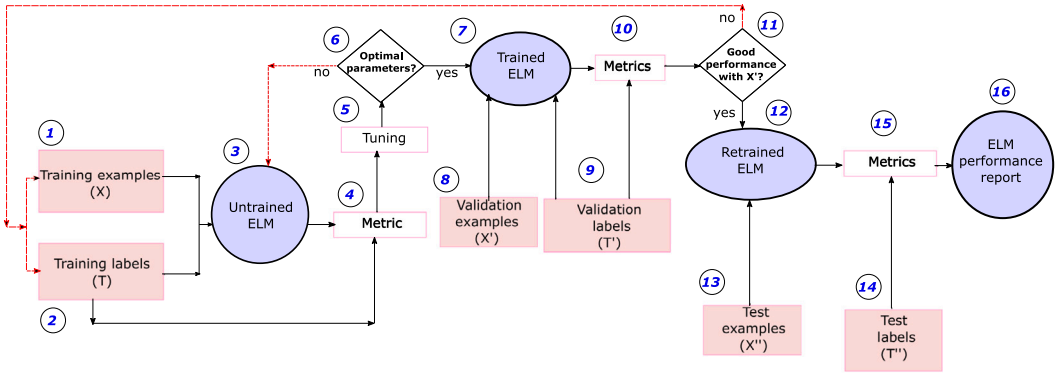


Fig. 4. General scheme for the generation of a trained ELM under the supervised learning approach.

ξ_i is the training error vector and C is a regularization parameter [96]. The solution of the output weights $\boldsymbol{\beta}$ is calculated as follows:

$$\boldsymbol{\beta} = \begin{cases} \left(\boldsymbol{I} + C\boldsymbol{H}^T\boldsymbol{H}\right)^{-1} C\boldsymbol{H}^T\boldsymbol{T}; N > L \\ \boldsymbol{H}^T\left(\boldsymbol{I} + C\boldsymbol{H}\boldsymbol{H}^T\right)^{-1} C\boldsymbol{T}; N \leq L \end{cases} \tag{6}$$

where $\boldsymbol{I}$ is an identity matrix of dimension $n \in \{N, L\}$.

The supervised training pseudocode is summarized in Algorithm 1 [14,33]:

Algorithm 1: SL-ELM

Input : A training set $\aleph = \{(x_i, t_i) : x_i \in R^n, t_i \in R^m, i = 1, ..., N\}$, g and L.

Output: β.

1. Assign pseudo-random values to (w_i) and (b_i).
2. Calculate $\boldsymbol{H}$ using Eq. (3).
3. Calculate β, using Equation (4) or (6).

To perform a complete evaluation of a supervised ELM, it is necessary to develop the phases or processes of training, validation, and testing, as shown in Fig. 4.

In Fig. 4, for the purpose of having information to train an ELM under the supervised learning approach, the database considered was partitioned into three subsets called training, validation, and testing. Items 1 and 2 have the training examples (X) and training labels (T), that serve as preliminary input information to the ELM. In point 3, the outline of the basic structure of an ELM, which is presented in Fig. 3, should be included. Then, in point 4, one or several reference metrics were used to evaluate the performance of the network. In point 5, a tuning process was performed to obtain optimal values for the parameters that control the network performance. If the optimal parameters are not obtained, it is necessary to return to point 3 or, otherwise, to advance considering the validation subset.

Points 8 and 9 are the validation examples (X') and the validation labels (T'). These labels go through the established metrics and the performance of the ELM is evaluated by considering X'. If the performance of the ELM is not adequate, the algorithm returns to points 1 and 2; otherwise, the algorithm generates a re-trained ELM which will be evaluated by the test subset. Only test examples (X'') and test labels (T'') are considered (items 13, 14 and 15). Finally, point 16 establishes whether the network performance was adequate or not.

It is important to mention that, after the creation of ELMs, and due to their proven efficiency and excellent generalization capability, several variants of this special type of neural network have emerged, in order to further explore and exploit these features in the context of supervised learning. Such is the case of H-ELMs [46], bidirectional ELMs (B-ELMs) [67], evolutionary ELMs (SaDE-ELMs) [97], fully complex ELMs (C-ELMs) [98] and OS-ELM [99], among others. Likewise, different implementations of the supervised ELM algorithm have been generated in different programming languages, e.g., Matlab, Python, C/C++, and Java, whose codes are available in web page.[2]

2.2. Unsupervised algorithm

ELMs have mainly been used in classification and regression tasks, in the context of supervised learning, which greatly limits their applicability. As an alternative solution to this limitation, intelligent strategies based on ELMs were developed in [15] by considering unsupervised and semi-supervised learning using multiple regularization.

In general, it can be stated that unsupervised learning (USL) is an interesting algorithm, since it does not require labeled data for the training process. An important task of USL is the creation of data groups (clustering) that allows for the analysis of information, leading to the detection and identification of hidden patterns in the data [100].

Unsupervised extreme learning machines (USL-ELMs) carry over the computational effectiveness and learning capacity of conventional ELMs [15]. In an unsupervised environment, all the training data are not labeled and, therefore, in the USL-ELM algorithm, $\boldsymbol{X} = \{\boldsymbol{x}_i\}_{i=1}^N$ are not labeled. This algorithm, can be based on various mathematical models such as the one given by [15]

$$\begin{aligned} &\min_{\beta} \left\{\|\boldsymbol{\beta}\|^2 + \lambda Tr\left(\boldsymbol{\beta}^T\boldsymbol{H}^T\boldsymbol{g}\boldsymbol{L}\boldsymbol{H}\boldsymbol{\beta}\right)\right\}, \\ &\text{s. t.: } (\boldsymbol{H}\boldsymbol{\beta})^T\boldsymbol{H}\boldsymbol{\beta} = \boldsymbol{I} \end{aligned} \tag{7}$$

where λ is a compensation parameter, $Tr(\cdot)$ denotes the trace of matrix $\left(\boldsymbol{\beta}^T\boldsymbol{H}^T\boldsymbol{g}\boldsymbol{L}\boldsymbol{H}\boldsymbol{\beta}\right)$, $\boldsymbol{gL}$ is the graph Laplacian matrix, and I is an identity matrix.

The following extended eigenvalue decomposition problem needs to be solved, in order to determine the best solution to Eq. (7):

$$\left(\boldsymbol{I} + \lambda\boldsymbol{H}^T\boldsymbol{g}\boldsymbol{L}\boldsymbol{H}\right)\boldsymbol{v} = \gamma\boldsymbol{H}^T\boldsymbol{H}\boldsymbol{v} \tag{8}$$

The network output weight matrix β can be formed by matching its columns with the vectors associated with the eigenvalues $(\boldsymbol{v}_i)$ of Eq. (8) [101]. If $L \leq N$, β is given by

$$\boldsymbol{\beta}^* = \tilde{\boldsymbol{v}}_i \tag{9}$$

where $\tilde{\boldsymbol{v}}_i$ are the generalized eigenvectors corresponding to (8) given by Eq. (11)

$$\tilde{\boldsymbol{v}}_i = \frac{\boldsymbol{v}_i}{\|\boldsymbol{H}\boldsymbol{v}_i\|}, i = 2, \ldots, n_o + 1 \tag{10}$$

If $L > N$, (9) would be underdetermined. In this case,

$$\left(\boldsymbol{I} + \lambda\boldsymbol{g}\boldsymbol{L}\boldsymbol{H}\boldsymbol{H}^T\right)\boldsymbol{u} = \gamma\boldsymbol{H}\boldsymbol{H}^T\boldsymbol{u} \tag{11}$$

where $\boldsymbol{u} = \boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_{n_o+1}$ are the generalized eigenvectors corresponding to the ith smallest eigenvalues of (11). Then the final solution would be given by

$$\boldsymbol{\beta}^* = \boldsymbol{H}^T[\tilde{\boldsymbol{u}}_i] \tag{12}$$

[2] http://www.extreme-learning-machines.org.html

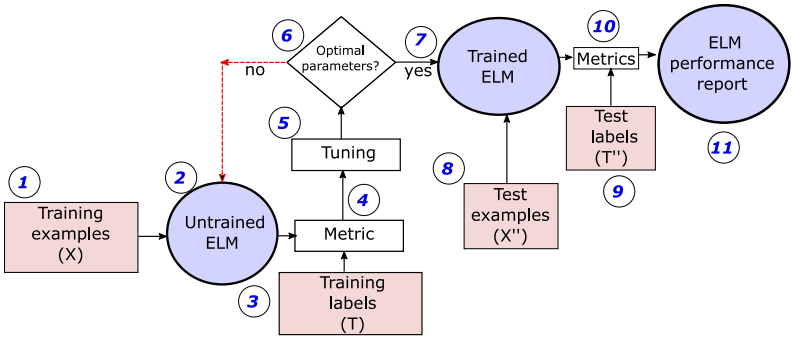


Fig. 5. General scheme for the generation of a trained ELM under the unsupervised learning approach.

where $\tilde{u}_i$ are the generalized eigenvectors associated to their eigenvalues (u_i) corresponding to Eq. (12), given by Eq. (13)

$$\tilde{\boldsymbol{u}}_i = \frac{\boldsymbol{u}_i}{\|\boldsymbol{H}\boldsymbol{H}_i^{Tu}\|}, i = 2, \ldots, n_o + 1 \tag{13}$$

The USL-ELM approach in this case often entails building an affinity matrix out of the input samples, applying an eigen-decomposition to the normalized affinity matrix, and utilizing d-eigenvectors to embed the original input data in a d-dimensional space. The data in the embedded space are then aggregated using straightforward k-means clustering [102].

The USL-ELM training pseudocode is summarized in Algorithm 2 [15,103].

Algorithm 2: USL-ELM

Input : A training set $\boldsymbol{X} \in R^{N \times n_i}$.
Output: The vector of cluster index labels and $\boldsymbol{y_i}$ corresponding to $\boldsymbol{x_i}$.
1. Calculate $\boldsymbol{L}$ of $\boldsymbol{X}$.
2. Initialize an ELM network of L with random input weights and calculate $\boldsymbol{H} \in R^{N \times L}$
3. If $L \leq N$;
Calculate $\tilde{\boldsymbol{v}}_i$ from (9) and let $\beta = [\tilde{v}_i]$; otherwise,
calculate $\tilde{\boldsymbol{u}}_i$ from (12) and let β as (13).
4. Calculate the embedding matrix: $\boldsymbol{E} = \boldsymbol{H}\beta$.
5. Construct vector $\boldsymbol{y}$ considering each row of $\boldsymbol{E}$ and agglomerating N points in K agroups using the k-means algorithm.

In a complementary manner, moving on to the training and testing processes that must be developed during the unsupervised learning of an ELM, as shown in Fig. 5, the structure is similar to that of Fig. 4, i.e., the relevant change is that, in point 1, only the training examples are considered and the labels at the input of the ELM are not presented. It can also be seen that the validation process has not been considered in this scheme, i.e., a more simplified scheme has been presented compared to Fig. 4.

Additionally, due to the reported high efficiency of ELMs, multiple studies have been carried out, in order to integrate ELMs and deep architectures. For example, the AE-ELM is one of the most used techniques when developing unsupervised learning processes [104]. The outputs of the AE are equal to the inputs and both the weights and biases of the hidden nodes must meet certain orthogonality conditions. Fig. 6 shows the general structure of the unsupervised AE-ELM method.

In the AE-ELM model, its input–output data, denoted by X, can be represented in 3 different attribute spaces (d), linked to the choice of the number of hidden neurons (L) [104]. These spaces are known as: compressed ($L < d$), egalitarian ($L = d$), and sparse ($L > d$). On the other hand, the random parameters of the hidden layer (weights and biases) are transformed, to endow them with mutual orthogonality, that

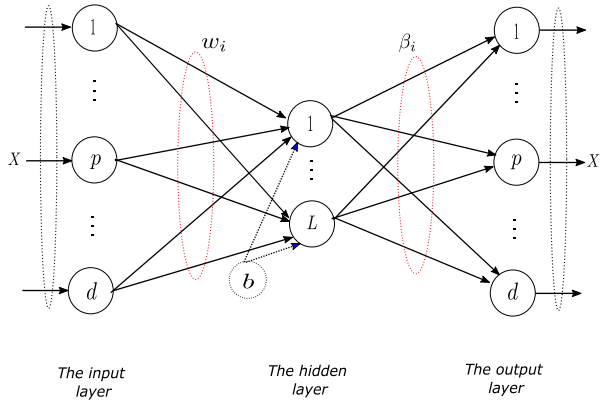


Fig. 6. Automatic orthogonal encoder generated considering the traditional ELM algorithm.

is:

$$\boldsymbol{w}^T\boldsymbol{w} = \boldsymbol{I}, \tag{14}$$

$$\boldsymbol{b}^T\boldsymbol{b} = 1 \tag{15}$$

where $\boldsymbol{I}$ is an identity matrix.

Then, matrix $\boldsymbol{H}$ is calculated using Eq. (3) and unsupervised learning is used to represent the attributes of $\boldsymbol{X}$ using Eq. (16)

$$\boldsymbol{H}\boldsymbol{\beta} = \boldsymbol{X} \tag{16}$$

$\boldsymbol{\beta}$ are the weights of the output layer and $\boldsymbol{X} = \{\boldsymbol{x}_i, \ldots, x_N\}$, with x_i being the ith example of the sample set. In this case, $\boldsymbol{\beta}$ can be calculated according to the type of attribute representation space.

For compressed, sparse or egalitarian spaces, a regularization parameter (C) is used, and the models are:

$$\boldsymbol{\beta} = \begin{cases} \left(\boldsymbol{H}^T\boldsymbol{H} + \frac{\boldsymbol{I}}{C}\right)^{-1}\boldsymbol{H}^T\boldsymbol{X}; N > L \\ \left(\boldsymbol{H}\boldsymbol{H}^T + \frac{\boldsymbol{I}}{C}\right)^{-1}\boldsymbol{H}\boldsymbol{X}; N < L \\ \boldsymbol{H}^{-1}\boldsymbol{X}, \quad \boldsymbol{\beta}^T\boldsymbol{\beta} = \boldsymbol{I}; N = L \end{cases} \tag{17}$$

Usually, matrix decomposition or factorization techniques are considered and, in the specific case that singular value factorization (SVD) is used, the attribute decomposition of (16) is given by Eq. (18)

$$\boldsymbol{H}\boldsymbol{\beta} = \sum_{i=1}^{N} \tilde{\boldsymbol{u}}_i \frac{\delta_i^2}{\delta_i^2 + C} \tilde{\boldsymbol{u}}_i{}^T \boldsymbol{X} \tag{18}$$

$\tilde{\boldsymbol{u}}_i$ are the eigenvectors of $\boldsymbol{H}\boldsymbol{H}^T$ and δ the singular values of $\boldsymbol{H}$, related to the SVD of $\boldsymbol{X}$. Finally, the output (f_L) of the AE-ELM, is given by Eq. (19)

$$f_L(x) = \sum_{i=1}^{L} \beta_i h_i(x) = \boldsymbol{g}(a_j\boldsymbol{x_i} + b_j) \tag{19}$$

where β_i is the weight between the ith L and the output layer, and $h_i(x)$ is the output of the ith L for the input x_i [104].

The pseudocode of the unsupervised training based on the AE-ELM model is summarized in Algorithm 3 [105,106]:

Algorithm 3: AE-ELM

Input : A training set $\boldsymbol{X} = \{\boldsymbol{x_i}\}_{i=1}^{N}$, L and C.
Output: The result of the grouping process contained in $f_L(x)$
1. Initialize an AE-ELM of L with random orthogonal input weights and biases.
2. Calculate β using (17), according to the case.
3. Calculate data attributes, using matrix decomposition techniques.
4. Calculate (f_L) by using (19)

On the other hand, different implementations of the USL-ELM and AE-ELM algorithm have been generated in the Matlab programming language. The codes of these algorithms are available on the web page mentioned above.

2.3. Semi-supervised algorithm

SSemi-supervised learning (SSL) is a technique that may employ both labeled and unlabeled data to produce higher prediction accuracy than either supervised learning or unsupervised learning. When working with real and large databases, there are often problems regarding labeled in its entirety, which generates problems when applying certain algorithms, such as supervised algorithms. To address this problem, the SSL technique is an appropriate solution.

When labeled data is sparse, an ELM approach, based on SSL learning (SSL-ELM) and multiple regularization, was developed by [15] to take advantage of unlabeled data and increase classification accuracy. The labeled data in the training set are referred to as $\{\boldsymbol{X_l}, \boldsymbol{Y_l}\} = \{\boldsymbol{x_i}, \boldsymbol{y_i}\}_{i=1}^{l}$, and the unlabeled data are $X_u = \{\boldsymbol{x_i}\}_{i=1}^{u}$, where l and u represent the number of labeled and unlabeled data, respectively. The SSL-ELM core problem is formulated in [15,107]

$$\begin{aligned} &\min_{\beta,\xi} \left\{ \tfrac{1}{2}\|\boldsymbol{\beta}\|^2 + \tfrac{C_i}{2}\sum_{i=1}^{l}\|\boldsymbol{\xi_i}\|^2 + \tfrac{\lambda}{2} Tr\left(\boldsymbol{F}^T \boldsymbol{gLF}\right) \right\}, \\ &\text{s. t.: } \boldsymbol{h}(\boldsymbol{x_i})\boldsymbol{\beta} = \boldsymbol{y_i}^T - \boldsymbol{\xi_i}^T, \qquad i = 1, \ldots, l \\ &\boldsymbol{f_i} = \boldsymbol{h}(\boldsymbol{x_i})\boldsymbol{\beta}, \qquad i = 1, \ldots, l+u \end{aligned} \tag{20}$$

where $Tr(\cdot)$ denotes the trace of the matrix $\left(\boldsymbol{F}^T \boldsymbol{gLF}\right)$, $\boldsymbol{gL}$ is the Laplacian graph constructed from labeled and unlabeled data, $F = [f(x_1), \ldots, f(x_{l+u})]^T$ is the output matrix of the network, C_i is the penalty coefficient with respect to the ith input example $(\boldsymbol{x_i})$ and λ is a compensation parameter.

By calculating the gradient of the unconstrained optimization problem of (20) with respect to $\boldsymbol{\beta}$, we obtain the solution $(\boldsymbol{\beta}^*)$ of the SSL-ELM

$$\boldsymbol{\beta}^* = \left(\boldsymbol{I_L} + \boldsymbol{H}^T \boldsymbol{CH} + \lambda \boldsymbol{H}^T \boldsymbol{LH}\right)^{-1} \boldsymbol{H}^T \boldsymbol{C\tilde{Y}} \tag{21}$$

where $\tilde{\boldsymbol{Y}} \in R^{(l+u)\times n_0}$ is the augmented training target with its first l rows equal to $\boldsymbol{Y}_l$ and the remainder equal to 0, $C \in R^{(l+u)\times(l+u)}$ is a diagonal matrix with $[C]_{ii} = C_i$, $i = 1, \ldots, l$ and remainder equal to zero. When $N < L$ the above SSL-ELM solution can be written as:

$$\boldsymbol{\beta}^* = \boldsymbol{H}^T \left(\boldsymbol{I} + \boldsymbol{CHH}^T + \lambda \boldsymbol{LHH}^T\right)^{-1} \boldsymbol{C\tilde{Y}} \tag{22}$$

where $\boldsymbol{I}$ is an identity matrix of dimension $l + u$.

The pseudocode of the semi-supervised training is summarized in Algorithm 4 [15,108]:

On the other hand, it is important to mention that code from the SSL-ELM algorithm in Matlab programming language is also available at the web page referenced above. Finally, Table 3 presents a summary of the three types of learning that have been developed, in order to visualize some differences and similarities among them.

Algorithm 4: SSL-ELM

Input : Data labeled $\{\boldsymbol{X_l}, \boldsymbol{Y_l}\} = \{\boldsymbol{x_i}, \boldsymbol{y_i}\}_{i=1}^{l}$, and unlabeled data $X_u = \{\boldsymbol{x_i}\}_{i=1}^{u}$.
Output: The SS-ELM mapping function: $\boldsymbol{f} : R^{n_i} \rightarrow R^{n_0}$
1. Calculate gL of both $\boldsymbol{X}_l$ and $\boldsymbol{X}_u$;
2. Initialize an ELM network of L with random input weights and biases and compute $\boldsymbol{H} \in R^{(l+u)\times L}$;
3. Choose the parameters C_i and λ;
4. If $L \leq N$;
Calculate β using (21); otherwise
Calculate β using (22)

3. Medical imaging

When a patient goes for a consultation with a clinical specialist, a preliminary interview and physical examination is carried out to establish the initial condition of that patient at that time. Once the interview and physical examination has occurred, the clinician develops a preliminary diagnosis [109]. A set of paraclinical tests are then ordered, to confirm the diagnosis and, sometimes, when the disease is more complex, the diagnosis must be validated by evidence, which generates resources such as medical imaging, which are obtained from various medical imaging modalities.

Medical imaging is a valuable tool for physicians because it can be used to evaluate, diagnose, treat, and monitor injuries and diseases by allowing visualization inside the human body [50,110]. There are several types of imaging modalities and each provides different information about the area of the body being observed and treated. For this reason, it is important for this paper to describe some of these modalities.

3.1. Acquisition modalities

Medical imaging modalities provide relevant information about a subject in order to detect diseases, make more accurate diagnoses, and plan more effective treatments and surgeries by physicians; they allow the morphology, physiology, and pathologies of the organs of the human body to be explored [110]. These image acquisition technologies are found throughout the range of the electromagnetic spectrum [50] and, over time, they have improved significantly, in terms of acquisition time, image quality, and resolution [111,112] and are now more widely used in many areas of medicine.

One classification that can be made for medical imaging modalities is if they are based on ionizing or non-ionizing radiation processes. X-ray, computed tomography (CT), single photon emission computed tomography (SPECT), positron emission tomography (PET), and mammography (MG) are the imaging modalities that use external sources of ionizing radiation. On the other hand, modalities that do not use ionizing radiation are ultrasound (US) and magnetic resonance (MR). This section will describe the modalities most frequently reported in the articles reviewed in Sections 4–6.

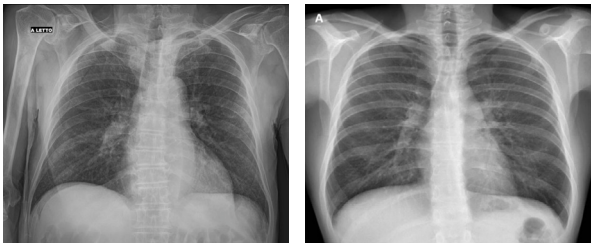
3.1.1. Radiography

Radiography began with the discovery of X-rays by Rontgen in 1895 and, due to the low cost and simplicity of production, they generate images that are widely used in medicine [113]. It is a diagnostic technique that uses ionizing electromagnetic radiation, called X-rays, to obtain information [114].

X-rays are absorbed or attenuated at varying levels, depending on the density and atomic number of the various tissues, and so radiographs or X-ray images display in black and white. The majority of the X-rays are absorbed by calcium in the bones, which is why bones appear white [114]. Fat and other soft tissues appear to be gray because they absorb less radiation. Lungs appear black because air absorbs the least

Table 3
Comparison between different types of learning based on ELMs.

Supervised learning	Unsupervised learning	Semi-supervised learning
It exhibits superior efficiency schemes to classical neural networks and excellent generalization capability [52].	It inherits the computational efficiency and learning capability of traditional ELMs [15].	It carries over the computational speed and capacity for learning of conventional ELMs [15].
It starts with pre-classified data and, therefore, it is mandatory to consider the labels in all the phases that involve the generation of the trained computational model.	The input data are not classified a priori, in this sense, it is impossible to consider pre-established labels in any of the phases involved in the generation of the trained learning model.	It can incorporate both labeled and unlabeled data in the learning process.
New labels are not generated during the training phase.	Labels are generated automatically to allow the identification of the classes created during training.	By employing heuristic approaches, re-enforcement learning occurs, which optimizes the generalization capability of ELMs under this type of learning.
In this learning type, ELMs have been shown to be superior with respect to support vector machines in classification [59], regression [65] and clustering [74] tasks.	It is widely used for clustering tasks, dimension reduction or data representation [15].	It features training efficiency and simple implementation for multiclass classification problems [15].



(a) (b)

Fig. 7. X-ray images of patients who are COVID-19 positive or suspicious [118].

amount of these rays [115]. In this regard, the most common use of X-rays is to see fractures (broken bones) [114], but it also has great utility in detecting pneumonia [116] and COVID-19 [117], as shown in Fig. 7.

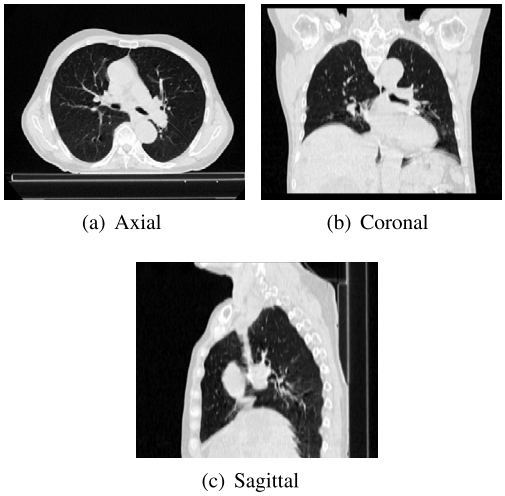
It is important to note that the main disadvantage of this type of modality is the exposure to ionizing radiation experienced by the patient when performing this type of radiological study. On the other hand, as an advantage, these X-ray images can reveal very subtle features, are generally inexpensive to produce, and the imaging procedure is simple [113].

3.1.2. Computed tomography

Computed tomography (CT) is a popular imaging technique for the diagnosis of many different disorders. This imaging technique reveals the bones, blood vessels, lungs, brain, heart, abdomen, and soft tissues in the pelvis. In addition, CT is frequently used to diagnose a variety of cancers, including pancreatic, lung, and liver cancer. The images created in a CT procedure are represented by thin slices of the body (a two-dimensional (2D) image). The complete sequence of images produced in a CT scan, i.e., the union of all the slices, forms a volume [three-dimensional (3D) image] [119].

Nowadays, CT requires multi-detector equipment, which provides a volumetric examination of the entire organ under study in a few seconds, with sub-millimeter slice thicknesses. This allows obtaining images in different visualizations of the 2D space, called the axial, coronal and sagittal orthogonal planes [120], with isotropic resolution as shown in Fig. 8.

When the CT imaging modality is used in medicine, the process of scanning the organ to be studied involves sending photon radiation



(a) Axial (b) Coronal

(c) Sagittal

Fig. 8. Thoracic CT images in different planes [121].

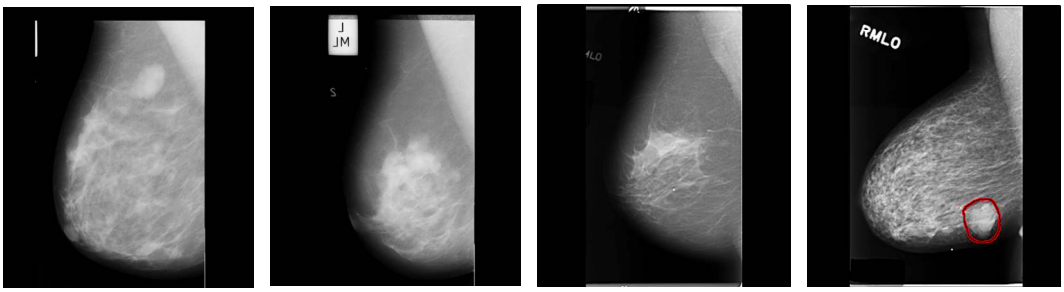
through the body from an X-ray tube that rotates around the patient and has a type of noise associated with it, which is statistically described by a Poisson distribution [122]. The attenuation along the path of such rays is measured by detectors [119].

There are certain methods for reconstructing the 3D images that comprise the volumes shown by CT, with the purpose of reducing noise and artifacts in the images [119]. These methods are classified as direct [123], algebraic-iterative [124] and statistical [125].

3.1.3. Mammography

Mammography (MG) is the standard imaging modality for breast cancer screening because it is easy to implement and standardize. It is used as a screening and diagnostic tool to examine the human breast using low-dose X-rays; it has the advantage of allowing reviews and direct comparison with previous mammographic studies [126].

The patient's breast is placed on a flat support plate throughout the MG acquisition operation, and the breast is compressed using a parallel plate. A brief burst of X-rays generated by an X-ray machine travels through the breast to a detector on the other side. The detector could be a photographic film plate for conventional mammography, which records the X-ray image on film, or a solid-state detector for digital



(a) (b) (c) (d)

Fig. 9. Sample of benign and malignant images from different datasets, such as MIAS [128]. [(a) Benign, (b) Malignant] and DDSM [129] [(c) Benign, (d) Malignant. In this case the shape of the lobulated mass with micro-lobulated margins is identified with a red circle].

mammography [127], which sends electronic signals to a computer to create a digital image. Fig. 9 shows digital MG images belonging to two databases from the MIAS [128] and DDSM [129] repositories which are described in Section 3.2.

There are three important techniques included in MG: digital mammography, computer-aided diagnostics (CAD), and digital breast tomosynthesis (DBT) [130]. A digital mammography has certain advantages since the processes of image acquisition, visualization, and storage are separated. This means that each of these processes can be optimized.

Next, CAD systems are used as a second reader, i.e., the computer flags candidate lesions, which should be re-evaluated by the clinician again. The goal of the CAD system is to automatically detect and diagnose candidate lesions within the mammographic images by providing an aid to the clinician looking for possible abnormalities [131]. On the other hand, breast tomosynthesis is a three-dimensional image representation of the breast using X-rays. This consists of taking several images of the compressed breast at multiple angles by scanning the X-ray tube which forms an arc around the breast [130].

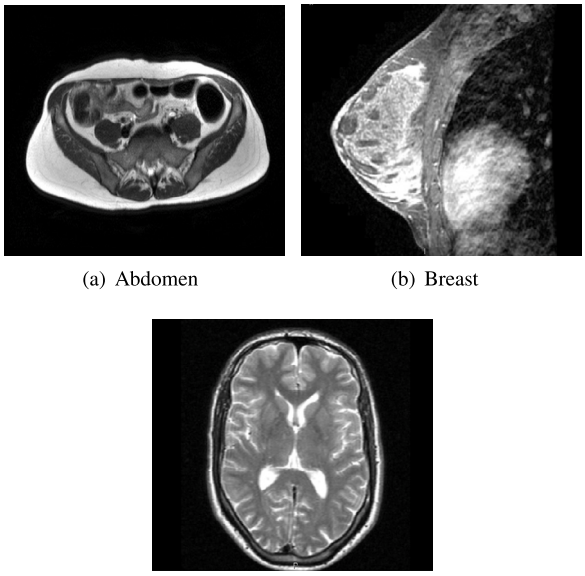
3.1.4. Nuclear magnetic resonance (NMR)

This modality makes it possible to create images with high spatial resolution without the danger of ionizing radiation, which is why its use has increased substantially in recent decades [132]. This type of imaging is very important from a clinical point of view because it provides morphological and functional images with a great variety of contrasts in tissues. By distinguishing between healthy and sick soft tissues in the body, magnetic resonance imaging (MRI) has been shown to be very useful in the diagnosis of a number of illnesses [133].

Additionally, images obtained by the MRI modality are also used to examine blood vessels, brain and breast tumors, abnormal tissues, spinal injuries, and other problems. Fig. 10 shows MRIs of the abdomen, breast, and brain. This type of modality can also be visualized in the axial, coronal and sagittal planes.

MRI produces finely detailed images of organs and tissues using radio frequency pulses and a magnetic field. These images are taken in a plane that is orthogonal to the indicated plane in question. By employing an intense magnetic field within the human body, different proton orientations are induced [134]. Basically, the MRI acquisition process consists of placing the patient on a gurney and introducing them into a large doughnut, which creates a magnetic field that acts to align the protons in the body. After the process of the magnetic field aligning the hydrogen protons, radio frequency waves are emitted from external equipment and transmitted to the patient, causing the alignment of the protons to change, e.g., at angles of 90° or 180° [135]. The protons in the patient's body then return to their neutral state once these radio frequency pulses are shut off, generating their own radio wave signals that are picked up by the receiver coils and used to create an image [135].

One of the advantages of this type of modality is that neither patients nor operators are exposed to ionizing radiation. On the other hand, the main disadvantages are its high cost and the impossibility of accessing the patient during the acquisition process. Another disadvantage, which is very important, is the fact that the scanner attracts ferromagnetic objects, turning them into projectiles that could cause serious injuries to the patient or the operator, as well as damaging the scanner.



(a) Abdomen (b) Breast

(c) Brain

Fig. 10. NMR images in different organs [121].

3.1.5. Ultrasound

Ultrasound (US) is a diagnostic technology that uses high frequency sound waves to image human bodily structures. Since the frequencies are higher than 20,000 Hz, the human ear cannot hear them. A specific gel is applied directly to the subject's skin to avoid air pockets from accumulating between the transducer and the skin, which could prevent the ultrasonic waves from entering the body. This makes it feasible to display elements like blood flow and organ movement that a still X-ray image cannot. According to [136], US pictures are highly helpful in the diagnosis and treatment of various diseases, including the treatment of prostate cancer, the detection of breast cancer, the diagnosis of liver tumors, and the diagnosis of esophageal cancer, among others. The pancreas, liver, and kidney are depicted in the US pictures in Fig. 11.

By converting electrical energy into mechanical energy, an ultrasonic transducer may often produce acoustic waves. The piezoelectric effect is the most effective way to do this when turning electrical energy into mechanical energy to produce medical images utilizing US. In a US scanner, a transducer projects a beam of sound waves into

(a) Pancreas (b) Liver (c) Kidney

Fig. 11. US images in different organs [121].

the body and these waves are reflected back to the transducer at the boundaries between tissues in the beam's path, such as the boundary between fluid and soft tissue or tissue and bone. Electrical signals are produced and supplied to the ultrasound scanner when these echoes hit the transducer [137,138].

Using the sound speed and the time between each echo, the scanner determines the distance between the transducer and the tissue boundary. These distances are utilized to produce 2D pictures of tissues and organs [138,139]. This modality has been widely used because of its low cost, as well as the fact that it has no known side effects for patients.

Clinicians sometimes need a 3D view of anatomical structures, which limits 2D imaging. Because of this, and the advances that have been made in the development of US-based techniques, 2D images have gradually been replaced by 3D US images. These 3D US images can be obtained in three main stages: acquisition, reconstruction, and visualization.

On the other hand, medical imaging is important in today's diagnostic medicine because it allows the detection, diagnosis, follow-up, and monitoring of a variety of diseases that affect humans. Table 4 shows the main advantages and disadvantages of each of the imaging modalities described in this section. Currently, great efforts are being made in the development of procedures and methodologies that allow the reduction of such disadvantages and exploiting the advantages of the benefits of each of the imaging techniques in each area of use.

3.2. Public databases in medical imaging

Databases for scientific research are of great importance because they contribute to its methodological development, both qualitative and quantitative. These databases are often difficult to obtain because many of them are private, which represents a difficulty when developing new research.

Working with public databases allows a comparison of research results. However, until now, there was no registry of the available repositories characterizing these datasets. Therefore, in this part of our review, we propose to present a description of the most used public databases, classified by organ type (lung, brain, skin, breast, cervix, and eye) (see Fig. 12).

3.2.1. Brain data sets

Four databases with magnetic resonance (MR) imaging modalities were registered. The databases considered were:

1. **TCIA-Brain**: contains images of early-diagnosed primary glioblastoma that were treated with surgery and chemoradiotherapy [140].
2. **BRATS**: Was initially built by the BraTS image segmentation challenge which started in 2012 [141]. This challenge has been carried out annually with the aim of evaluating the state-of-the-art segmentation and classification method.
3. **ADNI**: The Alzheimer's disease neuroimaging initiative dataset, conducted its studies on subjects aged 18 to 96 years. This database contains sets of synthetic images, cognitive test results, cerebrospinal fluid samples and blood biomarker tests, as a set of possible predictors of the disease. All subjects were grouped into four different classes: (i) Alzheimer's disease; (ii) converting; (iii) non-converting; and (iv) normal control.
4. **MCHD**:Magnetic resonance scans of the brain are contained in the Multiclass Harvard Dataset (MCHD) collection. Each of the four illness types represented by the pathological brains (stroke, degenerative, infectious, and brain tumor) has 40 pictures [142].

Table 5 presents each dataset, indicating their location on the web, together with the amount of data and subjects.

3.2.2. Cervix data sets

Three databases were collected, of which only one is reported in the articles reviewed [143,144]. Table 5 shows the URLs of the repositories and the number of images contained in each of them. A brief description of each is given below:

1. **Herlev**: Developed at Herlev University Hospital (Denmark), contains optical colposcopy images (OCC) divided into 242 images for normal cases, and 675 images for abnormal cases.
2. **TCGA-CESC**: A bigger initiative to create a research community focused on linking cancer characteristics to genetics includes the collecting of data for Cervical Squamous Cell Carcinoma and Endocervical Cell Adenocarcinoma (TCGA-CESC).
3. **ca_cervix**: This database contains 19 attributes, which are related to the risk of cervical cancer behavior with a class label ca cervix with 1 and 0 values.

3.2.3. Breast data sets

Four databases were found, all of which were used in some of the studies reviewed and are summarized in Table 4. The databases are:

1. **MIAS**: Mammographic Image Analysis Society, a digital mammography database which contains a variety of breast cancer images.
2. **DDSM**: In this database, each breast is represented by two images, along with some patient data (age at the time of the study, American College of Radiology breast density classification, subtle classification of abnormalities, etc.).
3. **WBCD**: The Wisconsin Breast Cancer Diagnostic Database classifies breast cancer into benign and malignant.
4. **INbreast**: This database acquired images in a Breast Center, located in a S. Joao Hospital (A Joa, Breast Center, Porto, Portugal). It consists of 90 cases of women with both breasts (4 images per case) and 25 cases of patients with a mastectomy (2 images per case).

Table 4
The listed imaging modalities' benefits and drawbacks.

Imaging modalities	Advantages	Disadvantages
Radiography	High availability in medical centers. It is economical. Low dimensionality of the data which implies reduced computation time.	Superposition of structures and/or tissues. Has ionizing radiation. Impossibility of obtaining sharpened images. Artifact imperfection.
Computed tomography	Greater detail of most of the anatomical structures of the human body. Reduced image acquisition time. Possibility of obtaining contrasted images.	Low contrast, Poisson-type noise, artifacts such as stair-step and streak. Less detail in soft tissue. Has ionizing radiation. High dimensionality of the data which implies high computational time.
Mammography	Early detection of breast cancer. Low dimensionality of the data which implies reduced computation time.	Painful for the patient. Poisson-type noise imperfections.
Magnetic resonance	Generates better quality images than other modalities. No ionizing radiation. Better detail in soft tissue. Possibility of obtaining contrasted images.	High dimensionality of the data which implies high computational time. Long scanning time. Imperfection of riciano noise. Partial volume and ring-type imperfections. It is expensive. It cannot be used in patients with metallic devices, such as pacemakers.
Ultrasound	Widely used. It is economical. No ionizing radiation. Possibility of obtaining contrasted images. Low dimensionality of the data which implies reduced computation time.	Depends on an operator. Moderate resolution. Very low signal-to-noise ratio. Speckled noise imperfection.

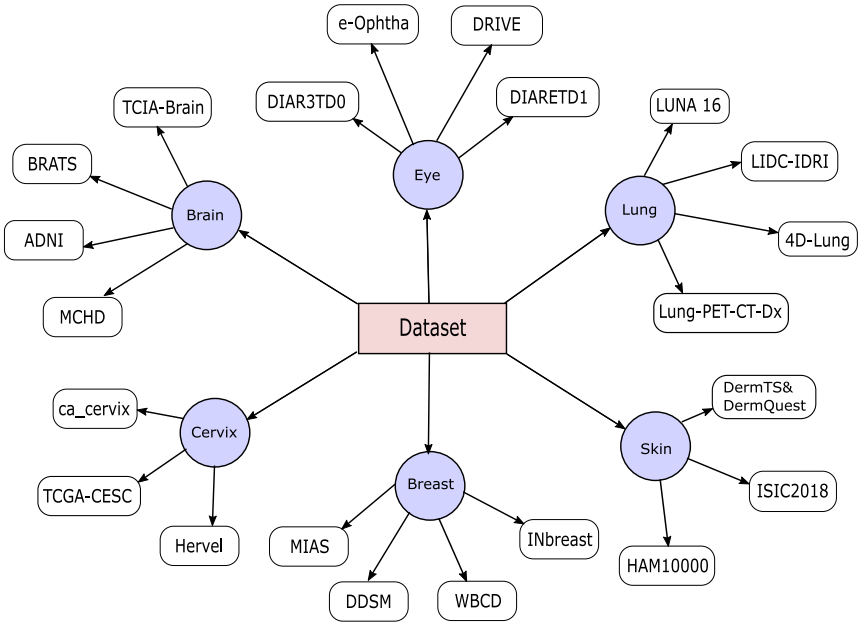


Fig. 12. Classification of the most used public databases in the reviewed literature.

3.2.4. Skin data sets

Four databases were compiled with the dermoscopy (DS) modality, two of which were used in the various studies reviewed in this document. Some of the information from these databases are summarized in Table 5. The databases considered are:

1. **HAM10000**: The Human Against Machine data set has 10,000 training images used to detect pigmented skin lesions.
2. **ISIC2018**: The International Skin Imaging Collaboration released this dataset. Lesion segmentation, attribute identification, and disease categorization are among the three tasks it completes.
3. **DermIS&DermQuest**: The DermIS dataset consists of skin cancer images, consisting of 43 melanoma cases and 36 normal cases. On the other hand, the DermQuest database consists of 76 melanoma cases and 58 normal cases.

3.2.5. Lung data sets

Three CT modality databases were collected, only one of which is reported in the articles reviewed. Some of the information from these databases is summarized in Table 5. The databases considered are:

1. **LIDC-IDRI**: Four radiologists independently assessed each image in the Lung Image Database Consortium (LIDC-IDRI) image collection and marked lesions that fit into one of three categories.
2. **LUNA-16**: The 2016 lung nodule analysis dataset (LUNA-16) was an open challenge conducted during 2016. For this challenge they used part of the LIDC/IDRI database and built an annotation file containing 1186 nodules.
3. **Lung-PET-CT-Dx**: Consists of lung cancer patient images in DICOM format, with the locations of each tumor identified by five academic thoracic radiologists with knowledge of the disease. This dataset can be used to construct algorithms for medical diagnosis.
4. **4D-Lung**: Images taken during chemoradiotherapy on patients with locally advanced non-small cell lung cancer make up this data collection.

3.2.6. Eye data sets

Four databases with ophthalmoscopy (OS) modality were compiled, two of which were used in the various studies reviewed in this document. Some of the information from these databases is summarized in Table 5. The databases considered are:

1. **DIARETDB0**: The goal of the calibration level 0 fundus images dataset (DI-ARETDB0) is to compare the ability of digital pictures to diagnose diabetic retinopathy (DR). The collection includes color fundus images, of which 110 show indications of DR and 20 are normal (hard exudates, soft exudates, micro-aneurysms, hemorrhages and neovascularization).
2. **DIARETDB1**: The calibration level 1 fundus imaging dataset (DIARETDB1) consists of color fundus images, 84 of which have at least mild non-proliferative indicators (micro-aneurysms) of DR, and 5 of which, in the opinion of all specialists involved in the evaluation, are normal and have no signs of DR.
3. **e-ophtha**: Is a database specially designed for scientific research in RD.
4. **DRIVE**: The digital retinal image dataset for vessel extraction consists of 7 cases of abnormal pathology.

Other public databases of the aforementioned bodies and others that may be of interest can be consulted at [145].

3.3. Metrics for performance evaluation of ELMs in the context of medical imaging

Metrics have been defined to evaluate the performance of a classifier applied to medical imaging, including: sensitivity (Sn), specificity (Sp), accuracy (Acc), area under the curve (AUC), F1 score, Dice coefficient (Csc), true positive rate, and false positive rate [146–148]. Table 6 summarizes three of the most commonly used metrics in the reviewed literature.

Sections 4–6 will present a description of the state-of-the-art related to ELM-based medical image processing and the types of supervised, unsupervised, and semisupervised learning. For each article reported, an analysis will be made emphasizing its methodology, modality, database, performance metrics (those that will not be reported in the summary tables), and results. Additionally, in each section, a summary table will be presented, showing the location of the pathology, year, authors, methodology, database, modality, and results. Within the results, only Sn, Sp, and Acc are considered, since they are the most reported metrics in the articles reviewed.

4. Medical image processing using supervised extreme learning machines

Usually, the data in medical imaging needs to be processed first, so that ELMs can function more effectively. In this regard, Lama et al. [149] examined various structural magnetic resonance imaging (smart MRI) methods for diagnosing Alzheimer's disease (AD) (sMRI). They used an SVM, an import vector machine (IVM), and a regularized extreme learning machine (RELM) to distinguish between AD, moderate cognitive impairment, and healthy control participants. They took into consideration the greedy score-based feature selection technique. They used the ADNI dataset to compare the performance of various classifiers and the results are shown in Table 7. In healthy, mildly cognitively impaired, and control participants, the trials showed that RELM can greatly increase the classification accuracy of AD with the feature selection strategy.

Zhu et al. [150] relied on a supervised ELM method to segment retinal vessels present in the public DRIVE database, as shown in Fig. 13, for which attributes such as width, tortuosity, branching pattern, and angles were taken into account. These attributes play an important role in the early diagnosis of the disease. The results of the experimental tests, shown in Table 7, demonstrate that the proposed method is more effective than alternative techniques for segmenting retinal arteries.

kuppili et al. [151] studied an ELM-based tissue characterization system for risk stratification of US images of livers. The authors considered four types of Kfold cross-validation protocols and stratified images of fatty liver disease present in the US modality using ELM and compared them with SVM. A liver database of 63 patients (27 healthy/36 diseased) was used. The findings showed that ELM outperforms SVM in terms of Sn, Sp, Acc, and AUC for all cross-validation methods and all classes of datasets.

Punithavathi et al. [152] proposed a method to diagnose the severity of DR, as shown in Fig. 13. Preprocessing, feature extraction, and classification were performed using morphological processing. Classification was performed using the ELM network. The images used were from the DRIVE and DIARETDB0 databases. The research showed an accuracy of 0.95 with a good training speed.

On the other hand, Luo et al. [153] developed an automatic left ventricle segmentation method using a Hierarchical Extreme Learning Machine (H-ELM). This network can achieve more compact and representative feature representations to perform segmentation tasks considering MR images. The private dataset consisted of cardiac MRI from Shenyang Military General Hospital, where there were a total of 30 cases (19 males, 11 females), including both healthy subjects and patients affected by diseases related to heart failure, obstructed coronary systems, and hypertrophy. The performance measurements in this study were segmentation time, data similarity coefficient (Dsc), mean absolute deviation (MAD), and maximum absolute deviation (MAXD). The results obtained, concerning the average segmentation of the proposed model, were MAD=2.31, MAXD=6.21, and Dsc=87.8, with a time of 0.36 s.

Toprak [154] developed his research on the classification of benign and malignant cells in breast cancer with an ELM. The data set used was WBCD. The results showed best performance at an Acc=0.98. This work highlights the excellent performance of the proposed ELM, considering accuracy and speed.

For the purpose of determining the ELM's input weights and hidden neurons' ideal values, Eshtay et al. [155], employed a competitive swarm optimizer (CSO). Both the regularized version (CSO-RELM) and the traditional ELM (CSO-ELM) were taken into consideration. The studies were conducted using 15 biometric medical categorization datasets that were taken from the UCI library (breast, diabetes, heart, etc.) The proposed model can achieve greater generalization performance with fewer hidden neurons and more stability, according to experimental results. In addition, it takes much less time to train than other metaheuristic-based ELMs.

Table 5
List of public brain databases and corresponding URLs.

Organ/Modality	Database	Subjects/Images	Location
Brain/MR	TCIA-Brain	20/8798	https://wiki.cancerimagingarchive.net/
	BRATS	2000/8000	http://braintumorsegmentation.org/
	ADNI	6000/–	https://adni.loni.usc.edu/
	MCHD	–/200	http://www.med.harvard.edu/AANLIB/
Cervix/OCC	Herlev	–/917	https://paperswithcode.com/dataset/herlev
	TCGA-CESC	–/19135	https://www.cancerimagingarchive.net/
	ca-cervix	–/72	https://archive.ics.uci.edu/ml/index.php
Breast/MG	MIAS	161/322	https://www.repository.cam.ac.uk/handle/1810/250394
	DDSM	2620/–	http://www.eng.usf.edu/cvprg/Mammography/Database.html
	WBCD	569/–	https://archive.ics.uci.edu/ml/datasets/breast+cancer+wisconsin+(diagnostic)
	INbreast	115/410	https://www.kaggle.com/datasets/martholi/inbreast
Skin/DS	HAM	–/10015	https://www.nature.com/articles/sdata2018161
	ISIC2018	–/12500	https://challenge.isic-archive.com/data/#2018
	DermIS	–/69	https://challenge.isic-archive.com/data/#2018
	DermQuest	–/134	https://www.isic-archive.com/#!/topWithHeader/wideContentTop/main
Lung/CT	LIDC-IDRI	1010/244527	https://https://www.cancerimagingarchive.net/
	LUNA-16	601/888	https://luna16.grand-challenge.org/Data/
	Lung-PET-CT-Dx	355/251135	https://https://www.cancerimagingarchive.net/
	4D-Lung	20/347330	https://https://www.cancerimagingarchive.net/
Eye/OS	DIARETDB0	–/130	https://www.it.lut.fi/project/imageret/diaretdb0/
	DIARETDB1	–/89	https://www.it.lut.fi/project/imageret/diaretdb1/index.html
	e-ophtha	–/463	https://www.it.lut.fi/project/imageret/diaretdb1/index.html
	DRIVE	–/40	https://drive.grand-challenge.org/

Table 6
Metrics most commonly used in the reviewed literature. In the formulas we denote by FP the number of false positives, TP the number of true positives, FN false negatives and TN true negatives.

Metrics	Definition	Formula
Sn	Measures the classifier's ability to identify positive examples.	$\frac{TP}{TP+FP} \times 100\%$
Sp	It represents the classifier's ability to identify negative examples.	$\frac{TN}{TN+FN} \times 100\%$
Acc	Accuracy represents the degree to which the predicted value is close to the true value.	$\frac{TP+FN}{TP+FN+TN+FP} \times 100\%$

The authors in [156–160] made use of supervised algorithms, for the classification of medical brain images, as shown in Fig. 13. Nayak et al. [156] employed a combined technique between Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) to find more discriminative and reduced feature sets. Finally, an ELM called 'modified sine cosine algorithm' (MSCA-ELM) was developed to classify MRI into two categories: diseased and healthy.

Gumaei et al. [157] used a normalized descriptor with the PCA feature extraction method to obtain the properties of brain tumors. Rahmat et al. [158] was created a methodology for feature extraction and categorization using a gray-level co-occurrence matrix. Devanathan and Kamarasan [160] was created an automated method for identifying and classifying brain tumors (RN-OKELM) and a residual network model (Res2Net) was applied as a feature extractor. The results are presented in Table 7.

Interesting works such as [161,162] focus on breast cancer diagnosis using a supervised ELM approach, as shown in Fig. 13. The authors reported that the ELM classifier was more efficient, compared to other reported classifier algorithms.

Aslan et al. [163] designed an ELM for retinal blood vessel detection from the DRIVE database. First, color retinal image features were extracted using adaptive thresholding, Gabor filtering, and top-hat transform techniques. These features constitute the input vectors of the ELM, which presents an accurate and efficient performance. The results obtained are shown in Table 7.

Huang et al. [164] worked on the automatic detection of neovascularization in retinal images using an ELM. A series of filters was used to obtain neovascularization features from retinal images. The detection approach was evaluated with images annotated by expert ophthalmologists: 16 MESSIDOR images, 21 DiaRetDB0 images, and 5 DiaRetDB1 images. The results obtained are shown in Table 7. The proposed approach shows suspicious regions of neovascularization and efficiently supports the ophthalmologist's decision making.

The authors in [165–168] studied image classification in the eye area, as shown in Fig. 13. In [165], was used the Modified Adaboost Extreme Learning Machine to perform supervised extraction of blood vessels from retinal images (MAD-ELM). They extracted and selected features using a maximum relevance and minimum redundancy approach. Moreover, in [166] was performed a feature extraction process with tetragonal local octal patterns. Classification of DR stages was performed by ELM.

In [167] the authors based their research on automated glaucoma classification using discrete wavelet transform and gradient histogram functions with ELM. In [168] was used an ELM as a classifier, to address the issues of early identification and categorization of healthy retinal images, as well as neovascularization in the optic disc and elsewhere. The results obtained are shown in Table 7.

Using ELM, Sousa et al. [169] described a quick technique for retinal vessel detection. The authors' suggestion is a concise and effective representation of retinal pictures that preserves representativeness while reducing the original data storage by up to 39%. To achieve this reduction, three features (local top hat, local average, and local variance) were combined. Images from the DRIVE, STARE, and HRF datasets were used in the simulations. The simulations showed that this approach outperformed most of the strategies examined in the study, achieving an average accuracy close to 0.95 (see Table 7).

The authors in [170–173] worked on breast image classification, as shown in Fig. 13. In [170] was performed ELM-based classification

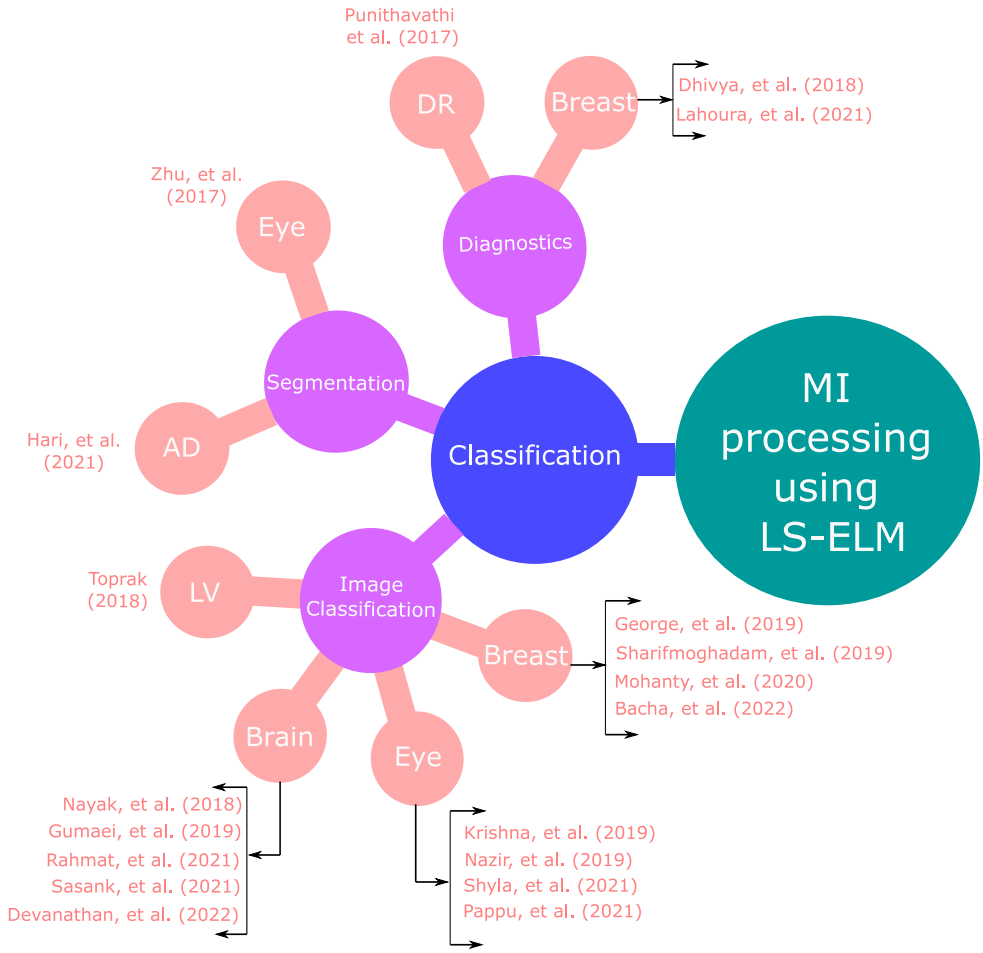


Fig. 13. Summary of some articles with supervised learning.

to detect microcalcifications in mammograms using multiscale features and applied different features, extracted with the Gabor filter. On the other hand, in [171] was proposed an algorithm for breast cancer classification based on AdaBoost and ELM algorithms. A threshold value was defined to reweight part of the misclassified data. In [172,173] was created a CAD-based model to classify digital mammograms as normal, benign or malignant. The Haralick method, PCA method (to reduce the dimensionality of the data) and discrete Chebyshev transform were used to extract features. The results are shown in Table 7.

Tandungan et al. [174] developed a technique for categorizing lung nodules using ELMs. The ELM algorithm was used to determine the ideal configuration and predict the degree of malignancy in lung cancer nodules using a number of hidden nodes and multiple activation functions. This investigation made use of the LIDC-IDRI database, and Table 7 presents the results that were obtained. The suggested algorithm successfully categorized lung nodules.

Ren et al. [96] proposed an ELM method based on the L_{21}-norm, in an effort to increase the robustness of the ELM (L_{21}-ELM). To increase robustness and reduce the impact of noise and outliers, the approach applied the L_{21}-norm on the loss function. The Cancer Genome Atlas (TCGA) and scRNA-seq datasets were used to classify cancer samples according to the proposed model. Seven cancer types served as the basis for the experiments: colon adenocarcinoma (COAD), esophageal carcinoma (ESCA), head and neck squamous cell carcinoma (HNSC), lung adenocarcinoma (LUAD), lung squamous cell carcinoma (LUSC), ovarian serous cystadenocarcinoma (OV), and pancreatic adenocarcinoma (PAAD). The results reported by the authors were contrasted with those of CNN, traditional ELM and RELM (the three other widely used networks). The results showed that $L_{2,1}$-ELM is more reliable and efficient than other approaches.

On the other hand, the authors in [175,176] focused their work on brain tumor detection, Alzheimer's disease detection and the classification of primary and secondary brain cancers, respectively. These authors applied the ELM algorithm considering MRI. For the case of [175], the authors applied the Gabor filter to extract texture features. The results obtained by these authors are shown in Table 7.

Pi et al. [147] used a mixed model to diagnose COVID-19. Prior to using ELM for classification, the gray co-occurrence matrix was employed to extract picture features. A total of 640 images, including 320 CT scans of the lungs taken from patients with COVID-19 and 320 control images, were used in this investigation, see Table 7 for the results. The proposed methodology, which the authors demonstrated as being workable, can aid medical professionals in accurately identifying probable COVID-19 patients for additional isolation and therapy.

A study, proposed by Govindarajan et al. [177], used integrated local feature descriptors and classifiers, such as ELMs and OS-ELMs, to detect and categorize tuberculosis situations in healthy patients, by using chest radiographs (X-rays). The digital X-ray pictures were taken from the Healthy-80/PTB-58 dataset (Table 7). In both healthy and TB pictures, the suggested approach can identify the lung fields. For the computational detection of pulmonary tuberculosis, the suggested methodology seems to be helpful.

By using ELM on Kaggle X-ray pictures, Nahiduzzaman et al. [116] created an automatic pneumonia diagnosis method. Three models were investigated: CNN-PCA-ELM with increased contrast on the X-ray pictures, ELM with a hybrid convolutional neural network principal component analysis (CNN-PCA), and classification using an ELM. The latter

model suggested an approach offering a positive outcome, earning a recall score of 0.98 and an accuracy score of 0.9832, for the multiclass pneumonia classification. A binary classification, on the other hand, earnt a 0.10 recovery score and a 0.9983 accuracy score.

Khan et al. [178] put out a technique to use ELM to predict COVID-19 infection in CT chest images. The database was initially set up comprising 2500 normal CTs, 3000 images with COVID-19 pneumonia, and 5500 photos from the Radiopedia COVID-19 website. The network's predictive power was then enhanced through canonical correlation analysis, to combine the various features that had been taken from the training data into a single vector. The ELM classifier's top prediction accuracy was 0.939. This study documented the successful categorization of COVID-19 infections using CT scans and ELM.

Suksmono et al. [179] created a cervical cancer classification algorithm utilizing a mix of gray-level co-occurrence matrix (GLCM) and ELM approaches, continuing the same line of cancer classification. The data used in this study were gathered from cervical cell microscope slides. 76 photos of squamous cells, adenocarcinoma, and high squamous intraepithelial lesions were taken into account. A training accuracy of 1 and a test accuracy of 0.80 were determined from the program's findings.

Next, Al-Hammouri et al. [180] used an ELM for the early diagnosis and treatment of melanoma, which is one of the most dangerous skin cancers in the world. The methodology of this research consisted of melanoma image processing, then segmentation, feature extraction, and, finally, ELM classifier. The dataset was generated by collecting labeled (Normal/Melanoma) dermoscopy (DS) type images, consisting of 70 melanoma (M) and 100 nevus (N) images from the digital image archive of the Dermatology Department of the private University Medical Center Groningen (UMCG). The results obtained reported an accuracy of 0.97 for ELM performance.

On the other hand, Elsayed et al. [181], developed a method for predicting diabetes using an ELM based on data from a questionnaire that can advise users to seek medical attention promptly, avoiding delayed diagnosis and the emergence of catastrophic conditions. It made use of a multivariate dataset that is accessible through the IEEE Dataport and UCI repository and comprises information on the signs and symptoms of newly diagnosed or potentially diabetic patients. The accuracy of the model was 0.9807 and there were no false positive rates.

Table 7 summarize most of the antecedents presented above, highlighting the organ, reference, year, algorithm, dataset, modality, and results.

5. Medical image processing using unsupervised extreme learning machines

This section presents an overview of the state-of-the-art related to ELM-based medical image processing and unsupervised learning.

To enhance automatic lung nodule diagnosis and to automatically carry out feature extraction, model training, and lung nodule detection, Hao et al. [182] presented an H-ELM-based computational model. Following that, the transformed images were input into an unsupervised multilayer automated encoder built on an ELM, to produce the lung nodule image's more compact and insightful high-level characteristics. Finally, the ELM classifier was updated with the implemented features. The dataset was built using 2800 CT scans of lung nodules. In Table 8, the results for BRATS 2014 are displayed.

Lingappa et al. [183] suggested a method for medical picture segmentation based on the extreme learning machine algorithm, using the artificial bee colony (ABC) method and Kernel Fuzzy C-means (KFCM) clustering methodology. Due to pattern ignorance, related objects were first grouped into a single division. Then, because of the subspace plane of the membership's typical structure, data misclassification was carried out. Finally, the efficiency of the suggested approach was assessed, based on a comparison of the centroids, segmentation accuracy, the pixel error of K-means, spatial FCM, and KFCM algorithms. The comparison was undertaken using a medical image from the BraTS collection. In comparison to the previous approaches, the proposed ABC-ELM-KFCM methodology increased segmentation accuracy up to 0.9703.

Wang et al. [103] presented a method for breast cancer detection using an ELM based on feature fusion with deep attributes (SF) from a neural network. The model presents a mass detection method based on deep CNN features and unsupervised ELM clustering. A feature set was built that fuses deep attributes, such as morphology, texture, and density, and then an ELM was implemented using the fused feature set to classify breast masses into benign and malignant. Four hundred mammograms were used in the image dataset, which contains two hundred images of malignant masses and two hundred images of benign masses. Additionally, detection and diagnostic metrics were considered. The reported results concerning SF with CNN are shown in Table 8.

In the context of ELMs, an important network topology is the multilayer ELM (ML-ELM) network. Nayak et al. [142] built the ML-ELM+LReLU model using this network with a leaky rectified linear unit (LReLU) activation function, to do away with the necessity for manual feature extraction and provide a more reliable and universal system for multiclass brain MRI classification. This study made use of the MCHD dataset. A training time of 46.13 s and an accuracy of 0.9375 were included in the results, outperforming their counterparts. A feature extractor and classifier are both functions of the suggested ML-ELM+LReLU.

The method proposed by Shiney and Rose [143] was based on the automatic detection and classification of cervical cancer in Pap smear images. These images come from the public Herlev Pap smear dataset. The methodology is based on image preprocessing, segmentation, feature extraction and classification. For preprocessing, an anisotropic diffusion filter was applied to remove noise present in the images. Segmentation was then performed using the proposed advanced map-based superpixel segmentation (AMBSS) algorithm. The results show that the accuracy (0.991) of the classification outperforms that of other previously applied intelligent methods, such as support vector mine and Quasi-Newton-based feed-forward neural networks.

Mohan et al. [184] suggested a brand-new approach based on modified features, a developing method to extract minutiae, and AE-ELM for reliable and quick localization of exudates (EX), which are important indicators of DR. The suggested method's key steps are dimensionality reduction using AE-ELM and preprocessing optic disc localization. Open-access retinal, DIARETDB0, DIARETDB1, e-Ophtha, MESSIDOR, and the local retinal database gathered from Silchar Medical College and Hospital were the databases taken into consideration (SMCH). Table 8, with a processing time of 3.19 s per image, displays the results.

Table 8, summarizes each of the antecedents presented above, highlighting the organ, reference, year, algorithm, dataset, modality, and results.

6. Medical image processing using extreme learning semi-supervised machines

This hybrid modality of supervised and unsupervised has allowed the development of high impact research, some of which will be explained below.

Wang et al. [185] proposed a lung nodule CAD, based on the SS-ELM model. The authors compared the proposed method with an SVM, a probabilistic neural network (PNN), and a multilayer perceptron (MLP). The dataset consisted of LIDC-IDRI, to test the effectiveness of the algorithm. From these, textural and morphological features of lung nodules were extracted. The results have shown that the proposed model achieved a higher accuracy of 0.9591. Overall, the proposed algorithm showed better performance at a faster learning speed and higher test accuracy than ELM, SVM, PNN, and MLP.

Table 7
Prior works related to medical image processing using SL-ELM.

Organ	Reference	Year	Algorithm	Dataset	Modality	Results (Sn/Sp/Acc)
Brain	Lama et al. [149]	2017	SVM/IVM/ RELM	ADNI	MR	0.6170/0.9063/0.7661
	Nayak et al. [156]	2018	ELM/MSCA	2550 images	MR	0.9914/0.9915/0.9986
	Gumaei et al. [157]	2019	RELM	3064 images	MR	–/–/0.9423
	Sharif et al. [175]	2020	ELM	BRATS	MR	0.98/1/0.99
	Hari et al. [176]	2021	ELM	OASIS1	MR	–/–/0.8002
	Rahmat et al. [158]	2021	ELM	100 images	MR	–/–/0.9667
	Devanathan y Kamarasan [160]	2022	RN-OKELM	253 images	MR	–/–/0.9793
	Sasank and Venkateswarlu [159]	2021	KSELM	BRATS	MR	0.94/0.963/0.9684
Breast	Toprak [154]	2018	ELM	WBCD	MG	–/–/0.9899
	Mohanty et al. [172]	2020	KELM	MIAS/DDSM	MG	–/–/0.9749
	Dhivya et al. [161]	2018	ELM	AASC	MG	–/–/0.92
	Lahoura et al. [162]	2021	ELM	WBCD	MG	/–/–/0.9868
	Bacha et al. [173]	2022	LELM	SAIM/INbreast	MG	–/–/0.9993
	Sharifmoghadam et al. [171]	2019	ELM	WBCD/UCI	MG	–/–/0.9962
	George et al. [170]	2019	ELM	–	MG	–/–/0.94
Eye	Pappu et al. [168]	2021	ELM	DRIVE/ STARE/ DIARETDB1	OS	0.985/1/0.98
	Zhu et al. [150]	2017	ELM	DRIVE	OS	0.7140/0.9868/0.9607
	Huang et al. [164]	2018	ELM	MESSIDOR/DiaRetDB0/ DiaRetDB1	OS	–/–/0.892
	Krishna et al. [165]	2019	MAD-ELM	DRIVE/DR HAGIS	OS	0.7435/0.9836/0.9619
	Nazir et al. [166]	2019	ELM	Kaggle-DR/ DRIVE/ Review-DB/STARE	OS	–/–/0.996
	Shyla et al. [167]	2021	ELM	DR HAGIS/HRF	OS	–/–/0.965
	Punithavathi et al. [152]	2017	ELM	DIARETDB0/ DRIVE	OS	–/–/0.95
	Aslan et al. [163]	2018	ELM	DRIVE	OS	–/–/0.9459
	Sousa et al. [169]	2019	ELM	DRIVE/STARE/ HRF	OS	0.9746/0.7816/0.95
Lung	Tandungan et al. [174]	2019	ELM	LIDC-IDRI	CT	0.8517/0.8558/0.8487
	Govindarajan et al. [177]	2021	ELM/ OSELM	Healthy-80/PTB-58	X-ray	0.9079/0.9932/0.9719
	Nahiduzzaman et al. [116]	2021	CNN-PCA-ELM	Kaggle X-ray	X-ray	–/–/0.9832
	Khan et al. [178]	2021	ELM	5500 images	CT	–/–/0.939
	Pi et al. [147]	2021	ELM	640 images	CT	0.7419/0.7781/0.7600
	Ren et al. [96]	2020	L_{21}-ELM	TCGA/ scRNA-seq	CT	–/–/0.9953
Cervix	Suksmono et al. [179]	2021	ELM	76 images	OCC	–/–/0.80
Skin	Al-Hammouri et al. [180]	2021	ELM	70 images (M)/ 100 images (N)	DS	–/–/0.97
Liver	kuppili et al. [151]	2017	ELM	63 patients	US	0.9423/0.9759/0.9674
Heart	Luo et al. [153]	2018	H-ELM	HGSM	MR	–

Table 8
Background on medical image processing using USL-ELM.

Organ	Reference	Year	Algorithm	Dataset	Modality	Results (Sn/Sp/Acc)
Brain	Nayak et al. [142]	2020	ML-ELM+LReLU	MCHD	MR	–/–/0.9375
	Lingappa et al. [183]	2018	ABC-ELM-KFCM	BraTS	MR	–/–/0.9703
Lung	Hao et al. [182]	2018	H-ELM	2800 images	CT	0.945/0.880/0.918
Breast	Wang et al. [103]	2019	CNN/USL-ELM	400 images	MG	0.7638/0.7612/0.7625
Cervix	Shiney and Rose [143]	2021	AMBSS/AE-ELM	Herlev	OCC	–/–/0.991
Eye	Mohan et al. [184]	2022	AE-ELM	DIARETDB0, DIARETDB1, e-Ophtha, MESSIDOR, SMCH	OS	0.965/0.964/0.97

Through the use of CNN and ELM, Pashaei et al. [186] performed a categorization method for brain cancers. It assessed the performance of the suggested strategy using a dataset made up of three classes and 3064 photos of three different cancer types. Meningioma, glioma, and pituitary tumors are among the forms of brain cancers seen on T1-weighted contrast-enhanced magnetic resonance imaging (CE-MRI). A variety of classifiers, including SVM, radial basis function, and other classifiers, were used to compare the outcomes of this CNN and KELM (KE-CNN) ensemble. These comparisons demonstrate that the KE-CNN has promising outcomes, with an accuracy of 0.9368 for the categorization of brain tumors.

She et al. [187]created the Safe-SSELM, a secure extreme learning machine, to classify electroencephalography (EEG) inputs. Following this method, a new level of risk is generated for each instance of unlabeled data, by comparing the similarity of the predictions provided by the ELM and SSELM algorithms using the Wasserstein distance. The SSELM objective function is then built up with a risk-based regularization term incorporated. The effectiveness of the suggested technique was assessed using the MI EEG dataset from BCI Competition IV Dataset 2a. The information came from MI EEG readings taken from 9 people, who were divided into 4 movement classes: tongue, feet, right hand, and left hand. Table 10 presents the experimental findings.

She et al. [188] then go on to present the hierarchical semi-supervised extreme learning machine approach (HSS-ELM), which is used to categorize motor imagery (MMI). The H-ELM approach's deep architecture was first used for automatic feature learning, and these new high-level features were then classified using the SS-ELM algorithm, which can make use of both labeled and unlabeled data information. In-depth tests were carried out on benchmark datasets and EEG MMI datasets to determine how well the suggested technique worked. The HSS-ELM approach was able to obtain greater classification accuracy (0.6776) when compared to a number of cutting-edge methods, including SVM, ELM, SAE, H-ELM, and SS-ELM.

Ronoud et al. [189] investigated two evolutionary techniques ($E_{(T)}$), $E_{(T)}$ -DBN-BP-ELM and $E_{(T)}$ -DBN-ELM-BP, which combine deep belief networks (DBN) with an ELM classifier and backward propagation to handle the first problem (BP). The second problem has been addressed in the proposed techniques, by applying the genetic algorithm (GA) for architecture optimization due to the wide solution space of DBN topologies. $E_{(TW)} - DBN$, the third approach put out in this study, employs GA to address both issues, allowing the topology and DBN weights $E_{(TW)}$ to evolve simultaneously. The two breast cancer datasets WBCO and WDBC are used to test the proposed models. Table 9 displays the outcomes DBN-ELM. The results show that the suggested approaches have very excellent diagnostic performance for classifying breast cancer.

A thorough ELM-based adaptive image classifier for pneumonia detection using chest radiographs is proposed Vijendran and Dubey et al. [190] for quicker diagnosis and earlier treatment. The OS-ELM network is compared against algorithms like CNN, SVM, and RELM, in order to assess its performance. For the MNIST dataset, there were 60,000 training images and 10,000 test images, while for the pneumonia chest X-ray dataset, there were 3516 training images and 2344 test images. The results were reported with an accuracy of 0.9095.

Fei et al. [191] suggested a projective model (PM) based on the ML-KELMPM method, which implements the parameter transfer approach in the classifier to carry out transfer learning in CAD, for US-based diagnosis of breast tumors. A bimodal US picture dataset with 264 pairs of US breast photos was employed (129 patients with benign tumors and 135 patients with malignant cancers). The Nanjing Drum Tower Hospital provided all of the bimodal images (DTN). Table 9 contains experimental findings on a set of bimodal US imaging data.

Research was carried out by Zhou et al. [192] in order to provide a unique method for automatic optical disc detection, dubbed a low-rank representation-based semi-supervised extreme learning machine (LRR-SSELM). First, the methodology involved identifying the optical disc from a semi-supervised learning perspective and resolving the issue of the sparse labeling of data. Second, the suggested model completely explored the nonlinear data by using a nonlinear classifier. DIARETDB0, DIARETDB1, and Messidor (three publicly accessible databases) were used to validate the performance of the proposed technique. The benefits and efficacy of the suggested approach are demonstrated by the outcomes of the experiments. Table 9 displays the experimental findings on a bimodal US picture dataset.

Huang et al. [148] investigated a diagnostic approach based on deep transfer convolutional neural network (DTCNN) and ELM, which combined the efficiency of two algorithms to handle the classification of benign and malignant nodules. First, after training with the ImageNet dataset, an optimum DTCNN was used to extract high-level characteristics of lung nodules. The classification of benign and malignant lung nodules was then undertaken using ELM. Two databases were employed to test the efficacy and efficiency of the suggested method: the private FAH-GMU and the public LIDC-IDRI. The experimental outcomes for the DTCNNELM model for the LIDC-IDRI dataset are displayed in Table 9.

Next, Ghoneim et al. [144] detected and classified cervical cancer cells using CNN and ELM. The CNN model was used to perform feature extraction through transfer learning and fine-tuning, and subsequent image classification was performed through ELM. The experimentation was implemented from the public database of Herlev University Hospital. The proposed CNN-ELM-based system achieved an accuracy of 0.995 on the detection problem and 0.912 on the classification problem. According to the authors, these accuracies were better than any previously reported.

On the other hand, Wang et al. [193] suggested a hybrid paradigm that combined interactive cross-task extreme learning with dual deep transfer learning (D2TL) (ICELM). Deep transfer learning and two-step deep transfer learning were used to extract high-level features. The classification performance in ICELM was then enhanced by using the high-level feature sets in combination as regularization terms. The database contained 134 histopathological pictures of breast cancer. Table 9 displays the results. In conclusion, the suggested approach appears to be a useful tool for identifying breast cancer in clinical settings.

Additionally, Saxena et al. [194] proposed a hybrid learning machine model to solve the class imbalance problem. The methodology consisted of patch generation, feature extraction and classification. A pre-trained ELM and a kernelized weighted ELM (KP-ELM) were employed for breast cancer CAD using histopathology (HP). Breast cancer histopathology images were extracted from the publicly available BreakHis and BisQue datasets. In this case, the authors did not indicate the size of the datasets. Within the results, they reported an accuracy of 0.9002.

For the purpose of extracting and classifying breast cancer features, Sannasi et al. [195] ooffered a CAD method that combined deep learning with ELM. The study improved the classification performance of ELM using the Crow-Search sinus–cosine optimization algorithm (SC-CSOA-ELM). The idea of transfer learning was used to extract robust features from the input mammograms. For this purpose, the work used the ResNet18, ResNet50 and ResNet101 architectures, which are the three most efficient CNN ResNet families. The INbreast dataset, which includes full-field digital mammography (FFDM) images, served as the input database for the evaluation. The performance of the proposed model is 0.9581 more accurate than other algorithms when comparing the research results with data from existing ELM methods and the K-nearest neighbors (KNN) approach.

For real-time COVID-19 diagnosis using X-ray pictures, Wu et al. [117] propose an evolving deep convolutional neutral network employing a hybrid ELM and sine-cosine. The proposed network is evaluated to the COVID-Xray-5k dataset, and the outcomes are supported by a study that compares it to canonical deep CNN, optimized ELM, ELM optimized by genetic algorithm, and ELM optimized by a whale optimization algorithm. With a final accuracy of 0.9883 on the COVID-Xray-5k dataset, the suggested method beats the benchmark comparisons and results in a relative error reduction of 2.33 percent when compared to a conventional deep CNN. It was determined that the suggested model is typically simpler and easier to apply.

Turkoglu [196], creates a Multiple Kernel ELM-based Deep Neural Network (MKs-ELM-DNN) system to recognize COVID-19 disease from CT images. Three key procedures make up the MKs-ELM-DNN model: deep feature extraction, multi-kernel classifier, and data augmentation. The CNN-based DenseNet201 architecture, which is based on the transfer learning methodology, was employed by the author. Additionally, a 746 picture public dataset including 349 COVID-19 photos and 397 images of cases with no discoveries was used to confirm the accuracy of the suggested model. As a result, the MKs-ELM-DNN model had an accuracy of 0.9836.

Similarly, Murugan et al. [197] has looked on a deep convolutional network-based ELM-based classifier for COVID-19 diagnosis. To autonomously diagnose COVID-19, the researchers employed an X-ray imaging dataset including COVID-19, bacterial pneumonia, and normal condition. On publicly accessible normal, pneumonia, and COVID-19 datasets (Chest Imaging 2020 (IT2020); SIRM COVID-19 2020 database), the suggested CNN approach has been trained and tested. Table 9 displays the ResNet CNN model's findings.

Afza et al. [198] introduce a new technique for classifying multiclass skin lesions that combines the best deep learning characteristics and an ELM. The methodology consists of feature selection using a hybrid whale optimization and entropy mutual information (EMI) approach, image acquisition and contrast enhancement, deep learning feature extraction using transfer learning, fusion of selected features using a modified canonical correlation-based approach, and ELM based classification. HAM10000 and ISIC2018, two publicly accessible datasets, were used in the experiment. On both datasets, accuracy was attained

Table 9
Background on medical image processing using SSL-ELM.

Organ	Reference	Year	Algorithm	Dataset	Modality	Results (Sn/Sp/Acc)
Brain	Pashaei et al. [186]	2018	KELM	3064 images	MR	–/–/0.9368
Lung	Wang et al. [185]	2018	SS-ELM	LIDC-IDRI	CT	–/–/0.9591
	Turkoglu [196]	2021	MKs-ELM-DNN	746 images	CT	–/–/0.9836
	Huang et al. [148]	2020	DTCNN/ ELM	LIDC-IDRI/ FAH-GMU	CT	0.9369/0.9515/0.9457
	Wu et al. [117]	2021	ELM	COVID-Xray-5k	X-ray	–/–/0.9883
	Murugan et al. [197]	2021	ELM	IT2020/ SIRM COVID-19 2020	X-ray	0.9815/0.9148/0.9407
	Vijendran et al. [190]	2019	OS-ELM	MNIST/Chest X-ray	X-ray	–/–/0.9095
Breast	Wang et al. [193]	2020	D^2TL/ ICELM	134 images	MG	0.9696/0.9818/0.9667
	Saxena et al. [194]	2021	KP-ELM	BreakHis/ BisQue	HP	–/–/0.9002
	Ronoud et al. [189]	2019	DBN-ELM	WBCO/WDBC	MG	1/0.9917/0.9945
	Sannasi et al. [195]	2021	SC-CSOA-ELM	INbreast	MG	–/–/0.9581
	Fei et al. [191]	2019	ML-KELM-PM	DTN	US	0.8526/0.8741/0.8636
Skin	Afza et al. [198]	2022	ELM	HAM10000/ ISIC2018	DS	–/–/0.9436
Eye	Zhou et al. [192]	2020	LRR-SSELM	DIARETDB0/ DIARETED1/ Messidor	OS	–/–/0.9740
Cervix	Ghoneim et al. [144]	2020	CNN/ELM	Herlev	OCC	–/–/0.995
Others	She et al. [187]	2018	Safe-SSELM	MI EEG	EEG	–/–/0.6812
	She et al. [188]	2019	HSS-ELM	MI EEG	EEG	-/-/0.6776

Table 10
Advantages and disadvantages of learning algorithms in a medical imaging context.

Learning algorithms	Advantages	Disadvantages
Supervised	This type of learning contains most of the articles reviewed, and this is because the databases used have all the tags. Addresses classification and regression problems. It is a type of machine learning in which the inputs and outputs are clearly identified.	It is difficult to collect images of any organ that are labeled by physicians. It does not consider grouping problems.
Unsupervised	It is an inductive algorithm and is based on the data clustering process. It can find previously unknown patterns in the data, which is impossible with supervised learning models.	It must handle the high complexity of unlabeled databases. Difficult to measure performance due to a lack of predefined responses during training. This is the section with the fewest articles.
Semi supervised	It has training efficiency and a simple implementation for multi-class classification problems It is considered an appropriate approach in the absence of labeled data and uses, simultaneously uses, this type of data and unlabeled data in the training stage. Labeled data can contribute significantly to accurate pattern extraction. It exhibits proper label function prediction with high accuracy. It can be used for speech recognition, data mining, video surveillance, and prediction. It is ideal for medical databases, since this type of data is usually not fully labeled.	It can be complicated to design and apply both the model and the algorithm suitable for the specific problem. Human effort to label the data.

at 0.9340 and 0.9436. It has been demonstrated that the suggested strategy is computationally effective.

Table 9 summarizes each of the antecedents presented above, highlighting the organ, reference, year, algorithm, dataset, modality and results.

The main advantages and disadvantages of each of the types of learning applied to medical imaging, presented in the previous sections, are shown in Table 10.

7. Discussion and future directions

7.1. Discussion

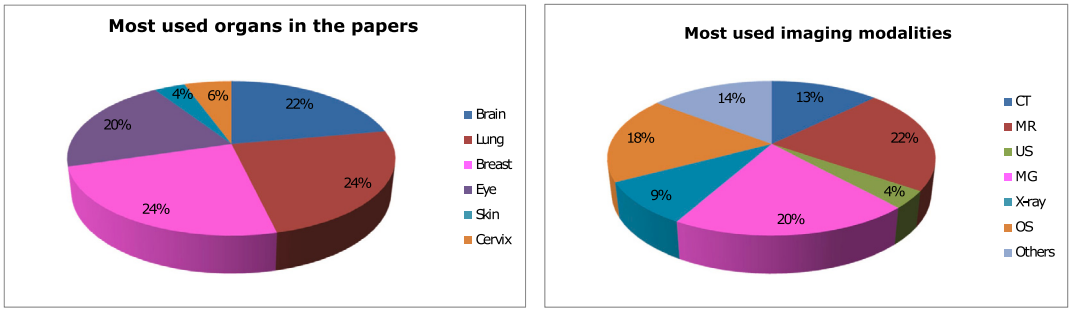
This section presents a discussion of the findings identified from this review.

Within the total number of articles reviewed in Sections 4, 5 and 6 it could be seen that the most studied organs, with the same number of articles, were lungs and breasts [see Fig. 14(a)], followed by eye, brain, cervix, and skin. This indicates that diseases of the lungs and breast are considered the most important. The most addressed diseases within the lungs were lung nodules and COVID-19, and all of the articles linked to the breasts only addressed breast cancer. Additionally, the pie chart in Fig. 14(b) identified that the most used imaging modalities in the aforementioned sections were MR followed by MG.

One thing to keep in mind is that the MR modality produces pictures with high spatial resolution, excellent soft tissue discrimination, and rich anatomical structural information. This imaging technique is more advantageous than other modalities and is a standard for brain studies, whereas MG is also the standard modality for breast cancer screening.

An exhaustive analysis of the articles reviewed in Sections 4, 5 and 6 showed that, to improve the accuracy in the detection and classification of different pathologies in MI, variants of ELMs have been developed and have reported excellent results, comparing both classical ELM and machine learning algorithms [149,157,159,172,173,182,186,190,194,199].

Currently, ELM paradigms have been strongly positioning themselves as an excellent option for the detection of diseases in medical images. However, it should not be ignored that other operators have done similar work in the context of machine learning, such as SVMs, KNN, CNN, PNN, DNN, and MLP. However, ELMs have shown better results in measuring classification accuracy in medical images [96,103,143,151,156,170,171,183,185].



(a) (b)

Fig. 14. Percentage representation of (a) Most reported organs and (b) Most reported imaging modalities.

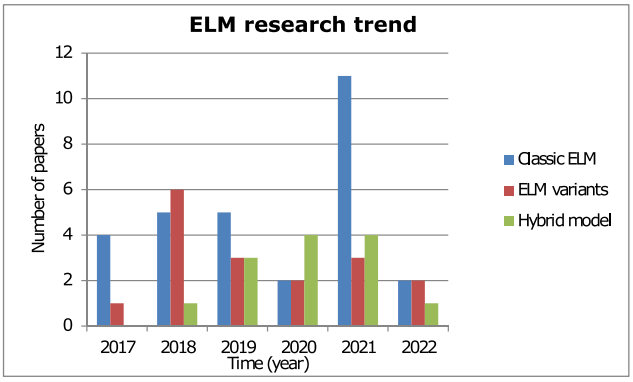


Fig. 15. Research trend among basic ELMs, ELM variants, and hybrid models.

In the last two years, medical-technological research has emerged in which a hybrid process that combines networks developed in the context of deep learning, specifically CNNs with ELMs, has reported good performance in the disease classification process [103,116,144,148].

Importantly, algorithms developed with transfer learning seem very useful in the medical image processing scenario for disease detection and/or classification [144,148,189,193]. However, there are difficulties in acquiring large-scale datasets in the medical area. This is because the available annotated datasets in medical imaging are small, so researchers have used pre-trained networks with non-medical datasets. This implies a higher probability of misclassification by the predictive model.

On the other hand, in a large percentage of the articles reviewed, the most reported metrics for evaluating the performance of the proposed classification models were sensitivity, specificity, and accuracy.

Another important point to highlight is that most of the articles analyzed were concerned with medical image processing using supervised ELMs. However, it is difficult to collect images of any organ that is labeled by clinicians.

In most cases, there are a significant number of unlabeled medical images and although they represent an important source of information, these images cannot be used for supervised learning. This is why Section 5 presents fewer reviewed articles, due to the complexity presented with unlabeled databases.

By analyzing the articles presented in Sections Section 4, 5 and 6 it was determined that the trend in ELM research is given by the introduction of modifications in the basic ELM algorithm for proposed hybrid ELM models, to improve the performance of the original ELM, as seen in Fig. 15.

On the other hand, many of the studies analyzed used different data sets obtained privately by clinics or universities. This produces disadvantages because there is no access to the data to replicate the study.

A review of the literature indicates that ELMs have been used for several purposes, including the use of attributes for the detection of diseases linked to brain tumors [160], melanoma [180], type 2 diabetes [200], diagnosis of liver cancer [201], breast cancer [162], and lung nodules [182], among others.

In some cases, ELM has been found to perform better than other conventional learning algorithms, especially in those databases contaminated with high levels of noise. Moreover, nowadays, the hybrid CNN-ELM model is trending, since CNN behaves as an excellent attribute extractor and selector, while ELM efficiently classifies such attributes [103,116,144,201,202].

However, it is important to realize that the performance of machine learning algorithms may depend on several factors, such as the specific dataset and associated features. Therefore, it is difficult to make a general statement on whether ELM performs better than other algorithms for specific human organs or datasets. This reveals that more research is required to rigorously evaluate the performance of the different intelligent algorithms available on a case-by-case basis and, thus, have sufficient empirical evidence to identify which of them is the most suitable for specific tasks.

Despite the advantages offered by ELMs, there are some gaps available for future research, such as improving the performance of ELM models using prior knowledge. For this, the following strategies can be considered:

- Incorporating prior knowledge into the model: prior knowledge can be used to inform the model about the problem domain, which can help improve its accuracy. For example, prior knowledge about the texture of the midbrain can be used to train a single or multilayer ELM model that is able to localize the substantia nigra, which is linked to Parkinson's disease. Another example is the use of prior knowledge in ELMs to address disease counting, the fundamental conclusion of which is that classical ELMs exhibit worse performance than ELMs with prior spatial knowledge (Spatial-ELM), especially in those cases in which information loss by omission or human procedural error is present (databases with missing information [203]).
- Feature selection: prior knowledge can be used to select relevant features for the model. This can help reduce the dimensionality of the input data and improve model performance. For example, the use of iterative schemes based on discarding attributes that contribute only low percentages in pattern detection will allow selecting features relevant to the model in an appropriate way. In this context, regularization by sparsity-introducing rules, such as L_1 and/or joint rules, have proven to be very useful and, at present, their use is widespread and constitutes a very active field of current research [96,204–207].
- Regularization by considering mathematical join norms: These mathematical norms have exhibited good performance on noisy data and hyperparameter tuning, and they can improve the performance of ELMs by selecting optimal hyperparameters that are

specific to the problem domain. Besides, join-norms are useful in hyperparameter tuning, which it is a key process in the performance of ELMs in specific problems. For example, in disease detection in medical images, prior knowledge about the type of noise in the images to be processed can be used to make it easier to train a monolayer or multilayer ELM model.

Additionally, the hyperparameter tuning process can be linked with ensemble assembly-based methods [33] to generate intelligent multimodality extreme learning machine models, e.g., ELMs that can simultaneously analyze images of any modality to detect diseases affecting human health. In this way, join-norms and prior knowledge can be used to create a set of ELM models that are trained on different subsets of data or with different hyperparameters, giving relevance to ensemble assembly methods in real-time learning tasks for the classification, clustering and regression of medical images.

On the other hand, unbalanced databases have the disadvantage, if not properly processed, of inducing serious problems in the generalizability of intelligent machine learning models, those based on ELMs in particular. Existing approaches to unbalanced learning only consider the effect of the number of class samples, ignoring the degree of sparsity of the data, and can lead to suboptimal learning results [208].

In that sense, the literature reports several studies that have the purpose of balancing samples in this type of database, linked to pattern classification, considering ELMs and variants as an extreme learning machine of class-specific cost regulation [208].

1. Train an adaptive ELM-multilayer model combined with dynamic generative antagonistic networks (GANs). This approach can produce synthetic samples to balance the dataset and improve classification performance [209].
2. Use active learning combined with ELM for online unbalanced multiclass classification, avoiding preliminary labeling of most of the data and allowing effective selection of informative samples for labeling, thus elevating the classification process of the ELM model [210].
3. Consider the strategy of a cross-weighted ELM for unbalanced classification, to transfer information from the source domain to the target domain and improve the classification performance of the ELM model, considering public domain databases [211].
4. Develop a hybrid ELM based on the Salp Swarm algorithm (ELM-SSA) to classify brain magnetic resonance images as healthy or diseased [212]. On the other hand, it is interesting to consider studies with a physics-based deep learning framework for the modeling of some diseases using medical images, as was the case in [213].

It can be seen that only the last reference presents the combination with an approach to address the problem of biased data towards the use of medical imaging. As a consequence, it is expected that research will be developed to apply the approaches listed in points 1 to 3 but in the context of medical imaging.

7.2. Future directions

Based on the findings achieved and the extensive state-of-the-art, future research trends can be synthesized in the following propositions:

It is true that the efficiency of a single hidden layer ELM model is based on the random initialization of the internal network weights during the learning process. However, it is important to note that the performance of the network may be affected by a single hidden layer, which may lead to inefficient results.

The idea of incorporating ML-ELM algorithms in the medical area is interesting because these algorithms have the ability to separate nonlinear data in feature space. This implies that they can extract complex features and more meaningful representations from medical data, which improves the performance of the single-layer ELM method. In a medical context, where data sets are often complex and nonlinear, the use of ML-ELM algorithms can be an excellent alternative.

These algorithms can help identify hidden patterns and complex relationships in medical data, which can lead to better understanding and prediction of diseases, more accurate diagnoses, and more effective treatment recommendations. For example, [214] proposed an ML-ELM variant, Deep ELM (DELM), which combines the best of both approaches, i.e., ML-ELM and DELM. This approach proved successful for the classification of patterns present in EEG, as it requires less training time and high efficiency. A variant of ML-ELM, PGM-ELM [209], has been efficiently used to classify imbalanced medical data.

Feature extraction is a very important process because it allows the network to perform better; knowing which techniques to use is very valuable. In the context of medical imaging, ELM auto-encoders are used to extract relevant features or attributes from images [43,44]. These models are designed to learn latent representations of images by compressing and reconstructing the original information. Considering these models for strengthening and improving deep and randomized algorithms could possibly allow for improvements in accuracy in certain applications, specifically in the medical field.

On the other hand, it would be interesting to further exploit the capabilities of the hybrid CNN-ELM model, in order to improve the training speed of the supervised CNN output classifier, while maintaining the accuracy of traditional schemes. In the literature, works in the medical area with this hybrid model are reported, which show an improvement in the classification performance of some pathologies and, additionally, effectively reduce computational costs [144,148].

The international scientific community linked to ELMs expects that, in the short and medium term, intelligent models based on the MI-ELM binomial will continue to exploit the advantages of explanatory machine learning and incorporate the foundations of federated learning and meta-learning to analyze large amounts of data, identify patterns efficiently and improve the current perception that physicians have about the suitability of the solutions generated by ELM models.

Federated learning is a technique that trains an algorithm using a distributed architecture consisting of multiple devices that contain their own local and private data [215] and, therefore, the combination of ELM and federated learning, in the context of medical imaging, may be an important trend in the future, as it will address the problem of privacy of the information contained in the data derived from these types of images.

In addition, meta-learning will probably make it possible to improve the performance of ELMs by refining learning mechanisms that are based on answering the following questions: how does one learn and what strategies should be applied to progressively improve learning? These are the two main questions in this type of learning.

Perhaps the aspect that will continue to strengthen the MI-ELM approach is the explanatory learning implicit in ELM, since in scenarios as complex and delicate as that of the health sector, it is necessary to have the greatest number of arguments, with a mathematical foundation, that explain any result produced by ELM. This is required in order to generate, in the medical and general community, the necessary confidence to set the standard as a consolidated learning paradigm in diagnosis considering medical imaging.

It is imperative to remember that learning paradigms mostly linked to deep learning cannot explain, in detail, the processes that produce their final results, due to the lack of explicit and accessible mathematical modeling.

Finally, another possible direction for future work could be the realization of a comparative study considering random networks, reported in the literature, in the context of disease detection that affects human health. In this way, empirical evidence could be obtained in order to demonstrate which of these networks exhibits the best performance when processing the information contained in medical images of any modality.

8. Conclusions and challenges

This work is intended to serve as a resource for researchers in pathology detection, considering medical imaging and random neural networks, specifically extreme learning machines and other randomly weighted neural networks such as RVFL. The authors expect that these networks will continue to advance in this area of medicine as a resource to support physicians and will be used more widely to assess the health of patients in the near future.

This review article focused on medical image processing using extreme learning machines. The theoretical–mathematical foundations of supervised, unsupervised, and semi-supervised learning, based on ELMs, were provided to highlight their high speed, good generalization, and easy implementation. Presenting these types of learning based on ELMs fills an information gap, in the sense that, without loss of formality, presenting direct, precise, clear, and concise information allows researchers to quickly find the fundamental characteristics, 'pros', and 'cons' of applying one or another type of learning.

In addition, the training, validation, and testing processes were analyzed in depth, using diagrams, which facilitate the researcher's understanding of the various ELM models considered in this work. Also, the most commonly used medical imaging modalities in the literature were described, and the advantages and disadvantages of these modalities were identified from a medical and computational point of view. To complement the information at this point, the public reference databases for brain, lung, breast, eye, skin, and cervical organs were described.

Besides, various research publications were collected, studied, and analyzed, which were classified into the different types of learning in the time interval from 2017 to 2022. In this case, it can be seen that ELMs have the advantages of fewer training parameters, fast learning speed, and high generalization ability. On the other hand, it has been identified that the majority of publications are presented in the supervised learning type, and a lesser number of publications are presented as unsupervised learning.

One of the reasons for this phenomenon is the fact that unlabeled databases are not often used; however, this work shows that it is possible to use unlabeled databases under the unsupervised learning approach. Despite the encouraging findings from the reviewed literature, there are still certain restrictions and difficulties with using ELMs and medical imagery to identify and categorize various illnesses. Several significant difficulties were found and they are listed below:

- How do we prove that the hybrid CNN Multi-resolution ELM approach can outperform the machine learning approaches to pathology detection reported so far in the literature? The totality of articles that handle the hybrid CNN-ELM approach do not consider multiscale versions for CNNs, thus missing the appropriate use of local attributes that could contribute to increasing the accuracy of the diagnoses. In other words, it is necessary to develop machine learning techniques that modify the size of the receptive fields in an iterative manner, simultaneously capturing local and global attributes present in the data. This statement is based on the fact that reviews of the multi-scale approach by [216,217] have not yet been proposed to address the issue analyzed in this paper.
- Will it be possible to increase the generalization capacity of ELMs which use medical imaging for disease detection, by developing strategies based on the identification, extraction, and selection of attributes from information obtained exclusively from different imaging modalities?
  In high correlation with the multi-resolution approach, raised in the previous paragraph, the process of feature extraction and selection linked to each imaging modality should be studied in depth. This is an interesting task to develop since these are processes that are considered to be a key part of the detection and identification of benign and malignant lesions. This is because the aforementioned processes are directly involved with the performance and generalization capacity of intelligent techniques in pattern classification and regularity detection present in the databases.
- Will it be possible to develop computational strategies for pathology monitoring using medical imaging and ELM, by replacing the technique established by Moore–Penrose for the calculation of the pseudo-inverse?
  Due to the fundamental role of pseudo-inverse matrices in the development of applications for disease detection using ELM, it is important to explore the evolution of the concept of these types of matrices in depth and evaluate the possibility of considering other methodologies for the calculation of pseudo-inverse matrices that allow a comparative study of the performance of ELMs concerning Moore–Penrose pseudo-inverse.
- Will it be possible to raise the quality of disease prognosis by considering medical imaging and pre-trained machine learning techniques, exclusively with big data taken from a strictly medical context, that serve for the implementation of semi-supervised ELMs?
  Considering the utility of algorithms with transfer learning in medical image processing, it is necessary to perform the training or evaluation process only on large-scale medical datasets.
  The issue, specifically in relation to this topic, is that it is challenging to create significant amounts of medical imaging data since annotating the data takes a lot of time and effort, not just from one clinical expert but from a number of them, in order to reduce human error. Additionally, the development of large-scale (big data) databases for medical imaging would be necessary for this initiative to be genuinely beneficial.
- It is a known fact that ELMs based on the concept of certain mathematical joint-norms, have successfully addressed problematic situations involving databases containing noisy non-medical information. Given this scenario, the following challenge arises: is it possible to successfully extend the application of strategies based on families of joint-norms to the ELM and Medical Imaging for the processing of this type of noisy database?
  To minimize the influence of noise and outliers associated with images from imaging modalities used in medicine, particularly the various types of noise, the information gap that exists regarding the use of joint standards could be explored to create strategies for disease detection and classification using MI and ELMs.
  In this sense, it is convenient to highlight that tackling the problem of noise in medical images by developing ELM-based algorithms could be feasible. This is because research developed with noisy databases in non-medical contexts have demonstrated the ability of joint-norms-based ELMs to perform adequately when processing noise-contaminated information [96,218,219].

CRediT authorship contribution statement

Yoleidy Huérfano-Maldonado: Methodology, Research, Writing – original draft. **Marco Mora:** Conceptualization, Methodology, Review and editing. **Karina Vilches:** Methodology, Review and editing. **Ruber Hernández-García:** Methodology, Review and editing. **Rodrigo Gutiérrez:** Review and editing. **Miguel Vera:** Methodology, Review and editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgments

This work was funded by the National Agency for Research and Development (ANID) / Scholarship Program/DOCTORADO BECAS CHILE/2021 - 21211760. The authors of the paper also thank to Research Project FONDECYT REGULAR 2020 N° 1200810 "Very Large Fingerprint Classification based on a Fast and Distributed Extreme Learning Machine", Ministerio de Ciencia, Tecnología, Conocimiento e Innovación, Gobierno de Chile.

References

[1] W. Cao, X. Wang, Z. Ming, J. Gao, A review on neural networks with random weights, Neurocomputing 275 (2018) 278–287.
[2] L. Zhang, P.N. Suganthan, A survey of randomized algorithms for training neural networks, Inform. Sci. 364 (2016) 146–155.
[3] P.N. Suganthan, R. Katuwal, On the origins of randomization-based feedforward neural networks, Appl. Soft Comput. 105 (2021) 107239.
[4] Q. Jiang, L. Zhu, C. Shu, V. Sekar, An efficient multilayer RBF neural network and its application to regression problems, Neural Comput. Appl. (2022) 1–18.
[5] J. Zhuo, C. An, J. Fei, Fuzzy multiple hidden layer neural sliding mode control of active power filter with multiple feedback loop, IEEE Access 9 (2021) 114294–114307.
[6] Z. Li, F. Liu, W. Yang, S. Peng, J. Zhou, A survey of convolutional neural networks: Analysis, applications, and prospects, IEEE Trans. Neural Netw. Learn. Syst. (2021).
[7] A.L. Caterini, D.E. Chang, A.L. Caterini, D.E. Chang, Recurrent neural networks, Deep Neural Netw. Math. Framew. (2018) 59–79.
[8] A. Liguori, R. Markovic, T.T.H. Dam, J. Frisch, C. van Treeck, F. Causone, Indoor environment data time-series reconstruction using autoencoder neural networks, Build. Environ. 191 (2021) 107623.
[9] Y.-H. Pao, Y. Takefuji, Functional-link net computing: Theory, system architecture, and functionalities, Computer 25 (5) (1992) 76–79.
[10] L. Zhang, P.N. Suganthan, A comprehensive evaluation of random vector functional link networks, Inform. Sci. 367 (2016) 1094–1105.
[11] N. Vuković, M. Petrović, Z. Miljković, A comprehensive experimental evaluation of orthogonal polynomial expanded random vector functional link neural networks for regression, Appl. Soft Comput. 70 (2018) 1083–1096.
[12] W.F. Schmidt, M.A. Kraaijveld, R.P. Duin, et al., Feed forward neural networks with random weights, in: International Conference on Pattern Recognition, IEEE Computer Society Press, 1992, p. 1.
[13] G.-B. Huang, Q.-Y. Zhu, C.-K. Siew, Extreme learning machine: A new learning scheme of feedforward neural networks, in: 2004 IEEE International Joint Conference on Neural Networks (IEEE Cat. No. 04CH37541), Vol. 2, Ieee, 2004, pp. 985–990.
[14] G.-B. Huang, Q.-Y. Zhu, C.-K. Siew, Extreme learning machine: Theory and applications, Neurocomputing 70 (1–3) (2006) 489–501.
[15] G. Huang, S. Song, J.N. Gupta, C. Wu, Semi-supervised and unsupervised extreme learning machines, IEEE Trans. Cybern. 44 (12) (2014) 2405–2417.
[16] J. Chen, Y. Zeng, Y. Li, G.-B. Huang, Unsupervised feature selection based extreme learning machine for clustering, Neurocomputing 386 (2020) 198–207.
[17] S. Ding, W. Jia, C. Su, L. Zhang, L. Liu, Research of neural network algorithm based on factor analysis and cluster analysis, Neural Comput. Appl. 20 (2) (2011) 297–302.
[18] Q. Shi, P.N. Suganthan, J. Del Ser, Jointly optimized ensemble deep random vector functional link network for semi-supervised classification, Eng. Appl. Artif. Intell. 115 (2022) 105214.
[19] H. Ye, F. Cao, D. Wang, A hybrid regularization approach for random vector functional-link networks, Expert Syst. Appl. 140 (2020) 112912.
[20] M. Hu, P.N. Suganthan, Representation learning using deep random vector functional link networks for clustering, Pattern Recognit. 129 (2022) 108744.
[21] S. Scardapane, D. Wang, Randomness in neural networks: an overview, Wiley Interdisc. Rev.: Data Min. Knowl. Discov. 7 (2) (2017) e1200.
[22] G.-B. Huang, L. Chen, C.K. Siew, et al., Universal approximation using incremental constructive feedforward networks with random hidden nodes, IEEE Trans. Neural Netw. 17 (4) (2006) 879–892.
[23] H. Zhong, C. Miao, Z. Shen, Y. Feng, Comparing the learning effectiveness of BP, ELM, I-ELM, and SVM for corporate credit ratings, Neurocomputing 128 (2014) 285–295.
[24] L. Zhang, D. Zhang, F. Tian, SVM and ELM: Who wins? Object recognition with deep convolutional features from ImageNet, in: Proceedings of ELM-2015 Volume 1: Theory, Algorithms and Applications (I), Springer, 2016, pp. 249–263.
[25] S.O. Olatunji, Extreme learning machines and support vector machines models for email spam detection, in: 2017 IEEE 30th Canadian Conference on Electrical and Computer Engineering, CCECE, IEEE, 2017, pp. 1–6.
[26] I.G. Maglogiannis, E.P. Zafiropoulos, Characterization of digital medical images utilizing support vector machines, BMC Med. Inform. Decision Making 4 (1) (2004) 1–9.
[27] K. Akyol, Comparing of deep neural networks and extreme learning machines based on growing and pruning approach, Expert Syst. Appl. 140 (2020) 112875.
[28] M. Jain, W. Andreopoulos, M. Stamp, Convolutional neural networks and extreme learning machines for malware classification, J. Comput. Virol. Hacking Tech. 16 (2020) 229–244.
[29] K. Kamnitsas, C. Ledig, V.F. Newcombe, J.P. Simpson, A.D. Kane, D.K. Menon, D. Rueckert, B. Glocker, Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation, Med. Image Anal. 36 (2017) 61–78.
[30] L. Alzubaidi, J. Zhang, A.J. Humaidi, A. Al-Dujaili, Y. Duan, O. Al-Shamma, J. Santamaría, M.A. Fadhel, M. Al-Amidie, L. Farhan, Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions, J. big Data 8 (2021) 1–74.
[31] M.A.A. Albadra, S. Tiuna, Extreme learning machine: A review, Int. J. Appl. Eng. Res. 12 (14) (2017) 4610–4623.
[32] B. Deng, X. Zhang, W. Gong, D. Shang, An overview of extreme learning machine, in: 2019 4th International Conference on Control, Robotics and Cybernetics, CRC, IEEE, 2019, pp. 189–195.
[33] J. Wang, S. Lu, S.-H. Wang, Y.-D. Zhang, A review on extreme learning machine, Multimedia Tools Appl. 81 (29) (2022) 41611–41660.
[34] G. Huang, G.-B. Huang, S. Song, K. You, Trends in extreme learning machines: A review, Neural Netw. 61 (2015) 32–48.
[35] O.A. Alade, A. Selamat, R. Sallehuddin, A review of advances in extreme learning machine techniques and its applications, in: International Conference of Reliable Information and Communication Technology, Springer, 2017, pp. 885–895.
[36] R. Nilesh, W. Sunil, Improving extreme learning machine through optimization a review, in: 2021 7th International Conference on Advanced Computing and Communication Systems, Vol. 1, ICACCS, IEEE, 2021, pp. 906–912.
[37] S. Ding, X. Xu, R. Nie, Extreme learning machine and its applications, Neural Comput. Appl. 25 (3) (2014) 549–556.
[38] S. Ghosh, H. Mukherjee, S.M. Obaidullah, K. Santosh, N. Das, K. Roy, A survey on extreme learning machine and evolution of its variants, in: Recent Trends in Image Processing and Pattern Recognition: Second International Conference, RTIP2R 2018, Solapur, India, December 21–22, 2018, Revised Selected Papers, Part I 2, Springer, 2019, pp. 572–583.
[39] P.V. De Campos Souza, L.C. Bambirra Torres, G.R. Lacerda Silva, A.d.P. Braga, E. Lughofer, An advanced pruning method in the architecture of extreme learning machines using l1-regularization and bootstrapping, Electronics 9 (5) (2020) 811.
[40] Z. Tian, G. Wang, S. Li, Y. Wang, X. Wang, Artificial bee colony algorithm–optimized error minimized extreme learning machine and its application in short-term wind speed prediction, Wind Eng. 43 (3) (2019) 263–276.
[41] S. Zang, X. Li, J. Ma, Y. Yan, J. Gao, Y. Wei, TSTELM: Two-stage transfer extreme learning machine for unsupervised domain adaptation, Comput. Intell. Neurosci. 2022 (2022).
[42] J. Zhang, Y. Li, W. Xiao, Adaptive online sequential extreme learning machine for dynamic modeling, Soft Comput. 25 (2021) 2177–2189.
[43] J. Zhang, Y. Li, W. Xiao, Z. Zhang, Non-iterative and fast deep learning: Multilayer extreme learning machines, J. Franklin Inst. B 357 (13) (2020) 8925–8955.
[44] J.A. Vásquez-Coronel, M. Mora, K. Vilches, A review of multilayer extreme learning machine neural networks, Artif. Intell. Rev. (2023) 1–52.
[45] R. Kaur, R.K. Roul, S. Batra, Multilayer extreme learning machine: A systematic review, Multimedia Tools Appl. (2023) 1–39.
[46] J. Tang, C. Deng, G.-B. Huang, Extreme learning machine for multilayer perceptron, IEEE Trans. Neural Netw. Learn. Syst. 27 (4) (2015) 809–821.
[47] R. Li, X. Wang, L. Lei, C. Wu, Representation learning by hierarchical ELM auto-encoder with double random hidden layers, IET Comput. Vis. 13 (4) (2019) 411–419.
[48] J. Wang, S. Lu, S.-H. Wang, Y.-D. Zhang, A review on extreme learning machine, Multimedia Tools Appl. (2021) 1–50.
[49] M. Eshtay, H. Faris, N. Obeid, Metaheuristic-based extreme learning machines: A review of design formulations and applications, Int. J. Mach. Learn. Cybern. 10 (6) (2019) 1543–1561.
[50] M.A. Azam, K.B. Khan, S. Salahuddin, E. Rehman, S.A. Khan, M.A. Khan, S. Kadry, A.H. Gandomi, A review on multimodal medical image fusion: Compendious analysis of medical modalities, multimodal databases, fusion techniques and quality metrics, Comput. Biol. Med. 144 (2022) 105253.
[51] Z. Wang, Y. Luo, J. Xin, H. Zhang, L. Qu, Z. Wang, Y. Yao, W. Zhu, X. Wang, Computer-aided diagnosis based on extreme learning machine: A review, IEEE Access 8 (2020) 141657–141673.
[52] G.-B. Huang, Q.-Y. Zhu, C.-K. Siew, Extreme learning machine: Theory and applications, Neurocomputing 70 (1–3) (2006) 489–501, http://dx.doi.org/10.1016/j.neucom.2005.12.126.

[53] G.-B. Huang, L. Chen, Convex incremental extreme learning machine, Neurocomputing 70 (16–18) (2007) 3056–3062.

[54] G.-B. Huang, L. Chen, Enhanced random search based incremental extreme learning machine, Neurocomputing 71 (16–18) (2008) 3460–3468.

[55] G.-B. Huang, H. Zhou, X. Ding, R. Zhang, Extreme learning machine for regression and multiclass classification, IEEE Trans. Syst. Man Cybern. B 42 (2) (2011) 513–529.

[56] D. Wang, G.-B. Huang, Protein sequence classification using extreme learning machine, in: Proceedings. 2005 IEEE International Joint Conference on Neural Networks, 2005, Vol. 3, IEEE, 2005, pp. 1406–1411.

[57] N.-Y. Liang, P. Saratchandran, G.-B. Huang, N. Sundararajan, Classification of mental tasks from EEG signals using extreme learning machine, Int. J. Neural Syst. 16 (01) (2006) 29–38.

[58] H.-J. Rong, G.-B. Huang, Y.-S. Ong, Extreme learning machine for multi-categories classification applications, in: 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence), IEEE, 2008, pp. 1709–1713.

[59] G. Vani, R. Savitha, N. Sundararajan, Classification of abnormalities in digitized mammograms using extreme learning machine, in: 2010 11th International Conference on Control Automation Robotics & Vision, IEEE, 2010, pp. 2114–2117.

[60] S. Suresh, S. Saras-wathi, N. Sundararajan, Performance enhancement of extreme learning machine for multi-category sparse data classification problems, Eng. Appl. Artif. Intell. 23 (7) (2010) 1149–1157.

[61] M. Pal, A.E. Max-well, T.A. Warner, Kernel-based extreme learning machine for remote-sensing image classification, Remote Sens. Lett. 4 (9) (2013) 853–862.

[62] W. Zheng, X. Fu, Y. Ying, Spectroscopy-based food classification with extreme learning machine, Chemometr. Intell. Lab. Syst. 139 (2014) 42–47.

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

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

[65] J.M. Martínez-Martínez, P. Escandell-Montero, E. Soria-Olivas, J.D. Martín-Guerrero, R. Magdalena-Benedito, J. GóMez-Sanchis, Regularized extreme learning machine for regression problems, Neurocomputing 74 (17) (2011) 3716–3721.

[66] Y. Yuan, Y. Wang, F. Cao, Optimization approximation solution for regression problem based on extreme learning machine, Neurocomputing 74 (16) (2011) 2475–2482.

[67] Y. Yang, Y. Wang, X. Yuan, Bidirectional extreme learning machine for regression problem and its learning effectiveness, IEEE Trans. Neural Netw. Learn. Syst. 23 (9) (2012) 1498–1505.

[68] G. Feng, Z. Qian, X. Zhang, Evolutionary selection extreme learning machine optimization for regression, Soft Comput. 16 (9) (2012) 1485–1491.

[69] A.R. Lima, A.J. Cannon, W.W. Hsieh, Nonlinear regression in environmental sciences using extreme learning machines: A comparative evaluation, Environ. Model. Softw. 73 (2015) 175–188.

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

[71] G. Huang, T. Liu, Y. Yang, Z. Lin, S. Song, C. Wu, Discriminative clustering via extreme learning machine, Neural Netw. 70 (2015) 1–8.

[72] Q. Wang, Y. Dou, X. Liu, Q. Lv, S. Li, Multi-view clustering with extreme learning machine, Neurocomputing 214 (2016) 483–494.

[73] J. Zhang, W. Xiao, Y. Li, S. Zhang, Residual compensation extreme learning machine for regression, Neurocomputing 311 (2018) 126–136.

[74] Q. He, X. Jin, C. Du, F. Zhuang, Z. Shi, Clustering in extreme learning machine feature space, Neurocomputing 128 (2014) 88–95.

[75] C.K.L. Lekamalage, T. Liu, Y. Yang, Z. Lin, G.-B. Huang, Extreme learning machine for clustering, in: Proceedings of ELM-2014 Volume 1, Springer, 2015, pp. 435–444.

[76] K. Javed, R. Gouriveau, N. Zerhouni, A new multivariate approach for prognostics based on extreme learning machine and fuzzy clustering, IEEE Trans. Cybern. 45 (12) (2015) 2626–2639.

[77] J. Huang, Z.L. Yu, Z. Gu, A clustering method based on extreme learning machine, Neurocomputing 277 (2018) 108–119.

[78] T. Liu, C.K.L. Lekamalage, G.-B. Huang, Z. Lin, Extreme learning machine for joint embedding and clustering, Neurocomputing 277 (2018) 78–88.

[79] J. Chen, Y. Zeng, Y. Li, G.-B. Huang, Unsupervised feature selection based extreme learning machine for clustering, Neurocomputing 386 (2020) 198–207.

[80] I. Fredholm, Sur une classe d'équations fonctionnelles, Acta mathematica 27 (1903) 365–390.

[81] D. Hilbert, Grundzüge einer allgemeinen Theorie der linearen Integralrechnungen.(Zweite Mitteilung), Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse 1904 (1904) 213–260.

[82] E.H. Moore, On the reciprocal of the general algebraic matrix, Bull. Amer. Math. Soc. 26 (1920) 394–395.

[83] R. Penrose, A generalized inverse for matrices, in: Mathematical Proceedings of the Cambridge Philosophical Society, Vol. 51, no. 3, Cambridge University Press, 1955, pp. 406–413.

[84] M. Drazin, Pseudo-inverses in associative rings and semigroups, Am. Math. Monthly 65 (7) (1958) 506–514.

[85] K. Manjunatha Prasad, K. Mohana, Core–EP inverse, Linear Multilinear Algebra 62 (6) (2014) 792–802.

[86] G. Luo, K. Zuo, L. Zhou, Revisitation of the core inverse, Wuhan Univ. J. Nat. Sci. 20 (5) (2015) 381–385.

[87] K. Zuo, Y. Li, G. Luo, A new generalized inverse of matrices from core-EP decomposition, 2020, arXiv preprint arXiv:2007.02364.

[88] Y. Chen, K. Zuo, Z. Fu, New characterizations of the generalized Moore-Penrose inverse of matrices, AIMS Math. 7 (3) (2022) 4359–4375.

[89] S.B. Malik, N. Thome, On a new generalized inverse for matrices of an arbitrary index, Appl. Math. Comput. 226 (2014) 575–580.

[90] M. Mehdipour, A. Salemi, On a new generalized inverse of matrices, Linear Multilinear Algebra 66 (5) (2018) 1046–1053.

[91] S.B. Malik, L. Rueda, N. Thome, The class of m-EP and m-normal matrices, Linear Multilinear Algebra 64 (11) (2016) 2119–2132.

[92] C.R. Rao, S.K. Mitra, Further contributions to the theory of generalized inverse of matrices and its applications, Sankhyā: Indian J. Stat. Ser. A (1971) 289–300.

[93] P.A. Alaba, S.I. Popoola, L. Olatomiwa, M.B. Akanle, O.S. Ohunakin, E. Adetiba, O.D. Alex, A.A. Atayero, W.M.A.W. Daud, Towards a more efficient and cost-sensitive extreme learning machine: A state-of-the-art review of recent trend, Neurocomputing 350 (2019) 70–90.

[94] W.-Y. Deng, Q.-H. Zheng, S. Lian, L. Chen, X. Wang, Ordinal extreme learning machine, Neurocomputing 74 (1–3) (2010) 447–456.

[95] P. Bartlett, The sample complexity of pattern classification with neural networks: The size of the weights is more important than the size of the network, IEEE Trans. Inform. Theory 44 (2) (1998) 525–536, http://dx.doi.org/10.1109/18.661502.

[96] L.-R. Ren, Y.-L. Gao, J.-X. Liu, R. Zhu, X.-Z. Kong, L2, 1-extreme learning machine: An efficient robust classifier for tumor classification, Comput. Biol. Chem. 89 (2020) 107368.

[97] J. Cao, Z. Lin, G.-B. Huang, Self-adaptive evolutionary extreme learning machine, Neural Process. Lett. 36 (3) (2012) 285–305.

[98] M.-B. Li, G.-B. Huang, P. Saratchandran, N. Sundararajan, Fully complex extreme learning machine, Neurocomputing 68 (2005) 306–314.

[99] N.-Y. Liang, G.-B. Huang, P. Saratchandran, N. Sundararajan, A fast and accurate online sequential learning algorithm for feedforward networks, IEEE Trans. Neural Netw. 17 (6) (2006) 1411–1423.

[100] K. El Bouchefry, R.S. de Souza, Learning in big data: Introduction to machine learning, in: Knowledge Discovery in Big Data from Astronomy and Earth Observation, Elsevier, 2020, pp. 225–249.

[101] Y. Peng, W.-L. Zheng, B.-L. Lu, An unsupervised discriminative extreme learning machine and its applications to data clustering, Neurocomputing 174 (2016) 250–264.

[102] S.P. Das, S. Padhy, Unsupervised extreme learning machine and support vector regression hybrid model for predicting energy commodity futures index, Memet. Comput. 9 (4) (2017) 333–346.

[103] Z. Wang, Y. Huang, M. Li, H. Zhang, C. Li, J. Xin, W. Qian, Breast mass detection and diagnosis using fused features with density, J. X-ray Sci. Technol. 27 (2) (2019) 321–342.

[104] L.L.C. Kasun, H. Zhou, G.-B. Huang, C.M. Vong, Representational learning with ELMs for big data, Intell. Syst. 28 (2013) 31–34.

[105] K. Sun, J. Zhang, C. Zhang, J. Hu, Generalized extreme learning machine autoencoder and a new deep neural network, Neurocomputing 230 (2017) 374–381.

[106] H.H. Nuha, A. Balghonaim, B. Liu, M. Mohandes, M. Deriche, F. Fekri, Deep neural networks with extreme learning machine for seismic data compression, Arab. J. Sci. Eng. 45 (3) (2020) 1367–1377.

[107] S. Li, S. Song, Y. Wan, Laplacian twin extreme learning machine for semi-supervised classification, Neurocomputing 321 (2018) 17–27.

[108] X. Gong, T. Zhang, C.P. Chen, Z. Liu, Research review for broad learning system: Algorithms, theory, and applications, IEEE Trans. Cybern. (2021).

[109] P. Drew, J. Chatwin, S. Collins, Conversation analysis: A method for research into interactions between patients and health-care professionals, Health Expect. 4 (1) (2001) 58–70.

[110] E.H. Houssein, M.M. Emam, A.A. Ali, P.N. Suganthan, Deep and machine learning techniques for medical imaging-based breast cancer: A comprehensive review, Expert Syst. Appl. 167 (2021) 114161.

[111] D. Marin, R.C. Nelson, S.T. Schindera, S. Richard, R.S. Youngblood, T.T. Yoshizumi, E. Samei, Low-tube-voltage, high-tube-current multidetector abdominal CT: Improved image quality and decreased radiation dose with adaptive statistical iterative reconstruction algorithm—initial clinical experience, Radiology 254 (1) (2010) 145–153.

[112] J.-B. Thibault, K.D. Sauer, C.A. Bouman, J. Hsieh, A three-dimensional statistical approach to improved image quality for multislice helical CT, Med. Phys. 34 (11) (2007) 4526–4544.

[113] M.A. Haidekker, X-ray projection imaging, in: Medical Imaging Technology, Springer, 2013, pp. 13–35.

[114] H. Kasban, M. El-Bendary, D. Salama, A comparative study of medical imaging techniques, Int. J. Inform. Sci. Intell. Syst. 4 (2) (2015) 37–58.

[115] MedLinePlus, National library of medicine, 2022, https://medlineplus.gov/xrays.html (Accedido en junio de 2022).
[116] M. Nahiduzzaman, M.O.F. Goni, M.S. Anower, M.R. Islam, M. Ahsan, J. Haider, S. Gurusamy, R. Hassan, M.R. Islam, A novel method for multivariant pneumonia classification based on hybrid CNN-PCA based feature extraction using extreme learning machine with CXR images, IEEE Access 9 (2021) 147512–147526.
[117] C. Wu, M. Khishe, M. Mohammadi, S.H. Taher Karim, T.A. Rashid, Evolving deep convolutional neutral network by hybrid sine–cosine and extreme learning machine for real-time COVID19 diagnosis from X-ray images, Soft Comput. (2021) 1–20.
[118] J.P. Cohen, P. Morrison, L. Dao, K. Roth, T.Q. Duong, M. Ghassemi, COVID-19 image data collection: Prospective predictions are the future, 2020, arXiv: 2006.11988 https://github.com/ieee8023/covid-chestxray-dataset.
[119] W. Stiller, Basics of iterative reconstruction methods in computed tomography: A vendor-independent overview, Eur. J. Radiol. 109 (2018) 147–154.
[120] E.R. Llorente, E.A. Alvarez, conceptos básicos en la tomografía computarizada de tórax, Medicina respiratoria 1 (11) (2018) 23–35.
[121] K. Clark, B. Vendt, K. Smith, J. Freymann, J. Kirby, P. Koppel, S. Moore, S. Phillips, D. Maffitt, M. Pringle, et al., The cancer imaging archive (TCIA): Maintaining and operating a public information repository, J. Digit. Imaging 26 (6) (2013) 1045–1057.
[122] M. Vera, A. Bravo, R. Medina, Description and use of three-dimensional numerical phantoms of cardiac computed tomography images, Data 7 (8) (2022) 115.
[123] J. Gore, J. Orr, Image formation by back-projection: A reappraisal, Phys. Med. Biol. 24 (4) (1979) 793.
[124] R. Gordon, R. Bender, G.T. Herman, Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and X-ray photography, J. Theoret. Biol. 29 (3) (1970) 471–481.
[125] K. Lange, R. Carson, et al., EM reconstruction algorithms for emission and transmission tomography, J. Comput. Assist. Tomogr. 8 (2) (1984) 306–316.
[126] R.d.O. Pereira, L.A.d. Luz, D.C. Chagas, J.R. Amorim, E.d.J. Nery-Júnior, A.C.B.R. Alves, F.T.d. Abreu-Neto, M.d.C.B. Oliveira, D.R.C. Silva, J.M. Soares-Júnior, et al., Evaluation of the accuracy of mammography, ultrasound and magnetic resonance imaging in suspect breast lesions, Clinics 75 (2020).
[127] NIBIB, National institute of biomedical imaging and bioengineering, 2022, https://www.nibib.nih.gov/ (Accedido en septiembre de 2022).
[128] J. Suckling, J. Parker, D. Dance, S. Astley, I. Hutt, C. Boggis, I. Ricketts, E. Stamatakis, N. Cerneaz, S. Kok, et al., Mammographic image analysis society (mias) database v1. 21, 2015.
[129] M. Heath, K. Bowyer, D. Kopans, P. Kegelmeyer, R. Moore, K. Chang, S. Munishkumaran, Current status of the digital database for screening mammography, in: Digital Mammography, Springer, 1998, pp. 457–460.
[130] D.S. Gowri, T. Amudha, A review on mammogram image enhancement techniques for breast cancer detection, in: 2014 International Conference on Intelligent Computing Applications, IEEE, 2014, pp. 47–51.
[131] A. Elmoufidi, Pre-processing algorithms on digital X-ray mammograms, in: 2019 IEEE International Smart Cities Conference, ISC2, IEEE, 2019, pp. 87–92.
[132] C.N. Valverde, J.R. Maqueda, M.J.R. Reyes, I.E. Ruiz, D.G. Medina, R.P. Jiménez, C. Rodríguez-Gómez, J.L. del Ojo, A. Cayuela, F.J. Molano-Casimiro, Magnetic resonance imaging in patients with cardiac implantable electronic devices: A prospective study, Magn. Res. Imaging 91 (2022) 9–15.
[133] MITA, Medical imaging & technology alliance, 2022, https://www.medicalimaging.org/about-mita/medical-imaging-primer/ (Accedido en junio de 2022).
[134] E.A. Zerhouni, D.M. Parish, W.J. Rogers, A. Yang, E.P. Shapiro, Human heart: Tagging with MR imaging–a method for noninvasive assessment of myocardial motion, Radiology 169 (1) (1988) 59–63.
[135] T. Karamitsos, S. Neubauer, Cardiovascular magnetic resonance imaging, Medicine 50 (6) (2022) 372–378.
[136] A.P. James, B.V. Dasarathy, Medical image fusion: A survey of the state of the art, Inform. Fusion 19 (2014) 4–19.
[137] R.W. Cootney, Ultrasound imaging: Principles and applications in rodent research, Ilar J. 42 (3) (2001) 233–247.
[138] A. Carovac, F. Smajlovic, D. Junuzovic, Application of ultrasound in medicine, Acta Inform. Medica 19 (3) (2011) 168.
[139] M. Ali, D. Magee, U. Dasgupta, Signal processing overview of ultrasound systems for medical imaging, SPRAB12, Texas Instrum. Texas 55 (2008).
[140] N.M. Dipu, S.A. Shohan, K.A. Salam, Brain tumor detection using various deep learning algorithms, in: 2021 International Conference on Science & Contemporary Technologies, ICSCT, IEEE, 2021, pp. 1–6.
[141] B. Menze, A. Jakab, S. Bauer, M. Reyes, M. Prastawa, K. Van Leemput, Proceedings of the miccai challenge on multimodal brain tumor image segmentation (brats) 2012, in: MICCAI Challenge on Multimodal Brain Tumor Image Segmentation, BRATS, MICCAI, 2012, p. 77.
[142] D.R. Nayak, D. Das, R. Dash, S. Majhi, B. Majhi, Deep extreme learning machine with leaky rectified linear unit for multiclass classification of pathological brain images, Multimedia Tools Appl. 79 (21) (2020) 15381–15396.
[143] T. Sheela Shiney, R.J. Rose, Deep auto encoder based extreme learning system for automatic segmentation of cervical cells, IETE J. Res. (2021) 1–21.
[144] A. Ghoneim, G. Muhammad, M.S. Hossain, Cervical cancer classification using convolutional neural networks and extreme learning machines, Future Gener. Comput. Syst. 102 (2020) 643–649.
[145] D. Painuli, S. Bhardwaj, et al., Recent advancement in cancer diagnosis using machine learning and deep learning techniques: A comprehensive review, Comput. Biol. Med. (2022) 105580.
[146] I. Düntsch, G. Gediga, Indices for rough set approximation and the application to confusion matrices, Internat. J. Approx. Reason. 118 (2020) 155–172.
[147] P. Pi, D. Lima, Gray level co-occurrence matrix and extreme learning machine for Covid-19 diagnosis, Int. J. Cogn. Comput. Eng. 2 (2021) 93–103.
[148] X. Huang, Q. Lei, T. Xie, Y. Zhang, Z. Hu, Q. Zhou, Deep transfer convolutional neural network and extreme learning machine for lung nodule diagnosis on CT images, Knowl.-Based Syst. 204 (2020) 106230.
[149] R.K. Lama, J. Gwak, J.-S. Park, S.-W. Lee, Diagnosis of Alzheimer's disease based on structural MRI images using a regularized extreme learning machine and PCA features, J. Healthcare Eng. 2017 (2017).
[150] C. Zhu, B. Zou, R. Zhao, J. Cui, X. Duan, Z. Chen, Y. Liang, Retinal vessel segmentation in colour fundus images using extreme learning machine, Comput. Med. Imaging Graph. 55 (2017) 68–77.
[151] V. Kuppili, M. Biswas, A. Sreekumar, H.S. Suri, L. Saba, D.R. Edla, R.T. Marinhoe, J.M. Sanches, J.S. Suri, Extreme learning machine framework for risk stratification of fatty liver disease using ultrasound tissue characterization, J. Med. Syst. 41 (10) (2017) 1–20.
[152] I.H. Punithavathi, P.G. Kumar, Severity grading of diabetic retinopathy using extreme learning machine, in: 2017 IEEE International Conference on Intelligent Techniques in Control, Optimization and Signal Processing, INCOS, IEEE, 2017, pp. 1–6.
[153] Y. Luo, B. Yang, L. Xu, L. Hao, J. Liu, Y. Yao, F.v.d. Vosse, Segmentation of the left ventricle in cardiac MRI using a hierarchical extreme learning machine model, Int. J. Mach. Learn. Cybern. 9 (10) (2018) 1741–1751.
[154] A. Toprak, Extreme learning machine (elm)-based classification of benign and malignant cells in breast cancer, Med. Sci. Monitor: Int. Med. J. Exper. Clin. Res. 24 (2018) 6537.
[155] M. Eshtay, H. Faris, N. Obeid, Improving extreme learning machine by competitive swarm optimization and its application for medical diagnosis problems, Expert Syst. Appl. 104 (2018) 134–152.
[156] D.R. Nayak, R. Dash, B. Majhi, S. Wang, Combining extreme learning machine with modified sine cosine algorithm for detection of pathological brain, Comput. Electr. Eng. 68 (2018) 366–380.
[157] A. Gumaei, M.M. Hassan, M.R. Hassan, A. Alelaiwi, G. Fortino, A hybrid feature extraction method with regularized extreme learning machine for brain tumor classification, IEEE Access 7 (2019) 36266–36273.
[158] R.F. Rahmat, R.M. Wijaya, S. Faza, E.B. Nababan, F. Nadi, et al., Classification of primary and secondary brain tumor using extreme learning machine, in: 2021 International Conference on Data Science, Artificial Intelligence, and Business Analytics, DATABIA, IEEE, 2021, pp. 101–105.
[159] V. Sasank, S. Venkateswarlu, Brain tumor classification using modified kernel based softplus extreme learning machine, Multimedia Tools Appl. 80 (9) (2021) 13513–13534.
[160] B. Devanathan, M. Kamarasan, Automated brain tumor diagnosis using residual network with optimal kernel extreme learning machine, in: 2022 4th International Conference on Smart Systems and Inventive Technology, ICSSIT, IEEE, 2022, pp. 860–865.
[161] S. Dhivya, A. Nithya, T. Abirami, Mamogram image classification using extreme learning machine, Indian J. Sci. Technol. 11 (17) (2018) 1–4.
[162] V. Lahoura, H. Singh, A. Aggarwal, B. Sharma, M.A. Mohammed, R. Damaševičius, S. Kadry, K. Cengiz, Cloud computing-based framework for breast cancer diagnosis using extreme learning machine, Diagnostics 11 (2) (2021) 241.
[163] M.F. Aslan, M. Ceylan, A. Durdu, Segmentation of retinal blood vessel using gabor filter and extreme learning machines, in: 2018 International Conference on Artificial Intelligence and Data Processing, IDAP, IEEE, 2018, pp. 1–5.
[164] H. Huang, H. Ma, H.J. van Triest, Y. Wei, W. Qian, Automatic detection of neovascularization in retinal images using extreme learning machine, Neurocomputing 277 (2018) 218–227.
[165] B. Krishna, T. Gnanasekaran, Retinal vessel extraction framework using modified adaboost extreme learning machine, CMC-Comput. Mater. Continua 60 (3) (2019) 855–869.
[166] T. Nazir, A. Irtaza, Z. Shabbir, A. Javed, U. Akram, M.T. Mahmood, Diabetic retinopathy detection through novel tetragonal local octa patterns and extreme learning machines, Artif. Intell. Med. 99 (2019) 101695.
[167] N.J. Shyla, W.S. Emmanuel, Automated classification of glaucoma using DWT and HOG features with extreme learning machine, in: 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks, ICICV, IEEE, 2021, pp. 725–730.
[168] G.P. Pappu, B. Biswal, T.K. Gandhi, M.V.S. Sai Ram, Classification of neovascularization on retinal images using extreme learning machine, Int. J. Imaging Syst. Technol. 31 (3) (2021) 1536–1550.
[169] L.S. Sousa, P.P. Rebouças Filho, F.N. Bezerra, A.R.R. Neto, S.A. Oliveira, An improved retinal blood vessel detection system using an extreme learning machine, Int. J. E-Health Med. Commun. (IJEHMC) 10 (3) (2019) 39–55.

[170] J. George, et al., Extreme learning machine based classification for detecting micro-calcification in mammogram using multi scale features, in: 2019 International Conference on Computer Communication and Informatics, ICCCI, IEEE, 2019, pp. 1–7.
[171] M. Sharifmoghadam, H. Jazayeriy, Breast cancer classification using AdaBoost-extreme learning machine, in: 2019 5th Iranian Conference on Signal Processing and Intelligent Systems, ICSPIS, IEEE, 2019, pp. 1–5.
[172] F. Mohanty, S. Rup, B. Dash, Automated diagnosis of breast cancer using parameter optimized kernel extreme learning machine, Biomed. Signal Process. Control 62 (2020) 102108.
[173] S. Bacha, K. Ben Abdellafou, A. Aljuhani, O. Taouali, N. Liouane, Early detection of digital mammogram using kernel extreme learning machine, Concurr. Comput.: Pract. Exper. (2022) e6971.
[174] S. Tandungan, I. Nurtanio, et al., Comparison of accuracy in extreme learning machine based on hidden node structure variation for lung cancer classification, IOP Conf. Ser.: Mater. Sci. Eng. 676 (1) (2019) 012014.
[175] M. Sharif, J. Amin, M. Raza, M.A. Anjum, H. Afzal, S.A. Shad, Brain tumor detection based on extreme learning, Neural Comput. Appl. 32 (20) (2020) 15975–15987.
[176] A. Hari, H. Vanamala, N. Sanjana, R. Jain, et al., An extreme learning machine based approach to detect the Alzheimer's disease, in: 2021 IEEE 4th International Conference on Computing, Power and Communication Technologies, GUCON, IEEE, 2021, pp. 1–6.
[177] S. Govindarajan, R. Swaminathan, Extreme learning machine based differentiation of pulmonary tuberculosis in chest radiographs using integrated local feature descriptors, Comput. Methods Programs Biomed. 204 (2021) 106058.
[178] M.A. Khan, A. Majid, T. Akram, N. Hussain, Y. Nam, S. Kadry, S.-H. Wang, M. Alhaisoni, Classification of COVID-19 CT scans via extreme learning machine, CMC-Comput. Mater. Continua (2021).
[179] A.B. Suksmono, R. Rulaningtyas, K. Triyana, I.S. Sitanggang, A.S. Rahaju, E.H. Kusumastuti, A.N.L. Nabila, R.N. Maharani, D.F. Ismayanto, Katherine, et al., Classification of adeno carcinoma, high squamous intraephithelial lesion, and squamous cell Carcinoma in pap smear images based on extreme learning machine, Comput. Methods Biomech. Biomed. Eng.: Imaging Visual. 9 (2) (2021) 115–120.
[180] S. Al-Hammouri, M. Fora, M. Ibbini, Extreme learning machine for melanoma classification, in: 2021 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology, JEEIT, IEEE, 2021, pp. 114–119.
[181] N. Elsayed, Z. ElSayed, M. Ozer, Early stage diabetes prediction via extreme learning machine, in: SoutheastCon 2022, IEEE, 2022, pp. 374–379.
[182] R. Hao, Z. Qiang, Y. Qiang, L. Ge, J. Zhao, Automatic diagnosis of pulmonary nodules using a hierarchical extreme learning machine model, Int. J. Bio-Inspired Comput. 11 (3) (2018) 192–201.
[183] H. Lingappa, H. Suresh, S. Manvi, Medical image segmentation based on extreme learning machine algorithm in kernel fuzzy c-means using artificial bee colony method, Int. J. Intell. Eng. Syst 11 (2018) 128–136.
[184] N.J. Mohan, R. Murugan, T. Goel, P. Roy, Fast and robust exudate detection in retinal fundus images using extreme learning machine autoencoders and modified KAZE features, J. Digit. Imaging (2022) 1–18.
[185] Z. Wang, J. Xin, P. Sun, Z. Lin, Y. Yao, X. Gao, Improved lung nodule diagnosis accuracy using lung CT images with uncertain class, Comput. Methods Programs Biomed. 162 (2018) 197–209.
[186] A. Pashaei, H. Sajedi, N. Jazayeri, Brain tumor classification via convolutional neural network and extreme learning machines, in: 2018 8th International Conference on Computer and Knowledge Engineering, ICCKE, IEEE, 2018, pp. 314–319.
[187] Q. She, B. Hu, H. Gan, Y. Fan, T. Nguyen, T. Potter, Y. Zhang, Safe semi-supervised extreme learning machine for EEG signal classification, IEEE Access 6 (2018) 49399–49407.
[188] Q. She, B. Hu, Z. Luo, T. Nguyen, Y. Zhang, A hierarchical semi-supervised extreme learning machine method for EEG recognition, Med. Biol. Eng. Comput. 57 (1) (2019) 147–157.
[189] S. Ronoud, S. Asadi, An evolutionary deep belief network extreme learning-based for breast cancer diagnosis, Soft Comput. 23 (24) (2019) 13139–13159.
[190] S. Vijendran, R. Dubey, Deep online sequential extreme learning machines and its application in pneumonia detection, in: 2019 8th International Conference on Industrial Technology and Management, ICITM, IEEE, 2019, pp. 311–316.
[191] X. Fei, W. Zhou, L. Shen, C. Chang, S. Zhou, J. Shi, Ultrasound-based diagnosis of breast tumor with parameter transfer multilayer kernel extreme learning machine, in: 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC, IEEE, 2019, pp. 933–936.
[192] W. Zhou, S. Qiao, Y. Yi, N. Han, Y. Chen, G. Lei, Automatic optic disc detection using low-rank representation based semi-supervised extreme learning machine, Int. J. Mach. Learn. Cybern. 11 (1) (2020) 55–69.
[193] P. Wang, Q. Song, Y. Li, S. Lv, J. Wang, L. Li, H. Zhang, Cross-task extreme learning machine for breast cancer image classification with deep convolutional features, Biomed. Signal Process. Control 57 (2020) 101789.
[194] S. Saxena, S. Shukla, M. Gyanchandani, Breast cancer histopathology image classification using kernelized weighted extreme learning machine, Int. J. Imaging Syst. Technol. 31 (1) (2021) 168–179.
[195] S.C. SR, H. Rajaguru, Deep features with improved extreme learning machine for breast cancer classification, in: 2021 8th International Conference on Soft Computing & Machine Intelligence, ISCMI, IEEE, 2021, pp. 237–241.
[196] M. Turkoglu, COVID-19 detection system using chest CT images and multiple kernels-extreme learning machine based on deep neural network, IRBM 42 (4) (2021) 207–214.
[197] R. Murugan, T. Goel, E-DiCoNet: Extreme learning machine based classifier for diagnosis of COVID-19 using deep convolutional network, J. Ambient Intell. Humaniz. Comput. 12 (9) (2021) 8887–8898.
[198] F. Afza, M. Sharif, M.A. Khan, U. Tariq, H.-S. Yong, J. Cha, Multiclass skin lesion classification using hybrid deep features selection and extreme learning machine, Sensors 22 (3) (2022) 799.
[199] X. Luo, X. Li, Z. Wang, J. Liang, Discriminant autoencoder for feature extraction in fault diagnosis, Chemometr. Intell. Lab. Syst. 192 (2019) 103814, http://dx.doi.org/10.1016/j.chemolab.2019.103814.
[200] A. Alharbi, M. Alghahtani, Using genetic algorithm and ELM neural networks for feature extraction and classification of type 2-diabetes mellitus, Appl. Artif. Intell. 33 (4) (2019) 311–328.
[201] A. Doğantekin, F. Özyurt, E. Avcı, M. Koc, A novel approach for liver image classification: PH-C-ELM, Measurement 137 (2019) 332–338.
[202] M.A. Khan, T. Akram, Y.-D. Zhang, M. Alhaisoni, A. Al Hejaili, K.A. Shaban, U. Tariq, M.H. Zayyan, SkinNet-ENDO: Multiclass skin lesion recognition using deep neural network and entropy-normal distribution optimization algorithm with ELM, Int. J. Imaging Syst. Technol. (2023).
[203] M.O. Prates, Spatial extreme learning machines: An application on prediction of disease counts, Statist. Methods Med. Res. 28 (9) (2019) 2583–2594.
[204] M. Jiang, Z. Pan, N. Li, Multi-label text categorization using L21-norm minimization extreme learning machine, Neurocomputing 261 (2017) 4–10.
[205] R. Li, X. Wang, Y. Song, L. Lei, Hierarchical extreme learning machine with L21-norm loss and regularization, Int. J. Mach. Learn. Cybern. 12 (5) (2021) 1297–1310.
[206] R. Li, X. Wang, L. Lei, Y. Song, $L\_\{21\}$-Norm based loss function and regularization extreme learning machine, IEEE Access 7 (2018) 6575–6586.
[207] Preeti, R. Bala, A. Dagar, R.P. Singh, A novel online sequential extreme learning machine with L 2, 1-norm regularization for prediction problems, Appl. Intell. 51 (2021) 1669–1689.
[208] W. Xiao, J. Zhang, Y. Li, S. Zhang, W. Yang, Class-specific cost regulation extreme learning machine for imbalanced classification, Neurocomputing 261 (2017) 70–82.
[209] L. Zhang, H. Yang, Z. Jiang, Imbalanced biomedical data classification using self-adaptive multilayer ELM combined with dynamic GAN, Biomed. Eng. Online 17 (1) (2018) 1–21.
[210] J. Qin, C. Wang, Q. Zou, Y. Sun, B. Chen, Active learning with extreme learning machine for online imbalanced multiclass classification, Knowl.-Based Syst. 231 (2021) 107385.
[211] Y. Guo, B. Jiao, Y. Tan, P. Zhang, F. Tang, A transfer weighted extreme learning machine for imbalanced classification, Int. J. Intell. Syst. 37 (10) (2022) 7685–7705.
[212] A. Pradhan, D. Mishra, K. Das, G. Panda, S. Kumar, M. Zymbler, On the classification of MR images using "ELM-SSA" coated hybrid model, Mathematics 9 (17) (2021) 2095.
[213] J. Zhang, Y. Zhao, F. Shone, Z. Li, A.F. Frangi, S.Q. Xie, Z.-Q. Zhang, Physics-informed deep learning for musculoskeletal modelling: Predicting muscle forces and joint kinematics from surface EMG, IEEE Trans. Neural Syst. Rehabil. Eng. (2022).
[214] S. Ding, N. Zhang, X. Xu, L. Guo, J. Zhang, Deep extreme learning machine and its application in EEG classification, Math. Probl. Eng. 2015 (2015).
[215] L. Li, Y. Fan, M. Tse, K.-Y. Lin, A review of applications in federated learning, Comput. Ind. Eng. 149 (2020) 106854.
[216] R. Wang, R. Shi, X. Hu, C. Shen, Remaining useful life prediction of rolling bearings based on multiscale convolutional neural network with integrated dilated convolution blocks, Shock Vib. 2021 (2021).
[217] H. Li, W. Zhao, Y. Zhang, E. Zio, Remaining useful life prediction using multi-scale deep convolutional neural network, Appl. Soft Comput. 89 (2020) 106113.
[218] S. Zhou, X. Liu, Q. Liu, S. Wang, C. Zhu, J. Yin, Random Fourier extreme learning machine with $l_{2,1}$-norm regularization, Neurocomputing 174 (2016) 143–153.
[219] X. Luo, X. Chang, X. Ban, Regression and classification using extreme learning machine based on $L1$–norm and $L2$–norm, Neurocomputing 174 (2016) 179–186.

Huérfano Maldonado Yoleidy obtained a B.Sc. degree in Education mention Physics and Mathematics in 2013, at the Universidad de Los Andes, Táchira-Venezuela. Then, She obtained the M.Sc. in Mathematics mention Mathematics Education in 2017 at the Universidad Nacional Experimental del Táchira (UNET), Táchira-Venezuela. Currently, She is a Ph.D student in Applied Mathematical Modeling at the Universidad Católica del Maule, Talca-Chile and a member of the Laboratory of Technological Research in Pattern Recognition (LITRP). Finally, She has several publications in international journals and conference proceedings. Her research interests include machine learning techniques, image analysis, image processing and medical image segmentation.

Mora Marco received the B.S. degree in Electronics Engineering and M.S. degree in Electrical Engineering from Universidad de Concepción, Concepción, Chile, in 1998 and 2004, respectively; and Ph.D. in Computer Science from Polytechnical National Institute of Toulouse (INPT), University of Toulouse, Toulouse, France, in 2008. He is a Full Professor in the Department of Computer Science and Industry at Universidad Católica del Maule, Talca, Chile. He is Head and Senior Researcher of the Laboratory of Technological Research in Pattern Recognition (LITRP), Universidad Católica del Maule. His research interests are digital image processing, neural networks, biometrics, and industrial applications of pattern recognition.

Vilches-Ponce Karina received the B.S. and M.Sc. degrees in mathematics from the Pontificia Universidad Católica de Valparaíso in 2007 and 2009, respectively, and Ph.D. in mathematical modeling from the Universidad de Chile and in mathematical sciences from Sorbonne University (both in 2014). She is an associate professor in the Faculty of Basic Sciences at the Universidad Católica del Maule, where she develops research in interdisciplinary contexts. She is also a researcher at the Laboratory of Technological Research in Pattern Recognition (LITRP). Her research interests consist mainly of the application of mathematical background to real problems in science, computation, and biology.

Hernández-García Ruber received an engineer in computer science degree from the Universidad de Ciencias Informáticas, Havana, Cuba, in 2007; M.Sc. degrees in audiovisual information systems and applied informatics from Universidad de Málaga (Spain) and Universidad de Ciencias Informáticas, in 2010 and 2011, respectively; and a Ph.D. degree in computer science from Universidad de Ciencias Informáticas in 2014. He currently serves as an academic researcher at the LITRP Laboratory and at the Centro de Investigación de Estudios Avanzados del Maule, Universidad Católica del Maule, Chile. His research interests include video and image processing, pattern recognition, and biometric systems under parallel platforms.

Gutiérrez Rodrigo received a degree in mathematics pedagogy from Universidad Católica del Maule, Chile, in 2015. He also received M.Sc. in Applied Mathematics from the Universidad del Bío-Bío, Chile, in 2017, and Ph.D. in Applied Mathematical Modeling from the Universidad Católica del Maule in 2021. He is currently an associate professor at the Faculty of Basic Sciences of the Universidad Católica del Maule, whose research focuses on construction and the interpretive analysis of mathematical models. His research interests included the mathematical modeling of biological and social systems, such as consumer behaviors and animal communication as a dynamics network.

Vera Miguel received a degree of accreditation in mathematics education from Universidad de Los Andes, Mérida, Venezuela, in 1996. He also received M.Sc. and Ph.D. degrees in mathematics from the Universidad de Los Andes, Mérida, Venezuela, in 2005 and 2014, respectively. Currently, he is a researcher at the Universidad Simón Bolívar, Cúcuta, Colombia. He is also an accredited researcher of the Minciencias organization in Colombia. He is a member of the Scientific Modeling and Business Innovation Research Group of the Universidad Simón Bolívar (GIMCINE). His research interests include machine learning techniques, image analysis, image processing, and medical image segmentation.