



# Survey on Deep Fuzzy Systems in Regression Applications: A View on Interpretability

Jorge S. S. Júnior<sup>1</sup> · Jérôme Mendes<sup>1</sup> · Francisco Souza<sup>2</sup> · Cristiano Premebida<sup>1</sup>

Received: 16 April 2022 / Revised: 17 November 2022 / Accepted: 3 May 2023 / Published online: 5 June 2023  
© The Author(s) 2023

**Abstract** Deep learning (DL) has captured the attention of the community with an increasing number of recent papers in regression applications, including surveys and reviews. Despite the efficiency and good accuracy in systems with high-dimensional data, many DL methodologies have complex structures that are not readily transparent to human users. Accessing the interpretability of these models is an essential factor for addressing problems in sensitive areas such as cyber-security systems, medical, financial surveillance, and industrial processes. Fuzzy logic systems (FLS) are inherently interpretable models capable of using nonlinear representations for complex systems through linguistic terms with membership degrees mimicking human thought. This paper aims to investigate the state-of-the-art of existing deep fuzzy systems (DFS) for regression, *i.e.*, methods that combine DL and FLS with the aim of achieving good accuracy and good interpretability. Within the concept of explainable artificial intelligence (XAI), it is essential to contemplate interpretability in the development of intelligent models and not only seek to promote

explanations after learning (post hoc methods), which is currently well established in the literature. Therefore, this work presents DFS for regression applications as the leading point of discussion of this topic that is not sufficiently explored in the literature and thus deserves a comprehensive survey.

**Keywords** Deep fuzzy systems · Deep regression · Explainable artificial intelligence (XAI) · Interpretability · Deep learning

## 1 Introduction

The goal in regression is to predict one or more variables  $\mathbf{y}(k) = (y_1(k), \dots, y_m(k))^T$  from the information provided by measurements  $\mathbf{x}(k) = (x_1(k), \dots, x_p(k))^T$  for a given sample  $k$ . Customary,  $\mathbf{y}(k)$  are referred as targets, outputs, or dependent variables, while  $\mathbf{x}(k)$  are commonly referred as predictors, inputs, covariates, regressors, or independent variables. Regression models covers several application areas, as economic growth problems [1, 2], air quality prediction [3, 4], medicine [5, 6], chemical industries [7, 8], and industrial processes [9, 10]. Recent studies show that regression models have become predominant in increasingly complex real-world systems due to the large availability of data, inclusion of nonlinear parameters, and other aspects intrinsic to the application area. For complex systems, traditional machine learning techniques (*i.e.*, non-deep/shallow techniques) may become limited due to two main challenges: big-data explosion (high-dimensionality, high number of observations) and increase in complexity, caused by dynamics of nowadays applications. Deep learning (DL) methods have gained prominence in recent

---

✉ Jorge S. S. Júnior  
jorge.silveira@isr.uc.pt

Jérôme Mendes  
jermendes@isr.uc.pt

Francisco Souza  
f.souza@science.ru.nl

Cristiano Premebida  
cpremebida@isr.uc.pt

<sup>1</sup> Department of Electrical and Computer Engineering, Pólo II, University of Coimbra, Institute of Systems and Robotics, 3030-290 Coimbra, Portugal

<sup>2</sup> Dept. Analytical Chemistry & Chemometrics, Radboud University, 6525 AJ Nijmegen, The Netherlands

years due to their ability to represent systems in complex structures with multiple levels of abstraction and high-level features. However, DL may have limitations such as dependence on a high number of samples, hyperparameter sensitivity, and interpretability issues [11, 12], which limits the application in critical systems or one that requires accountability in its results. In this sense, deep fuzzy systems (DFS) have emerged as a viable method over DL to balance accuracy and interpretability in complex real-world systems.

Regression models can be categorized as [13]: (i) “white-box” when the input–output mapping is built upon first-principle equations, (ii) “black-box” when the mapping is derived from the data (also referred to as data-driven modeling), or (iii) “grey-box” when the knowledge about the input–output mapping is known beforehand and integrated along with data-driven modeling. White-box models are advantageous for promoting interpretations of the internal mechanisms associated with input–output mapping. On the other hand, black-box models can address complex systems with predictive analysis without prior knowledge of the system. Grey-box models can combine the interpretability presented by “white-box” and the ability to learn from data given by “black-box” models (*e.g.*, fuzzy systems). Regarding data-driven modeling, the dependency between input and output can be built by linear or nonlinear models. A regression model is linear when the rate of change between input–output is constant due to the linear combination of the inputs. Examples include the multivariate linear regression models. Data-driven nonlinear regression is adopted when the input–output dependence is nonlinear and can not be covered by linear modeling. There is a plethora of methods for nonlinear regression, and its applicability is problem-dependent. Examples include fuzzy systems, support vector regression, artificial neural networks (*e.g.*, non-deep/shallow and deep networks), and rule-based regression (*e.g.*, decision trees and random forest).

Following are brief discussions of recent surveys and reviews that have showcased the applicability of DL to various regression application domains. The work of Han et al. [14] reviews DL models for time-series forecasting, where the DL models are categorized as: (i) discriminative, where the learning stage is based on the conditional probability of the output/target given an observation, (ii) generative, which learn the joint probability of both output and observation, with the generation of random instances), or (iii) hybrid, a combination of generative and discriminative DL methods. The authors demonstrated that DL models are effective at discriminating complex patterns in time series with high-dimensional data by implementing them in benchmark systems and using a real-world use case from the steel industry. Sun et al. [11] discuss the use of

DL for soft sensor applications, showing the trends, the scope of applications in industrial processes, and the best practices for model development. The authors established some directions for future research, such as working on DL solutions to overcome the limitation in learning in scenarios with a lack of labeled samples (*e.g.*, semi-supervised methods), hyperparameter optimization, solutions to improve model reliability (*e.g.*, model visualization), and the development of DL methods with distributed and parallel modeling. Torres et al. [15] survey DL architectures for time-series forecasting. Furthermore, the authors discuss the practical aspects that must be considered when using DL methods to solve complex real-world problems in big-data settings, which include the existing libraries/toolbox, the techniques for automatic optimization of model structures, and the hardware infrastructure. Pang et al. [16] and Chalapathy et al. [17] present an overview of DL studies for anomaly detection; they also discuss the complexities and different types of DL models for different domains (*e.g.*, classification, autoregressive, unsupervised, and semi-supervised) with application to cyber-security systems, medical monitoring, financial surveillance, and industrial processes. Additional review papers on DL for regression are referred to here: [18–20]. It is noted from these works that DL techniques have some advantages over non-deep methods, such as the ability to learn complex representations with automatic feature engineering, not requiring prior experience or knowledge, good performance by increasing the dimensionality of the data, among other specific advantages depending on the type of framework and application [21].

Despite the advantages portrayed in the literature, deep learning has limitations, such as the need for a sufficiently large dataset for model training, sensitivity to hyperparameter selection, and lack of interpretability [11, 12]. Also, there is a lack of proper explanations of the internal structure of deep model structures, which raises concern in applications that directly and indirectly impact human life as well for operational decisions [22, 23]. Having this concern in mind, recent works address the interpretability issue of DL from the following principles of “eXplainable Artificial Intelligence” (XAI) systems [24]: the existence of an appropriate explanation for each decision made; each explanation must be meaningful to the user; the process must accurately consider what happens in the system; identify the situations in which the system may or may not function properly (knowledge limits).

Fuzzy logic systems (FLS), composed of IF-THEN rules with linguistic terms mimicking human thought [25], is one of the research areas that contemplates the XAI principles. FLS has a wide range of approaches to nonlinear systems, primarily in terms of interpretability, whether related to the complexity and semantics of fuzzy rules, notation

readability, coverage of input data space (operation regions), and so on [26, 27]. Moral et al. [28] show the main benefits of adopting FLS for the development of explainable methodologies, with a discussion of fundamental concepts and definitions associated with XAI and FLS to how to design increasingly interpretable models. Therefore, DL techniques can be complemented with FLS, developing deep fuzzy systems (DFS) and then providing an easy-to-understand and easy-to-implement interface to efficiently address the main drawbacks of DL and FLS, thus ensuring good accuracy and good interpretability. The surveys of [29] and [30] investigated some recent trends in DFS models and real-world applications (e.g., time series forecasting, natural language processing, traffic control, and automatic control). However, despite the benefit of adopting FLS principles to DL systems, no comprehensive survey or review has been conducted focusing exclusively on deep fuzzy for regression problems.

This paper surveys and discusses the state-of-the-art on deep fuzzy techniques developed to deal with a diverse range of regression applications. Initially, Sect. 2 presents fundamental concepts about FLS and XAI. Then, an overview of DL techniques commonly used for regression will be presented in Sect. 3, namely Convolutional Neural Networks, Deep Belief Networks, Multilayer Autoencoders, and Recurrent Neural Networks. Next, Sect. 4 shows the literature on deep fuzzy systems in two ways: (i) standard deep fuzzy systems, based on fundamental FLS; and (ii) hybrid deep fuzzy systems, with the combination of FLS and the conventional deep models discussed in Sect. 3. Finally, Sect. 5 presents general discussions based on the state-of-the-art surveyed.

## 2 Background

### 2.1 Fuzzy Logic Systems

Developed initially by Lotfi A. Zadeh [31], FLS are rule-based systems composed mainly of an antecedent part, characterized by an “IF” statement, and a consequent part, characterized by a “THEN” statement, allowing the transformation of a human knowledge base into mathematical formulations, thus introducing the concept of linguistic terms related to the membership degree.

The basic configuration of a fuzzy logic system, shown in Fig. 1, depends on an interface that transforms the real input variables into fuzzy sets (fuzzifier), which are interpreted by a fuzzy inference model to perform an input–output mapping based on fuzzy rules. Thus, the mapped fuzzy outputs go through an interface that transforms them into real output variables (defuzzifier) [32, 33].

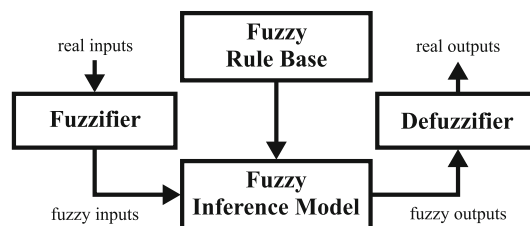


Fig. 1 Basic configuration of fuzzy logic systems

Some well-known fuzzy systems are Mamdani fuzzy systems [34], Takagi-Sugeno (T-S) fuzzy systems [35], and Angelov-Yager’s (AnYa) fuzzy rule-based systems using data clouds [36]. Among these, the T-S fuzzy models stand out for their ability to decompose nonlinear systems into a set of linear local models smoothly connected by fuzzy membership functions [37]. T-S fuzzy models are universal approximators capable of approximating any continuous nonlinear system, that can be described by the following fuzzy rules [38]:

$$R^i : \mathbf{IF} \ x_1(k) \text{ is } F_1^i \text{ and } \dots \text{ and } x_p(k) \text{ is } F_p^i \quad (1)$$

$$\mathbf{THEN} \ y^i(k) = f^i(x_1(k), \dots, x_p(k)),$$

where  $R^i$  ( $i = 1, \dots, N$ ) represents the  $i$ -th fuzzy rule,  $N$  is the number of rules,  $x_1(k), \dots, x_p(k)$  are the input variables of the T-S fuzzy system,  $F_j^i$  are the linguistic terms characterized by fuzzy membership functions  $\mu_{F_j^i}^i$ , and  $f(x_1(k), \dots, x_p(k))$  represents the function model of the system of the  $i$ -th fuzzy rule [39].

### 2.2 Explainable Artificial Intelligence

Although the explainable artificial intelligence (XAI) concept is often associated with a homonymous program formulated by a group of researchers from the Defense Advanced Research Projects Agency (DARPA) [40], the principles related to explainability gained strength from the 1970 s onwards. The earliest works presented rule-based structures and decision trees with human-oriented explanations, such as the MYCIN system proposed in [41] developed for infectious disease therapy consultation, the tutoring program GUIDON proposed in [42] based on natural language studies, among numerous other systems [43–46]. Although many authors commonly use the terms “explainability” and “interpretability” as synonyms, Rudin [47] discusses the problem of using purely explainable methods to only provide explanations from the results obtained in black-box models (post hoc analysis), demystifying the importance of developing inherently interpretable methodologies with causality relationships that are understandable to human users.

### 3 Deep Regression Overview

Deep neural networks (DNN) have emerged due to their architecture with multiple levels of representation and their remarkable performance in a variety of tasks [48]. This section presents the DL techniques commonly employed in regression problems to provide a better understanding and context for the literature review of deep fuzzy regression performed in Sect. 4.

#### 3.1 Convolutional Neural Networks

Convolutional neural networks (CNNs or ConvNets) are feedforward neural networks with a grid-like topology that are used for applications such as time-series data processing in 1-D grids and image data processing in 2-D pixel grids [49]. Figure 2 presents a CNN architecture for processing time-series data in 1-D grids.

The “neocognitron” model proposed in [50] is frequently referred to as the inspiration model for what is currently known about CNNs. First proposed in [51], the neocognitron aimed to represent simple and complex cells from the visual cortex of animals which present a mechanism capable of detecting light in receptive fields [52].

CNNs use a math operation on two functions called “convolution”, with the first function referred to as input, the second function as kernel, and the convolution’s output as the feature map [49]. Outputs from the convolution layer go through a pooling layer (downsampling), which performs an overall statistic of the adjacent outputs by reducing the size of the data and associated parameters via weight sharing [11, 53]. After the data is processed along with the layers that alternate between convolution and pooling, the final feature maps go through a fully connected (dense) layer to extract high-level features. For regression problems, the extracted features can be combined in a prediction mechanism with an activation function or a supervised learning model (e.g., support vector regression) to estimate the final output [54, 55].

CNNs in the context of regression have been explored in several domains, for example, traffic flow forecasting of real road networks [56, 57], prediction of natural environmental factors using data from meteorological institutes

[58–61], industrial process optimization [62–64], electric load forecasting [65–69], and chemical process analysis [70, 71]. Despite showing good performance with the extraction of spatially organized features, essentially for pre-processing, CNN’s performance depends on a large amount of data and the correct choice of hyperparameters, being computationally intensive [21, 72].

#### 3.2 Deep Belief Networks

Deep Belief Networks (DBNs) are probabilistic generative models proposed by Hinton et al. [73], which have a hierarchical structure as illustrated in Fig. 3.

The DBNs are composed of multiple layers of latent variables (hidden units) of binary values, organized into multiple learning modules called restricted Boltzmann machines (RBMs).

Each RBM comprises a layer of visible units for data representation and a layer of hidden units for feature representation, learned by capturing higher-order correlations from the data. The two RBM layers, with no connections within layers, are connected by a matrix of symmetrically weighted connections, with a total of  $L$  weight matrices  $W = \{w_1, w_2, \dots, w_L\}$ , considering a DBN of  $L$  hidden layers. All units in each layer have a bidirectional connection with all units in neighboring layers, except for the last two layers,  $L$  and output, that have a unidirectional connection [49]. In the DBN architecture of Fig. 3, the layer of visible units in the first RBM represents the input variables  $x$  and the subsequent layers represent hidden units  $h$ , progressing hierarchically until reaching the estimated output, with the output weights  $w_{out}$ .

Recent DBN literature for regression has primarily addressed industrial problems in the context of process monitoring [74–79], soft sensing [80, 81], and prognostics [82, 83]. Also, DBNs were applied in time series forecasting problems such as traffic flow [84, 85], environmental prediction [86, 87], and stock price [88], as well as modeling benchmark systems [89, 90]. Studies with DBNs have difficulties in explaining the effect of hidden units on the dynamics of the system, leading to interpretability issues. As Fig. 3 shows, a DBN needs to undergo unsupervised and supervised learning, whose training process

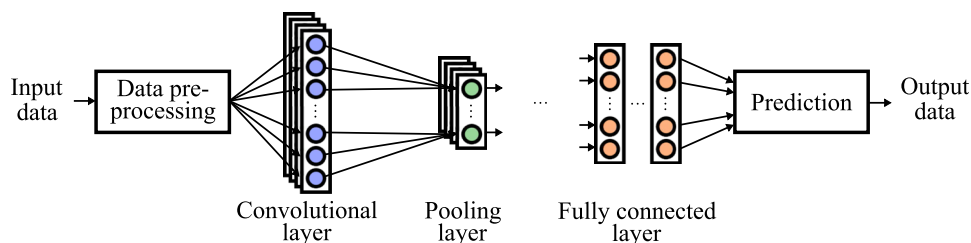


Fig. 2 Convolutional neural network, with 1-D architecture

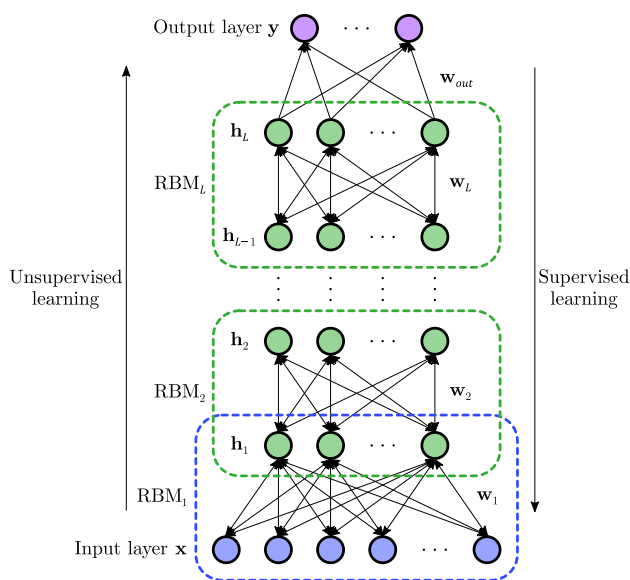


Fig. 3 Deep belief networks architecture

becomes slower with increasing DBN structure. In this way, the DBN becomes sensitive to noisy inputs by not correctly readjusting its low-level parameters [91].

### 3.3 Multilayer Autoencoders

Autoencoders (AEs) are feedforward neural networks used for dimensionality reduction and representation learning, whose training is aimed at mimicking inputs to outputs [49]. A historical overview of deep learning in [92] presented some early works of AEs in the literature, such as the work in [93] that proposes unsupervised architectures to reconstruct the inputs through internal representations. Another early work was published in [94], where the authors explore the effect of hidden units in simple two-layer associative networks, in which they want to map input patterns to a set of output patterns. As presented in

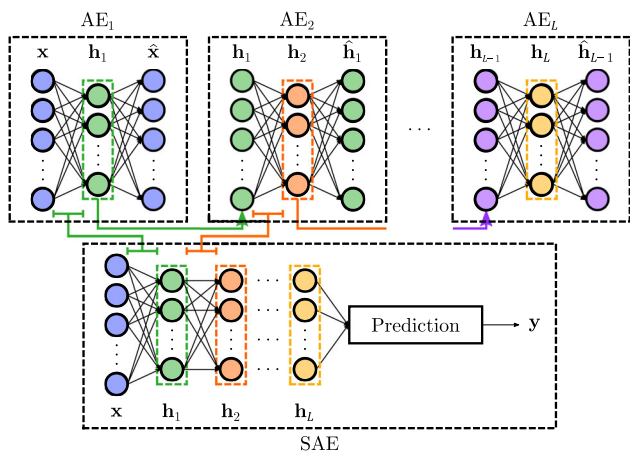


Fig. 4 Modified stacked autoencoder architecture for regression

Fig. 4, a single AE essentially has a structure composed of a layer with input variables  $x$ , a layer with hidden units  $h$  (which performs an encoding used to represent the input), and an output layer with the reconstructed inputs represented with a “hat” symbol (e.g.,  $\hat{x}$ ). These layers are interconnected by an encoder function (between input and hidden) and by a function decoder (between hidden and output) [49]. As an evident constraint, the number of neurons in the input layer must be the same number of neurons in the output layer, setting AE to unsupervised pre-training or feature extraction [95].

A viable alternative to deal with increasingly complex data is to increase the number of hidden layers in standard AEs, enabling the development of deep network architectures. A well-known configuration of multilayer autoencoders found in the literature is stacked autoencoder (SAE). As illustrated in Fig. 4, an SAE is developed from the grouping of  $L$  AEs, where the hidden layers of the autoencoders are stacked hierarchically, performing an unsupervised layerwise learning algorithm. Thus, the reconstruction of inputs after the  $L$ -th hidden layer can be disregarded to address regression problems. As for CNNs, a prediction mechanism or a supervised learning model can be included after the  $L$ -th hidden layer to estimate the system output. Related parameters, such as weights  $W = \{w_1, w_2, \dots, w_L\}$  between layers, are fine-tuned by a supervised method (e.g., backpropagation algorithm) [96].

Recent studies using AEs within a deep architecture for regression have considered soft sensing applications for quality variable prediction [97–101]. Other cases of applications in industrial processes include CNC turning machines [102, 103], end-point quality prediction [104, 105], and prognostics [106–109]. Other works were developed for time-series forecasting applications, such as natural environmental factors [110–112], electricity load forecasting [113], and tourism demand [114]. Some of the limitations of multilayer autoencoders include the sensitivity to errors or loss of information from the first layer, impairing learning as it progresses through the hidden layers. Thus, the nature of encoding and decoding by hidden layers can cause a loss of interpretability and an increase in computational cost [21].

### 3.4 Recurrent Neural Networks

Recurrent Neural Networks (RNNs) are artificial neural networks with an internal state that allow the use of feedback signals between neurons [115]. One of the early works that culminated in the popularization of RNNs was proposed in [116], with the development of content-addressable memory systems called Hopfield networks whose dynamics have a Lyapunov function (or energy function) to direct to a local minimum of “energy” associated with the

system states [117]. Another early work, published in [118], presented a new learning procedure, backpropagation, which adjusts the weights of connections of recurrent networks with “internal hidden units”. Due to the presence of an “internal memory”, RNNs are often used to process data in the time domain (sequential information), with weight sharing through hidden states [119]. Figure 5 shows the architecture of an RNN, whose hidden states  $h$  represent the “internal memory” of the system, and as the inputs are sequentially observed, the corresponding outputs are estimated. The weights  $W^{in}$ ,  $W^h$  and  $W^{out}$  are auxiliary parameters that are shared across time [120]. The states are updated at every temporal instant until completing all the input sequences of length  $K$  [121, 122]. The RNN learning process depends on a “backpropagation through time” algorithm, which is a gradient-based technique that updates parameters recursively starting from the last temporal instant and going backward in time [123].

In deep learning, there are many variations of standard RNNs, such as Long Short-Term Memory networks (LSTMs), Gated Recurrent Units (GRUs), and Echo State Network (ESNs) [49]. Some of these variations were proposed to cope with some limitations present in standard RNNs, such as exploding and vanishing gradients (instability in networks caused by a large variation in model parameters), overfitting, and difficulty to store low-level features in long data sequences [124]. In this document, only LSTMs and ESNs employed for regression problems will be discussed.

### 3.4.1 Long Short-Term Memory

Long Short-Term Memory networks (LSTMs) were initially developed in [125] to overcome the issues of RNNs associated with vanishing/exploding gradients. These issues can occur during the training of long temporal sequences with the backpropagation through time algorithm. In this sense, during successive operations in compound functions with the weight matrices, gradients can exponentially reach very low values close to zero

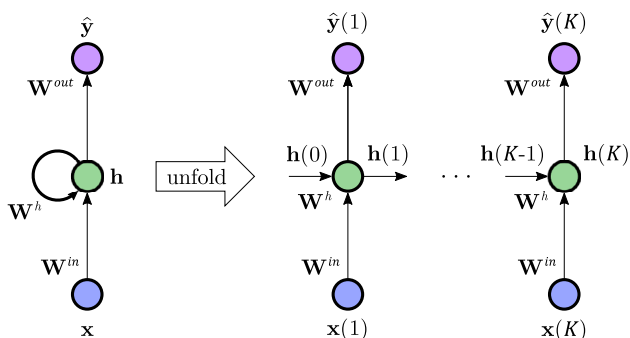


Fig. 5 Recurrent neural network architecture

(vanishing) or very high values (exploding) [49]. Figure 6 illustrates the architecture of an LSTM.

LSTMs introduced “gates”, nonlinear elements that control memory cells using sigmoidal functions  $\sigma$ , hyperbolic tangent functions, current observation  $x(k)$  and hidden units  $h(k - 1)$  from the previous time instant [122]. Each of these memory cells has input, output and forget gates ( $i$ ,  $o$  and  $f$ , respectively), that protect the information from perturbations caused by irrelevant inputs and irrelevant memory contents [72]. The information stored in the cells represents the states  $s(k)$  obtained from the data processed in the time domain. Despite having the same inputs and outputs as a standard RNN, an LSTM cell has an internal recurrence (self-loop) to propagate the information flow through a long sequence and, therefore, more parameters to be adjusted [49].

Due to the ability to process data sequentially and ensure long-term dependencies, LSTMs are primarily used for time-series forecasting in traffic flow [126, 127] and natural environment factors [128–133]. LSTM methodologies have also been applied in the industrial context for soft-sensing with attention mechanism [134, 135] and key quality prediction [136, 137], process monitoring [138, 139], and prognostics [140–142]. Other applications of LSTMs include electric load forecasting [143, 144]. LSTMs in regression problems, despite the feasibility, can suffer from vanishing gradient from the saturation of cell states, which must be reset occasionally. In addition, there is a computational cost issue in the application of LSTMs as they require a high memory bandwidth to perform the associated functions in complex systems [145].

### 3.4.2 Echo state network

Echo State Networks (ESNs) are variations of RNNs that were developed in [146] and share the basic ideas of reservoir computing of Liquid State Machines from [147]. The term “reservoir computing” stands for a homonymous

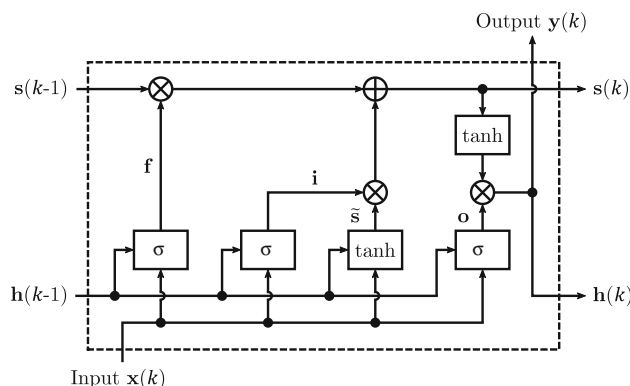


Fig. 6 The long short-term memory cell

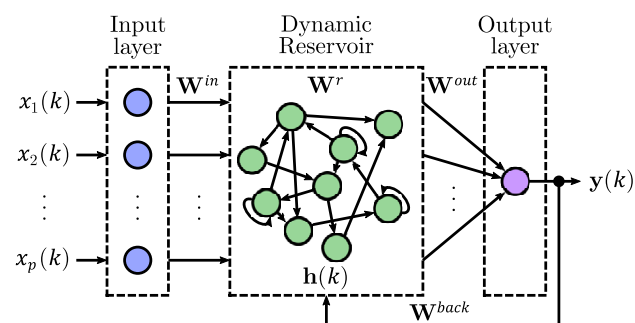
research stream that introduced the concept of a dynamic reservoir, in place of the hidden layer, with many sparsely connected neurons [148]. In addition, this reservoir must have a stability condition known as “echo state property”, which allows for a gradual reduction in the effect of previous states and inputs on future states over time [149]. Figure 7 illustrates the architecture of an ESN.

An ESN induces nonlinear response signals from the input signals  $\mathbf{x}(k)$ , whose resulting state  $\mathbf{h}(k)$  echoes the input information, estimating a desired output signal  $y(k)$  [146]. In addition to the input weight  $\mathbf{W}^{in}$  and output weight  $\mathbf{W}^{out}$  that are present in a standard RNN, the ESN features the reservoir weights  $\mathbf{W}^r$  and, optionally, the output-to-reservoir feedback weights  $\mathbf{W}^{back}$ . The fundamental idea for the functioning of an ESN is to adjust only  $\mathbf{W}^{out}$  during training, while the rest are randomly assigned and fixed before learning [150].

Recent applications of ESNs are time-series forecasting in chaotic systems [151–155], wind power systems [156–160], solar energy systems [161–164], and benchmark systems [165–167]. Other applications include chemical processes [168, 169], soft sensing [170], electrical/power systems [171–173], and the prognosis of turbofan engines [174, 175]. The recent literature shows that random connectivity within the dynamic reservoir of an ESN can lead to interpretability issues. Furthermore, ESN hyperparameters, such as the number of units in the reservoir and input scaling, are limited to a short operating space that ensures maximum model performance [176].

## 4 Deep Fuzzy Regression

Deep fuzzy systems (DFSs) are models built on top of DL structures with fuzzy logic systems (FLSs). DFS aims to overcome the lack of interpretability of DL systems and the limitations of FLS when dealing with high-dimensional data. DFS is often referred to in the literature as an explainable model due to the incorporation of fuzzy logic



**Fig. 7** The echo state network

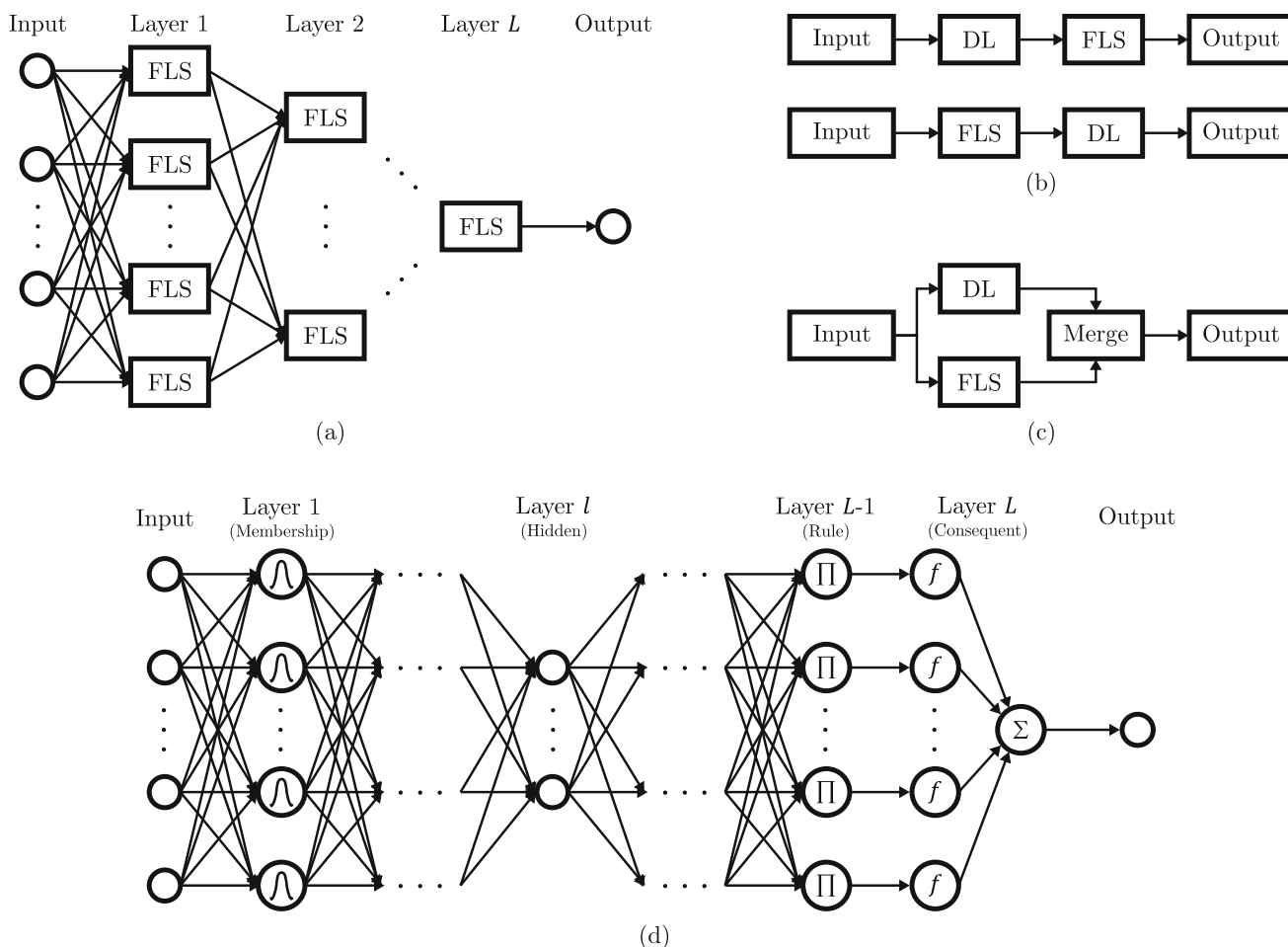
into its core. From the definition from [177], an explainable AI (XAI) system should provide capabilities accessible to human understanding, reflecting positively on the health of the system’s processes and being able to operate even in unforeseen situations. Although DFS is promoted as an XAI system by definition, this is not the case, as evidenced by works in the literature. This section surveys recent applications of DFS for regression, with the main focus on its structure and if it adheres to the XAI principles.

### 4.1 Survey on Deep Fuzzy Systems

This survey will follow two stages to review the DFS for regression applications. First, the DFS will be categorized according to its structure. Secondly, the models will be categorized according to whether they follow the XAI principles. The structures of deep fuzzy models can be represented in several forms, such as those illustrated in Fig. 8. Here, deep fuzzy structures are summarized into two categories: (i) Standard DFS and (ii) Hybrid DFS. A model belongs to the first category when the blocks of fuzzy systems are stacked in series, in parallel, or hierarchically (see Fig. 8a). Also, there are cases where the architecture of such systems resembles neural network architecture, such as the dense DNN architecture (see Fig. 8d). The second category includes hybrid methodologies, where conventional DL models are combined with FLS. The combination of DL and FLS is commonly in an ensemble form (see Fig. 8b,c) or mixed form (see Fig. 8d).

Aside from the deep fuzzy structures, the discussed works will be investigated whether they are in synergy with the XAI principles. If so, these works will be classified following the categories defined in [22]: understanding and scope of the explainability. In terms of understanding, the models will be classified as (i) transparent or (ii) opaque. Transparent models are those in which the decisions, predictions, or inner functioning are perceptible or are visible; they are considered opaque otherwise. The scope is related to accessing the model’s interpretability through post hoc explanations, classified as (i) local, (ii) global, and (iii) visual. Local explanations facilitate comprehension of small regions of interest in the input space for a given decision/prediction. Global explanations when such desired understanding considers the entire sample space. Visual explanations are required when visual interfaces are needed to demonstrate the influence of features on decisions.

The following sections will discuss the methods containing DFS for regression problems. Section 4.1.1 will discuss Standard DFS, while Sect. 4.1.2 will discuss Hybrid DFS.



**Fig. 8** Examples of deep fuzzy system frameworks: **a** multiple FLSs organized hierarchically; **b** sequential ensemble models; **c** parallel ensemble models; and **d** fuzzy neural network fully integrated into a deep architecture or mixed with parts of a conventional deep model

### 4.1.1 Standard Deep Fuzzy Systems

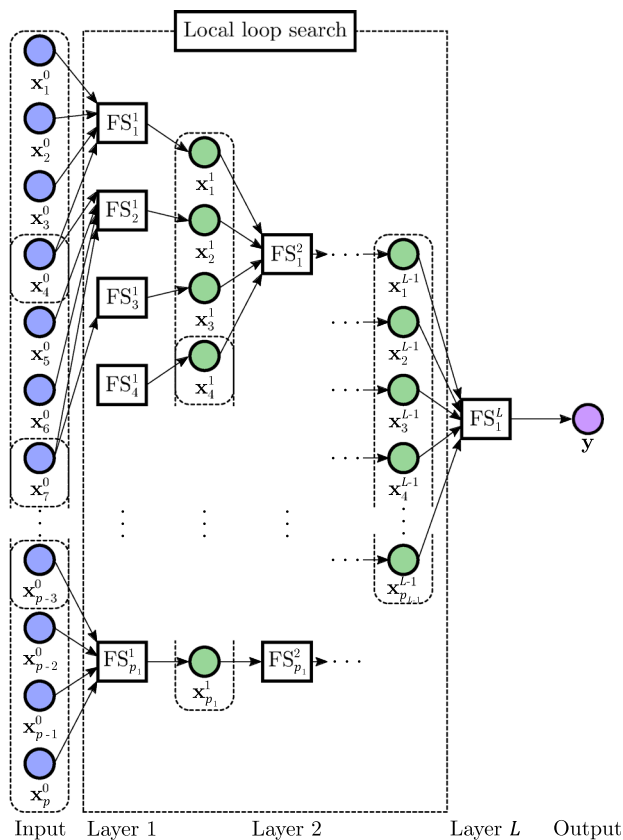
The methods discussed in this section have a DFS structure and were used in any regression problem domain while adhering to the Standard DFS structure. They are discussed in order of structural similarity, with structurally similar methods following each other.

In [178], a model called Randomly Locally Optimized Deep Fuzzy System (RLODFS) is proposed. It is composed of a hierarchical structure of a bottom-up layer-by-layer type with FLS, similar to a fully connected DNN; the structure of RLODFS is shown in Fig. 9. In RLODFS, the input variables are divided into several fuzzy subsystems, allowing the decrease of the size of the fuzzy rule’s antecedent, thus reducing the learning complexity when compared to learning with all input variables at once. The structure of RLODFS shows several groupings with sharing of input variables performed randomly for the fuzzy subsystems of level 1, whose outputs are used as inputs of the fuzzy subsystems of the subsequent levels until

reaching the last level fuzzy system (used to estimate the model output). The training of each fuzzy subsystem was followed by the Wang-Mendel algorithm [179], which performs the construction of fuzzy rules from a small-scale observation set (data pairs) with input–output mapping. Finally, a random local loop optimization strategy is performed to remove feature combinations and corresponding subsystems with low correlation to achieve fast convergence [178]. The performance of RLODFS on the prediction of 12 real-world datasets from the UCI repository<sup>1</sup> was compared with other methods such as DBN, LSTM, and generalized regression neural network (GRNN). From the authors’ perspective, RLODFS has good interpretability due to its structure, clear physical meaning of its parameters, and the ease of locating fuzzy rules that may fail for future optimizations. However, there is a lack of transparency in using input sharing strategies, which increases the method complexity. Furthermore, the selection of the number of features per fuzzy subsystem needs to be

<sup>1</sup> <http://archive.ics.uci.edu/ml>





**Fig. 9** The structure of RLODFS based on grouping and input sharing, adapted from [178]

carefully done, whose manual/arbitrary choice may not reflect the real needs of the case study under analysis.

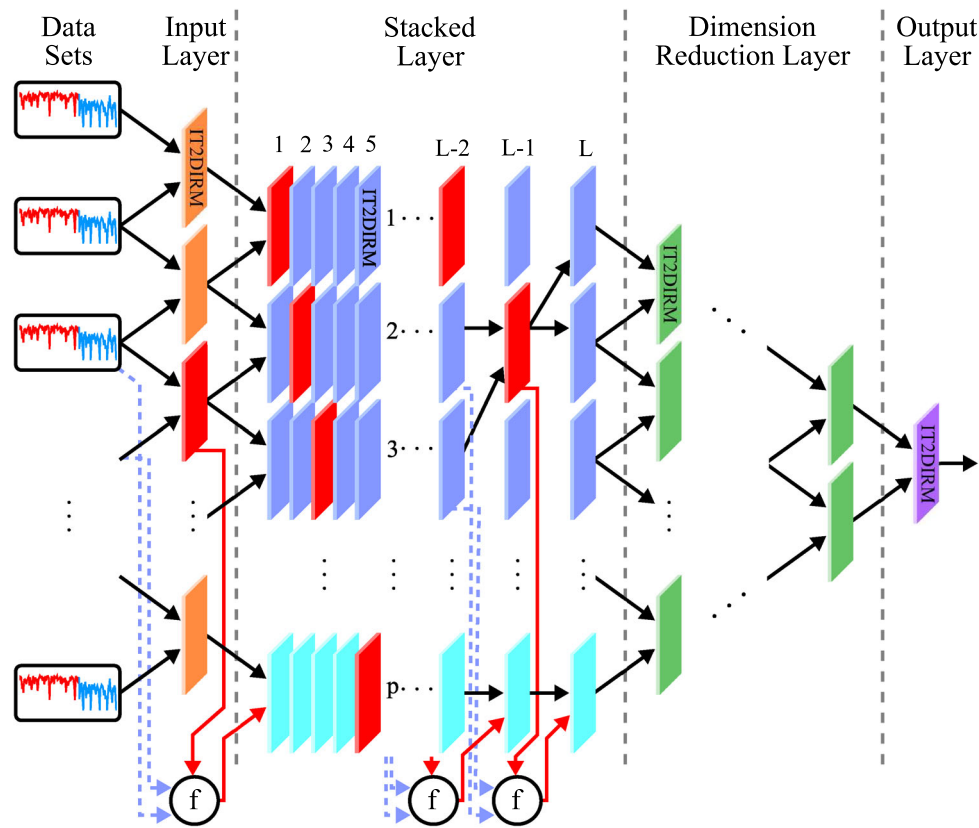
The study in [180] proposes a stacked structure composed of double-input rule modules and interval type-2 fuzzy models, abbreviated as IT2DIRM-DFM. The proposed model is illustrated in Fig. 10, where four layers are presented: the input layer, which deals directly with the original input data, grouped two-by-two in each rule module; the stacked layer, where the signals coming from the input layer become the inputs of the first layer, whose output becomes the input of the second layer, and so on; the dimension reduction layer, where the width of the IT2DIRM-DFM is hierarchically reduced until it becomes two; and the output layer, where the latest IT2DIRM-DFM model produces the final forecasting results. Still, the authors discuss the interpretability of the model showing layered learning and fuzzy rules composed of only two variables in the antecedent, partitioned into interval type-2 second-order rule partitions. The resulting model was evaluated in two real-world applications, using subway passenger data from Buenos Aires, Argentina, and traffic flow data from California Highway System. The results allowed to verify the interpretability from consistent readability of which partitions and bounds of the inputs are

used in each fired rule from each rule module. However, the proposed method is only interpretable locally in each rule module and not globally, whose structure does not reflect the depth or number of layers needed to address the experiments. Furthermore, it is not clear the motivations for considering a function, denoted as  $f$ , in each layer related to the worst-performing module (this is shown at the bottom of Fig. 10).

Another work that explores double-input rule modules within a stacked deep fuzzy model (in a hierarchical way) was proposed in [181] using datasets related to photovoltaic power plants from Belgium and China. The authors investigate the interpretability of the resulting model, called DIRM-DFM, with conclusions similar to [180], mainly regarding the composition of fuzzy rules. However, in DIRM-DFM, they promote more transparency and simplicity. Some other studies deal with interval type-2 fuzzy models for deep learning in regression problems. In [182], a novel dynamic fractional-order deep learned type-2 FLS was proposed and constructed using singular value decomposition and uncertainty bounds type-reduction. The resulting model was implemented with two chaotic benchmark system simulations, a simulation for the prediction of the glucose level of type-1 diabetes patients, and a dataset of a heat transfer system with an experimental setup. In addition to determining the limit values of the input data (upper and lower singular values), the authors used stability criteria of fractional-order systems, allowing to reduce the necessary number of fuzzy rules and reduce the complexity of nonlinear systems. An evolving recurrent interval type-2 intuitionistic fuzzy neural network (FNN) was proposed in [183], and it was evaluated using regression datasets from the KEEL repository,<sup>2</sup> Mackey–Glass time series, and a simulated second-order time-varying system. Intuitionistic evaluation, fire strength of membership degree and strategies for adding and removing fuzzy rules were considered to improve uncertainty modeling. Both studies in [182] and [183] did not present an analysis of the interpretability of their models.

Methods that use multiple neuro-fuzzy systems in a deep hierarchical structure were proposed in [184] and [185]. The work in [184] proposed a deep model that cascades multiple neuro-fuzzy systems modified as multivariable generalized additive models, with application in real ecological time series, the Darwin sea level pressure. The resulting model manages to locally detail the mechanisms of each neuro-fuzzy system in each layer. However, it presents incomplete discussions and results regarding the increase in the network depth and the influence of the inputs and partial outputs of the layers on the final prediction. In [185], a hybrid cascade neuro-fuzzy network

<sup>2</sup> <https://sci2s.ugr.es/keel/datasets.php>



**Fig. 10** The structure of the double-input rule modules stacked deep interval type-2 fuzzy model (IT2DIRM-DFM), adapted from [180]

was proposed, which is composed of multiple extended neo-fuzzy neurons with adaptive training designated for online non-stationary data stream handling. Each layer has a generalization node that performs a weighted linear combination to obtain an optimal output signal. The experimental results in electrical loads prediction, using Southern Ukraine’s data from 2012, show the authors’ search for better accuracy, although at the cost of increasing membership functions to cover the input space and increasing adjusted parameters (weights).

The work in [186] proposed a deep learning recurrent type-3 fuzzy system applied for modeling renewable energies (*i.e.*, power generation of a 660kW wind turbine and solar radiation generated by sunlight simulator). The proposed methodology showed good performance compared to other methods, such as multilayer perceptron, type-1 FLS, type-2 FLS, and interval type-3 FLS, despite the lack of a more elaborate discussion of the presented results. Furthermore, transparency is not guaranteed regarding the modeling steps and the influence of various parameters optimized during learning. The authors in [187] proposed a self-organizing FNN with incremental deep pre-training, abbreviated as IDPT-SOFNN, to promote efficient feature extraction and dynamic adaptation in the structure according to error-reduction rate. IDPT-SOFNN was

implemented for the prediction of Mackey–Glass time series, total phosphorus concentration in wastewater treatment plant, and air pollutant concentration. In [188], a deep fuzzy cognitive map was proposed for multivariable time-series forecasting applications, such as air quality indexes, traffic speed of six road segments in China, and two benchmark datasets from the UCI repository, the electric power consumption and the temperature of a monitor system. The analysis of the model’s interpretability was based on nonlinear and nonmonotonic influences of unknown exogenous factors. The work in [189] proposed a deep FNN composed of an input layer, four hidden layers (membership functions, T-norm operation, linear regression, and aggregation), and an output layer, designed exclusively for intra- and inter-fractional variational prediction for multiple patients’ breathing motion.

Table 1 summarizes the literature on Standard DFS. They are categorized according to the application domain and regarding its interpretability categories. The survey shows that the application domain is vast, with applications in the domain of industrial systems, power systems, traffic systems, and multivariable benchmark systems.

Table 1 shows that only four out of eleven works discuss interpretability, despite most of their proposed methods being transparent. Also, not all methods provide post hoc

**Table 1** State-of-the-art on methods with standard deep fuzzy systems for regression problems. XAI: explainable artificial intelligence; Disc.: discussion by the authors (Yes/No); Und.: how understandable is the model, whether it is transparent (T) or opaque (O); Scope: in post hoc scope, if the model promotes local explanations (L), global explanations (G) or visual explanations (V)

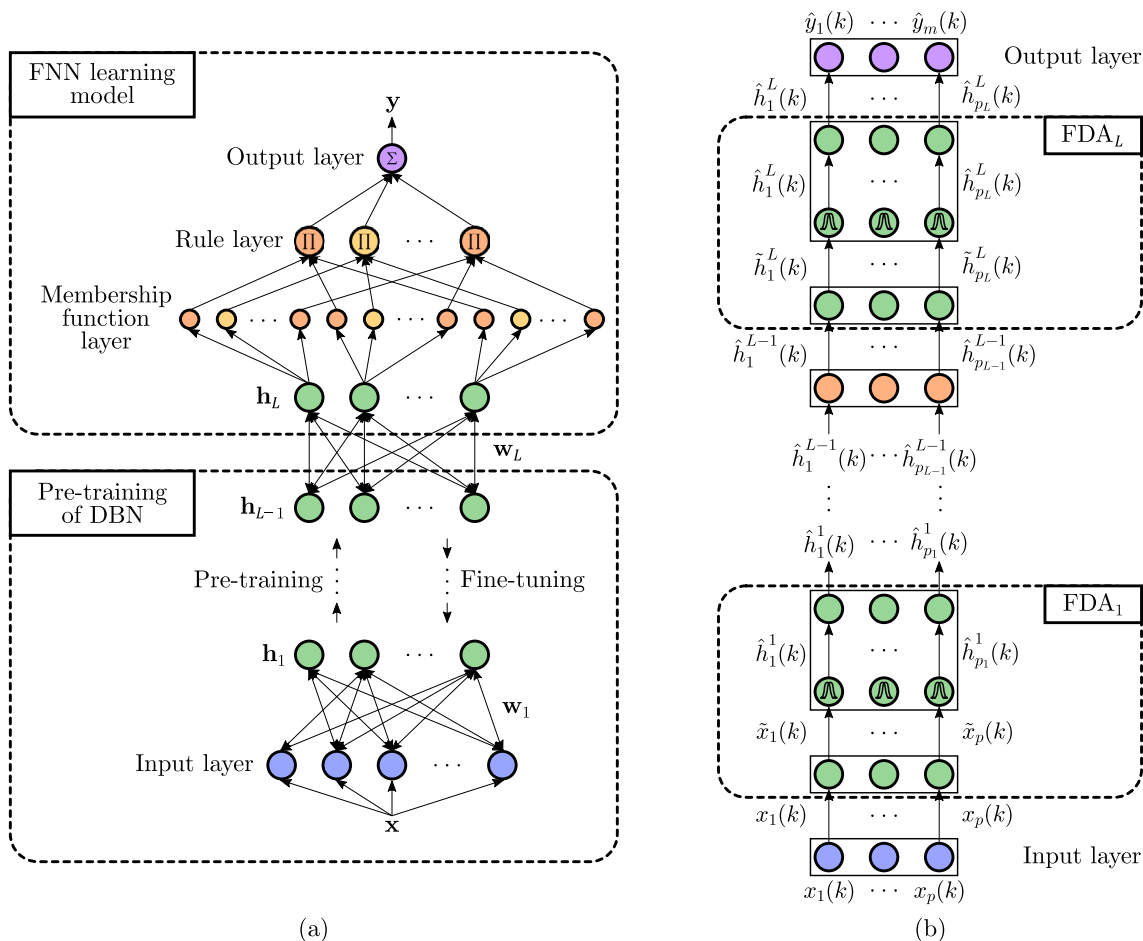
Reference	Approach	Application Domain	XAI		
			Disc.	Und.	Scope
[178]	Hierarchical Randomly Locally Optimized Deep Fuzzy System with input sharing	Prediction of 12 real-world datasets from the UCI repository	Yes	T	L
[180]	Double-input rule modules stacked deep interval type-2 fuzzy model	Time-series forecasting using subway passenger data and traffic flow data	Yes	T	L
[181]	Double-input-rule-modules stacked deep fuzzy model	Prediction of short-term photovoltaic power generation	Yes	T	L,G
[182]	Dynamic fractional-order deep learned type-2 fuzzy logic system	Prediction of two chaotic benchmark systems, glucose level of type-1 diabetes patients, and dataset of a heat transfer system	No	O	–
[183]	Evolving recurrent interval type-2 intuitionistic FNN	Time-series forecasting using KEEL datasets, Mackey–Glass time series, and a simulated second-order time-varying system	No	T	–
[184]	Deep Stacking Convex Neuro-Fuzzy System	Real ecological time-series forecasting (Darwin sea level pressure)	No	T	L
[185]	Hybrid cascade neuro-fuzzy scheme (ensemble of extended neuro-fuzzy neurons)	Time-series forecasting (electrical loads)	No	T	–
[186]	Deep learned recurrent type-3 fuzzy system	Renewable energy modeling (power generation of a 660kW wind turbine and solar radiation generated by sunlight simulator)	No	O	–
[187]	Self-Organizing Deep FNN	Prediction of Mackey–Glass time series, total phosphorus concentration in wastewater treatment plant, and air pollutant concentration	No	T	V
[188]	Deep Fuzzy cognition map model (with an alternate function gradient descent algorithm)	Multivariate time-series forecasting applications (air quality indexes, traffic speed, UCI datasets)	Yes	T	V
[189]	Fuzzy Deep Learning architecture (with four hidden layers)	Intra- and Inter-fractional variation prediction of Lung Tumors	No	T	L,V

explanations, and of these, the local scope is more frequent. The Standard DFS architecture is easy to implement, with flexibility in the construction of fuzzy rules, having a similar structure to feedforward neural networks. However, these methods may suffer from loss of interpretability when using a non-intuitive hierarchical structure, the lack of investigation of interconnections between the variables involved, and the limitation of fuzzy rules that can disturb the estimation by not covering all operating regions (coverage of input data space) [190].

#### 4.1.2 Hybrid Deep Fuzzy Systems

The Hybrid DFSs are discussed in this section. The selection and order of works follow the same principles used for Standard DFS. The common DL architectures used in combination with fuzzy systems are the Deep Belief Networks, Autoencoders, Long Short-Term Memory networks, and Echo State Networks.

In [191], it is proposed a sparse Deep Belief Network (SDBN) with FNN for nonlinear system modeling (benchmark) and total phosphorus prediction in wastewater treatment plant. The SDBN is considered for unsupervised learning and pre-training to perform fast weight-initialization and improve modeling robustness. The FNN is used as supervised learning to reduce layer-by-layer complexity. As shown in Fig. 11, the structure of the DBN resembles the structure shown in Fig. 3, except for considering additional constraints (sparsity) used to penalize fluctuations of values along the hidden neurons. The proposed method performed better compared to other similar methods, such as transfer learning-based growing DBN [192], DBN-based echo-state network [164], and self-organizing cascade neural network [193]. However, the authors noticed various fluctuations in the assignment of hyperparameters, which can compromise the stability of the proposed model, making it necessary to dynamically and robustly improve its structure to these fluctuations. In terms



**Fig. 11** Examples of FLS application with conventional deep models: **a** with deep belief networks, adapted from [191]; and **b** with denoising autoencoders, adapted from [194]

of interpretability, the proposed model guarantees a moderate number of membership functions and rules, allowing good consistency and readability of what happens within the FNN structure. The same cannot be said for the DBN structure, which is not intuitive in sparse representation to decide which features are more valuable than others.

The authors in [194] proposed a novel robust deep neural network (RDNN) for regression problems involving nonlinear systems, with the implementation of three strategies: a fuzzy denoising autoencoder (FDA) as a base-building unit for RDNN, improving the ability to represent uncertainties; a compact parameter strategy (CPS), designed to reconstruct the parameters of the FDA, reducing unnecessary learning parameters; and an adaptive backpropagation (ABP) algorithm to update RDNN parameters with fast convergence. As shown in Fig. 11, FDA has an input layer, a fuzzy layer (built by Gaussian membership functions), and an output layer, which can be divided into an encoder (“input” to “fuzzy”) and a decoder (“fuzzy” to “output”). Furthermore, at each FDA, due to the nature of this autoencoder, the data is partially

corrupted by noises (represented with a “tilde” symbol) and is reconstructed using CPS (represented with a “hat” symbol) with associated parameters (*e.g.*, encoder/decoder weights and biases) are adjusted via ABP. The resulting model was evaluated through four prediction examples: air quality from the UCI repository, wind speed from National Renewable Energy Laboratory, housing price from the KEEL repository, and water quality from a wastewater treatment plant in Beijing, China. Regarding the interpretability of the proposed model, there was no appropriate discussion by the authors, as they focused mainly on model performance/accuracy in the presence of uncertainties. In addition to a fuzzy rule base not being defined, the number of fuzzy neurons in each FDA is manually initialized and adapts through redundancies during the reconstruction of parameters via CPS. Finally, the architecture of the proposed model is not intuitive based on the mechanisms related to the representation ability of each FDA, which would be crucial to determine the influence of input variables and learning parameters on the outputs.

**Table 2** State-of-the-art on hybrid methods using fuzzy logic systems and conventional deep models for regression problems. XAI: explainable artificial intelligence; Disc.: discussion by the authors (Yes/No); Und.: how understandable is the model, whether it is transparent (T) or opaque (O); Scope: in post hoc scope, if the model promotes local explanations (L), global explanations (G) or visual explanations (V)

Reference	Approach	Application Domain	XAI		
			Disc.	Und.	Scope
[191]	Sparse Deep Belief Network with FNN	Nonlinear system modeling (benchmark) and total phosphorus prediction in wastewater treatment plant	No	T	–
[194]	Fuzzy denoising autoencoder with a compact parameter strategy and an adaptive backpropagation algorithm	Air quality, wind speed, housing price, and water quality from a wastewater treatment plant	No	O	–
[195]	Chronological Dolphin Echolocation-Fuzzy and Deep Belief Network	Traffic flow prediction in intelligent transportation system (traffic-major roads and PEMS-SF)	No	O	–
[197]	Fuzzy information granulation with Stacked Autoencoder	Traffic flow prediction	No	T	–
[198]	Fuzzy seasonal long short-term memory network	Seasonal time-series forecasting in a monthly wind power output dataset	No	O	–
[200]	Long Short-Term Memory model hybridized with rough and fuzzy set theory	Short-term wind speed forecasting	No	O	–
[202]	T-S fuzzy model with fuzzy c-means and Deep Belief Network	Photovoltaic power forecasting (with a 433 kW photovoltaic matrix database)	No	T	–
[203]	Deep Type-2 Fuzzy Logic System, trained as Stacked Autoencoder	Benchmark system prediction and binary classification	Yes	T	G,V
[204]	Differential Evolution based Fuzzy Echo State Network	Time-series forecasting (Mackey-Glass, NARMA time series with chaotic behavior, and Lorenz attractor)	No	O	–
[205]	T-S fuzzy model with Echo State Network	Benchmark system prediction (approximation of a nonlinear function, Henon chaotic system, and dynamic system with and without noise signal)	No	T	–
[206]	Fuzzy Cognitive Maps with Sparse Autoencoders	Time-series forecasting (ten datasets, including sunspots, Mackey-Glass, S&P 500 index, and Dow-Jones index)	No	T	–
[207]	Fuzzy rule reduction with Stacked Sparse Autoencoders	Time-series prediction (Mackey-Glass) and classification (Iris dataset)	No	O	–

Some hybrid methods using FLS and conventional deep models have been applied in the literature for traffic flow prediction. In [195], the authors developed an algorithm based on Dolphin Echolocation optimization [196], where the input features are fuzzified into membership functions to obtain chronological data, whose integration goes into the weight update process of the proposed algorithm converging to a globally optimal solution with a Deep Belief Network. The method in [195] was evaluated using datasets of traffic-major roads in Great Britain and PEMS-SF (San Francisco bay area freeways). The study in [197] combined fuzzy information granulation and a deep neural network to represent the temporal-spatial correlation of mass traffic data and be able to adapt to noisy data. A Stacked Autoencoder is used to obtain the prediction results based on processed granules that have a good capacity for interpretation, which have not been discussed by the authors. The method in [197] was evaluated using traffic flow data archived for the Portland–Vancouver Metropolitan region.

Other methods were implemented for energy forecasting, with the usual application of LSTM networks. A novel fuzzy seasonal LSTM was proposed in [198], where a fuzzy seasonality index [199] and a decomposition method were employed to solve the seasonal time-series problem in a monthly wind power output dataset from the National Development Council in Taiwan. In [200], the authors used an LSTM network with rough set theory [201] and interval type-2 fuzzy sets for short-term wind speed forecasting (dataset from Bandar-Abbas City, Iran), with the aid of mutual information approach for efficient variable input selection. A novel ultra-short-term photovoltaic power forecasting method was proposed in [202], where a T-S fuzzy model comprises a fuzzy c-means clustering algorithm and DBNs, with evaluation tests using a 433 kW photovoltaic matrix database. The studies in [198, 200] and [202] did not discuss the interpretability of their models, which have a structure with an ensemble characteristic whose fuzzy part has an affinity for good coverage of the input space and good capture of data uncertainty.

The work in [203] proposed a deep type-2 FLS (D2FLS) architecture with greedy layer-wise training for high-dimensional input data. The D2FLS model was applied to two regression datasets (prediction of the performance of British Telecom's work area and health insurance premium) and two binary classification datasets (Santander Customer Transaction prediction and British Telecom's customer service). The authors showed how to extract interpretable explanations related to the contribution of fuzzy rules to the final prediction developed for a two-layer D2FLS. However, the authors opted for a large number of fuzzy rules (100, in this case) that impair interpretability by increasing the complexity of the model. The authors in [204] proposed an ensemble model composed of an Echo State Network (ESN), a T-S fuzzy model, and differential evolution for time-series forecasting problems: Mackey-Glass time series, nonlinear auto-regressive moving average (NARMA) time series, and Lorenz attractor. The differential evolution method, used to optimize the weight coefficients of the model, managed to reduce the number of fuzzy rules. The interpretability of the model can be impaired due to the structure with an ensemble characteristic, which does not provide enough transparency for learning. A method based on the T-S fuzzy model and ESN was developed in [205] and tested with three benchmark examples: approximation of a nonlinear function, prediction of Henon chaotic system, and identification of a dynamic system with and without noise signal. The authors chose to balance model complexity and performance based on the parameters involved for learning (*e.g.*, number of fuzzy rules and reservoir size), which resulted in better results in comparison with other methods (*e.g.*, traditional ESN and hybrid fuzzy ESN), despite the lack of discussion on interpretability.

In [206], a framework based on a sparse autoencoder (SpAE) and a high-order fuzzy cognitive map (HFCM) is proposed for time-series forecasting problems (*e.g.*, sunspots, Mackey-Glass, S&P 500 stock index and Dow-Jones industrial index). SpAE is used to extract features from the original data, and these features are via HFCM. Another study that uses SpAE with FLS was proposed in [207] and implemented in Mackey-Glass time series and Iris dataset (classification). The proposed method uses a method to reduce fuzzy rules by reducing the data dimensionality with SpAE. Both studies in [206] and [207] do not consider addressing the interpretability of the proposed models, which present good directions in data partitioning and construction of fuzzy rules but are not intuitive in feature extraction with sparse representation.

Table 2 summarizes the works presented in this section, in addition to evaluating them within the context of explainable artificial intelligence (XAI) systems.

Only two works discuss the interpretability in Hybrid DFS, half of the discussed works are opaque, and only one presents post hoc explanations. This fact occurs since the combination with DNN brings a new layer of black-box to the system. However, the Hybrid DFS methods categorized as transparent show that the fuzzy component can promote good interpretability with efficient input–output mappings and the construction of rules to cover the universe of discourse for a given system. Inherently interpretable methodologies are difficult to achieve only with an ensemble of multiple methods, as seen in recent studies, in addition to the challenges of reducing the time complexity of systems, which is slightly mitigated by the reduction of fuzzy rules [208].

## 5 Conclusion

This study surveyed the literature on deep fuzzy systems (DFSs) for regression applications with an emphasis on interpretability. For the survey, the DFSs were categorized as (i) Standard DFS and (ii) Hybrid DFS. Regarding the interpretability, each method was categorized as to whether it is transparent or not and whether it has post hoc explanations (under the definition of [22]). Standard DFSs have been shown to promote more interpretability of their models when compared to Hybrid DFSs, according to the survey. Indeed, Standard DFSs are based on fundamental fuzzy logic systems, which are inherently interpretable, whereas Hybrid DFSs include conventional deep learning (DL) methods, which lack flexibility in promoting interpretability. In terms of applications, DFS can be flexible in its implementation either by simulation or in real-time systems, whose most recurrent applications involve time-series forecastings, such as traffic flow and energy modeling (*e.g.*, photovoltaic and wind). Furthermore, the DFS is frequently referred to as interpretable by default, but only 5 of the 23 works surveyed here actually addressed this issue. The remaining works had a common goal: to improve prediction accuracy using their proposed methods. However, this survey presented the potential of using Standard DFS as a base for developing accurate models while promoting interpretability since hybrid models are not straightforward.

**Acknowledgements** This work has been supported by the ERDF and national funds through the project iProMo (CENTRO-01-0247-FEDER-069730). Jorge S. S. Júnior is supported by Fundação para a Ciência e a Tecnologia (FCT) under the grant ref. 2021.04917.BD.

**Funding** Open access funding provided by FCTIFCCN (b-on).

**Declarations**

**Competing Interest** The authors declare no competing interests.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Busu, M., Trica, C.L.: Sustainability of circular economy indicators and their impact on economic growth of the European Union. *Sustainability* (2019). <https://doi.org/10.3390/su11195481>
- Botev, J., Égert, B., Jawadi, F.: The nonlinear relationship between economic growth and financial development: evidence from developing, emerging and advanced economies. *Int. Econ.* **160**, 3–13 (2019). <https://doi.org/10.1016/j.inteco.2019.06.004>
- Liu, B., Zhao, Q., Jin, Y., Shen, J., Li, C.: Application of combined model of stepwise regression analysis and artificial neural network in data calibration of miniature air quality detector. *Sci. Rep.* **11**(1), 1–12 (2021). <https://doi.org/10.1038/s41598-021-82871-4>
- Gu, K., Qiao, J., Lin, W.: Recurrent air quality predictor based on meteorology- and pollution-related factors. *IEEE Trans. Ind. Informatics* **14**(9), 3946–3955 (2018). <https://doi.org/10.1109/THI.2018.2793950>
- Zhu, J., et al.: Prevalence and influencing factors of anxiety and depression symptoms in the first-line medical staff fighting against COVID-19 in Gansu. *Front. Psychiatr.* **11**, 386 (2020). <https://doi.org/10.3389/fpsy.2020.00386>
- Shi, B., et al.: Nonlinear heart rate variability biomarkers for gastric cancer severity: a pilot study. *Sci. Rep.* **9**(1), 1–9 (2019). <https://doi.org/10.1038/s41598-019-50358-y>
- Orlandi, M., Escudero-Casao, M., Licini, G.: Nucleophilicity prediction via multivariate linear regression analysis. *J. Org. Chem.* **86**(4), 3555–3564 (2021). <https://doi.org/10.1021/acs.joc.0c02952>
- Yang, Y., et al.: A support vector regression model to predict nitrate-nitrogen isotopic composition using hydro-chemical variables. *J. Environ. Manag.* **290**, 112674 (2021). <https://doi.org/10.1016/j.jenvman.2021.112674>
- Souza, F., Mendes, J., Araújo, R.: A regularized mixture of linear experts for quality prediction in multimode and multiphase industrial processes. *Appl. Sci.* (2021). <https://doi.org/10.3390/app11052040>
- Liu, H., Yang, C., Carlsson, B., Qin, S.J., Yoo, C.: Dynamic nonlinear partial least squares modeling using gaussian process regression. *Ind. Eng. Chem. Res.* **58**(36), 16676–16686 (2019). <https://doi.org/10.1021/acs.iecr.9b00701>
- Sun, Q., Ge, Z.: A survey on deep learning for data-driven soft sensors. *IEEE Trans. Ind. Informatics* **17**(9), 5853–5866 (2021). <https://doi.org/10.1109/THI.2021.3053128>
- Angelov, P., Soares, E.: Towards explainable deep neural networks (xDNN). *Neural Netw.* **130**, 185–194 (2020). <https://doi.org/10.1016/j.neunet.2020.07.010>
- Sjöberg, J., et al.: Nonlinear black-box modeling in system identification: a unified overview. *Automatica* **31**(12), 1691–1724 (1995). [https://doi.org/10.1016/0005-1098\(95\)00120-8](https://doi.org/10.1016/0005-1098(95)00120-8)
- Han, Z., Zhao, J., Leung, H., Ma, K.F., Wang, W.: A review of deep learning models for time series prediction. *IEEE Sens. J.* **21**(6), 7833–7848 (2021). <https://doi.org/10.1109/JSEN.2019.2923982>
- Torres, J.F., Hadjout, D., Sebaa, A., Martínez-Álvarez, F., Troncoso, A.: Deep learning for time series forecasting: a survey. *Big Data* **9**(1), 3–21 (2021). <https://doi.org/10.1089/big.2020.0159>
- Pang, G., Shen, C., Cao, L., Hengel, A.V.D.: Deep learning for anomaly detection: a review. *ACM Comput. Surv.* (2021). <https://doi.org/10.1145/3439950>
- Chalapathy, R., Chawla, S.: Deep learning for anomaly detection: A survey. arXiv preprint [arXiv:1901.03407](https://arxiv.org/abs/1901.03407) (2019). <https://doi.org/10.48550/arXiv.1901.03407>
- Dong, S., Wang, P., Abbas, K.: A survey on deep learning and its applications. *Comput. Sci. Rev.* **40**, 100379 (2021). <https://doi.org/10.1016/j.cosrev.2021.100379>
- Sengupta, S., et al.: A review of deep learning with special emphasis on architectures, applications and recent trends. *Knowl.-Based Syst.* **194**, 105596 (2020). <https://doi.org/10.1016/j.knsys.2020.105596>
- Pouyanfar, S., et al.: A survey on deep learning: algorithms, techniques, and applications. *ACM Comput. Surv.* (2018). <https://doi.org/10.1145/3234150>
- Navamani, T.: Chapter 7-Efficient deep learning approaches for health informatics. In: Sangaiah, A.K. (ed.) *Deep learning and parallel computing environment for bioengineering systems*, pp. 123–137. Academic Press, Cambridge (2019). <https://doi.org/10.1016/B978-0-12-816718-2.00014-2>
- Angelov, P.P., Soares, E.A., Jiang, R., Arnold, N.I., Atkinson, P.M.: Explainable artificial intelligence: an analytical review. *WIREs Data Mining Knowl. Discov.* (2021). <https://doi.org/10.1002/widm.1424>
- Confalonieri, R., Coba, L., Wagner, B., Besold, T.R.: A historical perspective of explainable artificial intelligence. *Wiley Interdiscip. Rev.: Data Mining Knowl. Discov.* (2021). <https://doi.org/10.1002/widm.1391>
- Phillips, P. J. et al.: Four Principles of Explainable Artificial Intelligence (National Institute of Standards and Technology, 2021). <https://doi.org/10.6028/nist.ir.8312>
- Chimatapu, R., Hagra, H., Starkey, A., Owusu, G.: Explainable AI and Fuzzy logic systems. In: Fagan, D., Martín-Vide, C., O'Neill, M., Vega-Rodríguez, M.A. (eds) *Theory and Practice of Natural Computing*, pp. 3–20 (2018). [https://doi.org/10.1007/978-3-030-04070-3\\_1](https://doi.org/10.1007/978-3-030-04070-3_1)
- Łapa, K., Cpałka, K., Rutkowski, L.: New aspects of interpretability of fuzzy systems for nonlinear modeling. In: Gawęda, A., Kacprzyk, J., Rutkowski, L., Yen, G. (eds) *Advances in Data Analysis with Computational Intelligence Methods. Studies in Computational Intelligence*, vol. 738. Springer, Cham. (2017). [https://doi.org/10.1007/978-3-319-67946-4\\_9](https://doi.org/10.1007/978-3-319-67946-4_9)
- Mendes, J., Maia, R., Araújo, R., Souza, F.A.A.: Self-evolving fuzzy controller composed of univariate fuzzy control rules. *Appl. Sci.* **10**(17), 5836 (2020). <https://doi.org/10.3390/app10175836>
- Moral, J.M.A., Castiello, C., Magdalena, L., Mencar, C.: Explainable fuzzy systems. Springer International Publishing, New York (2021). <https://doi.org/10.1007/978-3-030-71098-9>
- Das, R., Sen, S., Maulik, U.: A survey on fuzzy deep neural networks. *ACM Comput. Surv.* (2020). <https://doi.org/10.1145/3369798>

30. Zheng, Y., Xu, Z., Wang, X.: The fusion of deep learning and fuzzy systems: a state-of-the-art survey. *IEEE Trans. Fuzzy Syst.* (2021). <https://doi.org/10.1109/TFUZZ.2021.3062899>
31. Zadeh, L.A.: Fuzzy sets. *Information Control* **8**(3), 338–353 (1965). [https://doi.org/10.1016/s0019-9958\(65\)90241-x](https://doi.org/10.1016/s0019-9958(65)90241-x)
32. Wang, L.-X.: *A course in fuzzy systems and control*. Prentice-Hall Inc., Hoboken (1997)
33. Mendes, J., Araújo, R., Sousa, P., Apóstolo, F., Alves, L.: An architecture for adaptive fuzzy control in industrial environments. *Comput. Ind.* **62**(3), 364–373 (2011). <https://doi.org/10.1016/j.compind.2010.11.001>
34. Mamdani, E., Assilian, S.: An experiment in linguistic synthesis with a fuzzy logic controller. *Int. J. Man-Mach. Stud.* **7**(1), 1–13 (1975). [https://doi.org/10.1016/S0020-7373\(75\)80002-2](https://doi.org/10.1016/S0020-7373(75)80002-2)
35. Takagi, T., Sugeno, M.: Fuzzy identification of systems and its applications to modeling and control. *IEEE Trans. Syst. Man Cybern.* **SMC-15**(1), 116–132 (1985). <https://doi.org/10.1109/TSMC.1985.6313399>
36. Angelov, P., Yager, R.: Simplified fuzzy rule-based systems using non-parametric antecedents and relative data density. In: *IEEE Workshop on Evolving and Adaptive Intelligent Systems (EAIS)*, pp. 62–69. Paris, France (2011). <https://doi.org/10.1109/EAIS.2011.5945926>
37. Qiu, J., Gao, H., Ding, S.X.: Recent advances on fuzzy-model-based nonlinear networked control systems: a survey. *IEEE Trans. Ind. Electron.* **63**(2), 1207–1217 (2016). <https://doi.org/10.1109/TIE.2015.2504351>
38. Ying, H.: General MISO Takagi-Sugeno fuzzy systems with simplified linear rule consequent as universal approximators for control and modeling applications. In: *IEEE International Conference on Systems, Man, and Cybernetics. Computational Cybernetics and Simulation*, Vol. 2, pp. 1335–1340 (1997). <https://doi.org/10.1109/ICSMC.1997.638158>
39. Júnior, J. S. S., Mendes, J., Araújo, R., Paulo, J. R., Premebida, C.: Novelty detection for iterative learning of MIMO fuzzy systems. In: *2021 IEEE 19th International Conference on Industrial Informatics (INDIN)*, pp. 1–7 (2021). <https://doi.org/10.1109/INDIN45523.2021.9557354>
40. Hall, P., Gill, N.: *An introduction to machine learning interpretability*. O'Reilly Media, Incorporated (2019)
41. Shortliffe, E.H., Axline, S.G., Buchanan, B.G., Merigan, T.C., Cohen, S.N.: An artificial intelligence program to advise physicians regarding antimicrobial therapy. *Comput. Biomed. Res.* **6**(6), 544–560 (1973). [https://doi.org/10.1016/0010-4809\(73\)90029-3](https://doi.org/10.1016/0010-4809(73)90029-3)
42. Clancey, W.J.: Tutoring rules for guiding a case method dialogue. *Int. J. Man-Mach. Stud.* **11**(1), 25–49 (1979). [https://doi.org/10.1016/S0020-7373\(79\)80004-8](https://doi.org/10.1016/S0020-7373(79)80004-8)
43. Weiss, S.M., Kulikowski, C.A., Amarel, S., Safir, A.: A model-based method for computer-aided medical decision-making. *Artif. Intell.* **11**(1), 145–172 (1978). [https://doi.org/10.1016/0004-3702\(78\)90015-2](https://doi.org/10.1016/0004-3702(78)90015-2)
44. Suwa, M., Scott, A.C., Shortliffe, E.H.: An approach to verifying completeness and consistency in a rule-based expert system. *Ai Mag.* **3**(4), 16–16 (1982). <https://doi.org/10.1609/aimag.v3i4.377>
45. Swartout, W.R.: XPLAIN: a system for creating and explaining expert consulting programs. *Artif. Intell.* **21**(3), 285–325 (1983). [https://doi.org/10.1016/S0004-3702\(83\)80014-9](https://doi.org/10.1016/S0004-3702(83)80014-9)
46. Swartout, W. R.: Explaining and justifying expert consulting programs. In: Reggia, J.A., Tuhim, S. (eds) *Computer-assisted medical decision making*, pp. 254–271 (1985). [https://doi.org/10.1007/978-1-4612-5108-8\\_15](https://doi.org/10.1007/978-1-4612-5108-8_15)
47. Rudin, C.: Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat. Mach. Intell.* **1**(5), 206–215 (2019). <https://doi.org/10.1038/s42256-019-0048-x>
48. Bengio, Y., Courville, A., Vincent, P.: Representation learning: a review and new perspectives. *IEEE Trans. Pattern Anal. Mach. Intell.* **35**(8), 1798–1828 (2013). <https://doi.org/10.1109/TPAMI.2013.50>
49. Goodfellow, I., Bengio, Y., Courville, A., Bengio, Y.: *Deep learning*. MIT press, Cambridge (2016)
50. Fukushima, K.: Neocognitron: a self organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biol. Cybern.* **36**(4), 193–202 (1980). <https://doi.org/10.1007/bf00344251>
51. Hubel, D.H., Wiesel, T.N.: Receptive fields and functional architecture of monkey striate cortex. *J. Physiol.* **195**(1), 215–243 (1968). <https://doi.org/10.1113/jphysiol.1968.sp008455>
52. Gu, J., et al.: Recent advances in convolutional neural networks. *Pattern Recognit.* **77**, 354–377 (2018). <https://doi.org/10.1016/j.patcog.2017.10.013>
53. Li, Z., Liu, F., Yang, W., Peng, S., Zhou, J.: A survey of convolutional neural networks: analysis, applications, and prospects. *IEEE Trans. Neural Netw. Learn. Syst.* (2021). <https://doi.org/10.1109/TNNLS.2021.3084827>
54. Ketkar, N.: Convolutional neural networks. In: *Deep Learning with Python: A Hands-on Introduction* 63–78 (2017). [https://doi.org/10.1007/978-1-4842-2766-4\\_5](https://doi.org/10.1007/978-1-4842-2766-4_5)
55. Mallat, S.: Understanding deep convolutional networks. *Philos. Trans. Royal Soc.: Mathemat. Phys. Eng. Sci.* **374**(2065), 20150203 (2016). <https://doi.org/10.1098/rsta.2015.0203>
56. Wu, S.: Spatiotemporal dynamic forecasting and analysis of regional traffic flow in urban road networks using deep learning convolutional neural network. *IEEE Trans. Intell. Trans. Syst.* **23**(2), 1607–1615 (2022). <https://doi.org/10.1109/TITS.2021.3098461>
57. Zhang, Y., Zhou, Y., Lu, H., Fujita, H.: Traffic network flow prediction using parallel training for deep convolutional neural networks on spark cloud. *IEEE Trans. Ind. Informatics* **16**(12), 7369–7380 (2020). <https://doi.org/10.1109/TII.2020.2976053>
58. Mukhtar, M., et al.: Development and comparison of two novel hybrid neural network models for hourly solar radiation prediction. *Appl. Sci.* (2022). <https://doi.org/10.3390/app12031435>
59. Heo, J., Song, K., Han, S., Lee, D.-E.: Multi-channel convolutional neural network for integration of meteorological and geographical features in solar power forecasting. *Appl. Energy* **295**, 117083 (2021). <https://doi.org/10.1016/j.apenergy.2021.117083>
60. Liu, T., et al.: Enhancing wind turbine power forecast via convolutional neural network. *Electronics* (2021). <https://doi.org/10.3390/electronics10030261>
61. Wan, R., Mei, S., Wang, J., Liu, M., Yang, F.: Multivariate temporal convolutional network: a deep neural networks approach for multivariate time series forecasting. *Electronics* (2019). <https://doi.org/10.3390/electronics8080876>
62. Gao, Z., et al.: Multitask-based temporal-channelwise CNN for parameter prediction of two-phase flows. *IEEE Trans. Ind. Informatics* **17**(9), 6329–6336 (2021). <https://doi.org/10.1109/TII.2020.2978944>
63. Fan, W., Zhang, Z.: A CNN-SVR hybrid prediction model for wastewater index measurement. In: *2020 2nd International Conference on Advances in Computer Technology, Information Science and Communications (CTISC)*, pp. 90–94 (2020). <https://doi.org/10.1109/CTISC49998.2020.00022>
64. Yuan, X., et al.: Soft sensor model for dynamic processes based on multichannel convolutional neural network. *Chemometr. Intell. Lab. Syst.* **203**, 104050 (2020). <https://doi.org/10.1016/j.chemolab.2020.104050>



65. Jalali, S.M.J., et al.: A novel evolutionary-based deep convolutional neural network model for intelligent load forecasting. *IEEE Trans. Ind. Informatics* **17**(12), 8243–8253 (2021). <https://doi.org/10.1109/TII.2021.3065718>
66. Eskandari, H., Imani, M., Moghaddam, M.P.: Convolutional and recurrent neural network based model for short-term load forecasting. *Electr. Power Syst. Res.* **195**, 107173 (2021). <https://doi.org/10.1016/j.epsr.2021.107173>
67. Zahid, M., et al.: Electricity price and load forecasting using enhanced convolutional neural network and enhanced support vector regression in smart grids. *Electronics* (2019). <https://doi.org/10.3390/electronics8020122>
68. Koprinska, I., Wu, D., Wang, Z.: Convolutional neural networks for energy time series forecasting. In: 2018 International Joint Conference on Neural Networks (IJCNN), pp. 1–8 (2018). <https://doi.org/10.1109/IJCNN.2018.8489399>
69. Tian, C., Ma, J., Zhang, C., Zhan, P.: A deep neural network model for short-term load forecast based on long short-term memory network and convolutional neural network. *Energies* (2018). <https://doi.org/10.3390/en11123493>
70. Gao, P., Zhang, J., Sun, Y., Yu, J.: Accurate predictions of aqueous solubility of drug molecules via the multilevel graph convolutional network (MGCN) and SchNet architectures. *Phys. Chem. Chem. Phys.* **22**(41), 23766–23772 (2020). <https://doi.org/10.1039/D0CP03596C>
71. Wu, K., Wei, G.-W.: Comparison of multi-task convolutional neural network (MT-CNN) and a few other methods for toxicity prediction. *arXiv Preprint* (2017). <https://doi.org/10.48550/arxiv.1703.10951>
72. Witten, I. H., Frank, E., Hall, M. A., Pal, C. J.: Chapter 10 - Deep learning. In: Witten, I. H., Frank, E., Hall, M. A., Pal, C. J. (eds) *Data Mining (Fourth Edition)*, pp. 417–466. Morgan Kaufmann (2017). <https://doi.org/10.1016/B978-0-12-804291-5.00010-6>
73. Hinton, G.E., Osindero, S., Teh, Y.-W.: A fast learning algorithm for deep belief nets. *Neural Comput.* **18**(7), 1527–1554 (2006). <https://doi.org/10.1162/neco.2006.18.7.1527>
74. Wang, Y., et al.: An ensemble deep belief network model based on random subspace for NOx concentration prediction. *ACS Omega* **6**(11), 7655–7668 (2021). <https://doi.org/10.1021/acsomega.0c06317>
75. Hao, X., et al.: Prediction of nitrogen oxide emission concentration in cement production process: a method of deep belief network with clustering and time series. *Environ. Sci. Pollut Res.* **28**(24), 31689–31703 (2021). <https://doi.org/10.1007/s11356-021-12834-9>
76. Yuan, X., Gu, Y., Wang, Y.: Supervised deep belief network for quality prediction in industrial processes. *IEEE Trans. Instrum. Meas.* **70**, 1–11 (2021). <https://doi.org/10.1109/TIM.2020.3035464>
77. Yuan, X., et al.: FeO content prediction for an industrial sintering process based on supervised deep belief network. *IFAC-PapersOnLine* **53**(2), 11883–11888 (2020). <https://doi.org/10.1016/j.ifacol.2020.12.703>
78. Hao, X., et al.: Energy consumption prediction in cement calcination process: a method of deep belief network with sliding window. *Energy* **207**, 118256 (2020). <https://doi.org/10.1016/j.energy.2020.118256>
79. Zhu, S.-B., Li, Z.-L., Zhang, S.-M., Zhang, H.-F.: Deep belief network-based internal valve leakage rate prediction approach. *Measurement* **133**, 182–192 (2019). <https://doi.org/10.1016/j.measurement.2018.10.020>
80. Wang, G., Jia, Q.-S., Zhou, M., Bi, J., Qiao, J.: Soft-sensing of Wastewater Treatment Process via Deep Belief Network with Event-triggered Learning. *Neurocomputing* **436**, 103–113 (2021). <https://doi.org/10.1016/j.neucom.2020.12.108>
81. Lian, P., Liu, H., Wang, X., Guo, R.: Soft sensor based on DBN-IPSO-SVR approach for rotor thermal deformation prediction of rotary air-preheater. *Measurement* **165**, 108109 (2020). <https://doi.org/10.1016/j.measurement.2020.108109>
82. Tian, W., Liu, Z., Li, L., Zhang, S., Li, C.: Identification of abnormal conditions in high-dimensional chemical process based on feature selection and deep learning. *Chin. J. Chem. Eng.* **28**(7), 1875–1883 (2020). <https://doi.org/10.1016/j.cjche.2020.05.003>
83. Shao, H., Jiang, H., Li, X., Liang, T.: Rolling bearing fault detection using continuous deep belief network with locally linear embedding. *Comput. Ind.* **96**, 27–39 (2018). <https://doi.org/10.1016/j.compind.2018.01.005>
84. Xu, H., Jiang, C.: Deep belief network-based support vector regression method for traffic flow forecasting. *Neural Comput. Appl.* **32**(7), 2027–2036 (2019). <https://doi.org/10.1007/s00521-019-04339-x>
85. Jia, Y., Wu, J., Du, Y.: Traffic speed prediction using deep learning method. In: 2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC), pp. 1217–1222 (2016). <https://doi.org/10.1109/ITSC.2016.7795712>
86. Tian, J., Liu, Y., Zheng, W., Yin, L.: Smog prediction based on the deep belief - BP neural network model (DBN-BP). *Urban Climate* **41**, 101078 (2022). <https://doi.org/10.1016/j.uclim.2021.101078>
87. Xie, T., Zhang, G., Liu, H., Liu, F., Du, P.: A hybrid forecasting method for solar output power based on variational mode decomposition, deep belief networks and auto-regressive moving average. *Appl. Sci.* (2018). <https://doi.org/10.3390/app8101901>
88. Li, X., Yang, L., Xue, F., Zhou, H.: Time series prediction of stock price using deep belief networks with intrinsic plasticity. In: 2017 29th Chinese Control And Decision Conference (CCDC), pp. 1237–1242 (2017). <https://doi.org/10.1109/CCDC.2017.7978707>
89. Qiao, J., Wang, G., Li, W., Li, X.: A deep belief network with PLSR for nonlinear system modeling. *Neural Netw.* **104**, 68–79 (2018). <https://doi.org/10.1016/j.neunet.2017.10.006>
90. Qiu, X., Zhang, L., Ren, Y., Suganthan, P. N., Amaratunga, G.: Ensemble deep learning for regression and time series forecasting. In: 2014 IEEE Symposium on Computational Intelligence in Ensemble Learning (CIEL), pp. 1–6 (2014). <https://doi.org/10.1109/CIEL.2014.7015739>
91. Salakhutdinov, R., Larochelle, H. (ed.): Efficient learning of deep Boltzmann machines. (ed.) In: *Proceedings of the Thirtieth International Conference on Artificial Intelligence and Statistics*, pp. 693–700 (2010)
92. Schmidhuber, J.: Deep learning in neural networks: an overview. *Neural Netw.* **61**, 85–117 (2015). <https://doi.org/10.1016/j.neunet.2014.09.003>
93. Ballard, D. H. (ed.): Modular learning in neural networks. (ed.) In: *Proceedings of the Sixth National Conference on Artificial Intelligence - Volume 1*, Vol. 647, pp. 279–284 (1987)
94. Rumelhart, D. E., Hinton, G. E., Williams, R. J.: Learning Internal Representations by Error Propagation. Technical report, California Univ., San Diego, La Jolla. Inst. for Cognitive Science (1985). <https://doi.org/10.21236/ada164453>
95. Baldi, P. (ed.): Autoencoders, Unsupervised Learning, and Deep Architectures. (ed.) In: *Proceedings of ICML Workshop on Unsupervised and Transfer Learning*, Vol. 27, pp. 37–49 (2012)
96. Liu, G., Bao, H., Han, B.: A stacked autoencoder-based deep neural network for achieving gearbox fault diagnosis. *Mathematical Problems Eng.* **2018**, 1–10 (2018). <https://doi.org/10.1155/2018/5105709>
97. Sun, Q., Ge, Z.: Deep learning for industrial KPI prediction: when ensemble learning meets semi-supervised data. *IEEE*

- Trans. Ind. Informatics **17**(1), 260–269 (2021). <https://doi.org/10.1109/TII.2020.2969709>
98. Sun, Q., Ge, Z.: Gated stacked target-related autoencoder: a novel deep feature extraction and layerwise ensemble method for industrial soft sensor application. *IEEE Trans. Cybern.* (2020). <https://doi.org/10.1109/TCYB.2020.3010331>
  99. Liu, C., Wang, Y., Wang, K., Yuan, X.: Deep learning with nonlocal and local structure preserving stacked autoencoder for soft sensor in industrial processes. *Eng. Appl. Artif. Intell.* **104**, 104341 (2021). <https://doi.org/10.1016/j.engappai.2021.104341>
  100. Yuan, X., Ou, C., Wang, Y., Yang, C., Gui, W.: A novel semi-supervised pre-training strategy for deep networks and its application for quality variable prediction in industrial processes. *Chem. Eng. Sci.* **217**, 115509 (2020). <https://doi.org/10.1016/j.ces.2020.115509>
  101. Wang, Y., Liu, C., Yuan, X.: Stacked locality preserving autoencoder for feature extraction and its application for industrial process data modeling. *Chemometr. Intell. Lab. Syst.* **203**, 104086 (2020). <https://doi.org/10.1016/j.chemolab.2020.104086>
  102. Shi, C., et al.: Using multiple-feature-spaces-based deep learning for tool condition monitoring in ultraprecision manufacturing. *IEEE Trans. Ind Electron.* **66**(5), 3794–3803 (2019). <https://doi.org/10.1109/TIE.2018.2856193>
  103. Bose, T., Majumdar, A., Chattopadhyay, T.: Machine load estimation via stacked autoencoder regression. In: 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 2126–2130 (2018). <https://doi.org/10.1109/ICASSP.2018.8461576>
  104. Zhang, X., Zou, Y., Li, S.: Enhancing incremental deep learning for FCCU end-point quality prediction. *Information Sci.* **530**, 95–107 (2020). <https://doi.org/10.1016/j.ins.2020.04.013>
  105. Liu, C., Tang, L., Liu, J.: A stacked autoencoder with sparse Bayesian regression for end-point prediction problems in steel-making process. *IEEE Trans. Autom. Sci. Eng.* **17**(2), 550–561 (2020). <https://doi.org/10.1109/TASE.2019.2935314>
  106. Wang, X., Liu, H.: Soft sensor based on stacked auto-encoder deep neural network for air preheater rotor deformation prediction. *Adv. Eng. Informatics* **36**, 112–119 (2018). <https://doi.org/10.1016/j.aei.2018.03.003>
  107. Wei, M., Ye, M., Wang, Q., Twajamahoro, J.P.: Remaining useful life prediction of lithium-ion batteries based on stacked autoencoder and gaussian mixture regression. *J. Energy Storage* **47**, 103558 (2022). <https://doi.org/10.1016/j.est.2021.103558>
  108. Li, Z., Li, J., Wang, Y., Wang, K.: A deep learning approach for anomaly detection based on SAE and LSTM in mechanical equipment. *Int. J. Adv. Manuf. Technol.* **103**(1–4), 499–510 (2019). <https://doi.org/10.1007/s00170-019-03557-w>
  109. Ren, L., Sun, Y., Cui, J., Zhang, L.: Bearing remaining useful life prediction based on deep autoencoder and deep neural networks. *J. Manuf. Syst.* **48**, 71–77 (2018). <https://doi.org/10.1016/j.jmsy.2018.04.008>
  110. Jin, X.-B., Gong, W.-T., Kong, J.-L., Bai, Y.-T., Su, T.-L.: PFVAE: a planar flow-based variational auto-encoder prediction model for time series data. *Mathematics* (2022). <https://doi.org/10.3390/math10040610>
  111. Xiao, X., et al.: SSAE-MLP: stacked sparse autoencoders-based multi-layer perceptron for main bearing temperature prediction of large-scale wind turbines. *Concurr. Comput.: Practice Exp.* (2021). <https://doi.org/10.1002/cpe.6315>
  112. Jiao, R., Huang, X., Ma, X., Han, L., Tian, W.: A model combining stacked auto encoder and back propagation algorithm for short-term wind power forecasting. *IEEE Access* **6**, 17851–17858 (2018). <https://doi.org/10.1109/ACCESS.2018.2818108>
  113. Li, M., Xie, X., Zhang, D.: Improved deep learning model based on self-paced learning for multiscale short-term electricity load forecasting. *Sustainability* (2022). <https://doi.org/10.3390/su14010188>
  114. Lv, S.-X., Peng, L., Wang, L.: Stacked autoencoder with echo-state regression for tourism demand forecasting using search query data. *Appl. Soft Comput.* **73**, 119–133 (2018). <https://doi.org/10.1016/j.asoc.2018.08.024>
  115. Grossberg, S.: Recurrent neural networks. *Scholarpedia* **8**(2), 1888 (2013). <https://doi.org/10.4249/scholarpedia.1888>
  116. Hopfield, J.J.: Neural networks and physical systems with emergent collective computational abilities. *Proc. Nat. Acad. Sci.* **79**(8), 2554–2558 (1982). <https://doi.org/10.1073/pnas.79.8.2554>
  117. Hopfield, J.J.: Hopfield network. *Scholarpedia* **2**(5), 1977 (2007). <https://doi.org/10.4249/scholarpedia.1977>
  118. Rumelhart, D.E., Hinton, G.E., Williams, R.J.: Learning representations by back-propagating errors. *Nature* **323**(6088), 533–536 (1986). <https://doi.org/10.1038/323533a0>
  119. Zhang, S., Bamakan, S.M.H., Qu, Q., Li, S.: Learning for personalized medicine: a comprehensive review from a deep learning perspective. *IEEE Rev. Biomed. Eng.* **12**, 194–208 (2019). <https://doi.org/10.1109/RBME.2018.2864254>
  120. Albertini, F., Pra, P. D.: Recurrent neural networks: identification and other system theoretic properties. In: *Neural Network Systems Techniques and Applications*, Vol. 3 pp. 1–49 (1998). [https://doi.org/10.1016/s1874-5946\(98\)80003-5](https://doi.org/10.1016/s1874-5946(98)80003-5)
  121. Pascanu, R., Gulcehre, C., Cho, K., Bengio, Y.: How to Construct Deep Recurrent Neural Networks. *arXiv preprint arXiv:1312.6026* (2013). <https://doi.org/10.48550/ARXIV.1312.6026>
  122. Theodoridis, S.: Chapter 18 - Neural networks and deep learning. In: *Machine learning* (Second Edition) pp. 901–1038 (2020). <https://doi.org/10.1016/B978-0-12-818803-3.00030-1>
  123. Werbos, P.: Backpropagation through time: what it does and how to do it. *Proc. IEEE* **78**(10), 1550–1560 (1990). <https://doi.org/10.1109/5.58337>
  124. Lalapura, V.S., Amudha, J., Satheesh, H.S.: Recurrent neural networks for edge intelligence: a survey. *ACM Comput. Surv.* **54**(4), 1–38 (2021). <https://doi.org/10.1145/3448974>
  125. Hochreiter, S., Schmidhuber, J.: Long Short-Term Memory. *Neural Comput.* **9**(8), 1735–1780 (1997). <https://doi.org/10.1162/neco.1997.9.8.1735>
  126. Lu, S., Zhang, Q., Chen, G., Seng, D.: A combined method for short-term traffic flow prediction based on recurrent neural network. *Alexandria Eng. J.* **60**(1), 87–94 (2021). <https://doi.org/10.1016/j.aej.2020.06.008>
  127. Yang, B., Sun, S., Li, J., Lin, X., Tian, Y.: Traffic flow prediction using LSTM with feature enhancement. *Neurocomputing* **332**, 320–327 (2019). <https://doi.org/10.1016/j.neucom.2018.12.016>
  128. Roy, D.K., et al.: Daily prediction and multi-step forward forecasting of reference evapotranspiration using LSTM and Bi-LSTM models. *Agronomy* (2022). <https://doi.org/10.3390/agronomy12030594>
  129. Kumari, P., Toshniwal, D.: Long short term memory-convolutional neural network based deep hybrid approach for solar irradiance forecasting. *Appl. Energy* **295**, 117061 (2021). <https://doi.org/10.1016/j.apenergy.2021.117061>
  130. Altan, A., Karasu, S., Zio, E.: A new hybrid model for wind speed forecasting combining long short-term memory neural network, decomposition methods and grey wolf optimizer. *Appl. Soft Comput.* **100**, 106996 (2021). <https://doi.org/10.1016/j.asoc.2020.106996>
  131. Ma, J., et al.: Air quality prediction at new stations using spatially transferred bi-directional long short-term memory

- network. *Sci. Total Environ.* **705**, 135771 (2020). <https://doi.org/10.1016/j.scitotenv.2019.135771>
132. Shen, Z., Zhang, Y., Lu, J., Xu, J., Xiao, G.: A novel time series forecasting model with deep learning. *Neurocomputing* **396**, 302–313 (2020). <https://doi.org/10.1016/j.neucom.2018.12.084>
  133. Zhang, Q., Wang, H., Dong, J., Zhong, G., Sun, X.: Prediction of sea surface temperature using long short-term memory. *IEEE Geosci. Remote Sens. Lett.* **14**(10), 1745–1749 (2017). <https://doi.org/10.1109/LGRS.2017.2733548>
  134. Zhu, X., Hao, K., Xie, R., Huang, B.: Soft sensor based on extreme gradient boosting and bidirectional converted gates long short-term memory self-attention network. *Neurocomputing* **434**, 126–136 (2021). <https://doi.org/10.1016/j.neucom.2020.12.028>
  135. Yuan, X., Li, L., Shardt, Y.A.W., Wang, Y., Yang, C.: Deep learning with spatiotemporal attention-based lstm for industrial soft sensor model development. *IEEE Trans. Ind. Electron.* **68**(5), 4404–4414 (2021). <https://doi.org/10.1109/TIE.2020.2984443>
  136. Yuan, X., Li, L., Wang, Y.: Nonlinear dynamic soft sensor modeling with supervised long short-term memory network. *IEEE Trans. Ind. Informatics* **16**(5), 3168–3176 (2020). <https://doi.org/10.1109/TII.2019.2902129>
  137. Salles, R., Mendes, J., Araújo, R., Melo, C., Moura, P.: Prediction of key variables in wastewater treatment plants using machine learning models. In: *International Joint Conference on Neural Networks (IJCNN)*, pp. 1–9. IEEE, Padova, Italy (2022). <https://doi.org/10.1109/IJCNN55064.2022.9892661>
  138. Cai, W., Zhang, W., Hu, X., Liu, Y.: A hybrid information model based on long short-term memory network for tool condition monitoring. *J. Intell. Manuf.* **31**(6), 1497–1510 (2020). <https://doi.org/10.1007/s10845-019-01526-4>
  139. Cheng, Y., Zhu, H., Wu, J., Shao, X.: Machine health monitoring using adaptive kernel spectral clustering and deep long short-term memory recurrent neural networks. *IEEE Trans. Ind. Informatics* **15**(2), 987–997 (2019). <https://doi.org/10.1109/TII.2018.2866549>
  140. Li, S., Fang, H., Shi, B.: Multi-step-ahead prediction with long short term memory networks and support vector regression. In: *2018 37th Chinese Control Conference (CCC)*, pp. 8104–8109 (2018). <https://doi.org/10.23919/ChiCC.2018.8484066>
  141. Cheng, Y., Wang, C., Wu, J., Zhu, H., Lee, C.: Multi-dimensional recurrent neural network for remaining useful life prediction under variable operating conditions and multiple fault modes. *Appl. Soft Comput.* **118**, 108507 (2022). <https://doi.org/10.1016/j.asoc.2022.108507>
  142. Guo, J., Lao, Z., Hou, M., Li, C., Zhang, S.: Mechanical fault time series prediction by using EFMSAE-LSTM neural network. *Measurement* **173**, 108566 (2021). <https://doi.org/10.1016/j.measurement.2020.108566>
  143. Moradzadeh, A., Zakeri, S., Shoaran, M., Mohammadi-Ivatloo, B., Mohammadi, F.: Short-term load forecasting of microgrid via hybrid support vector regression and long short-term memory algorithms. *Sustainability* (2020). <https://doi.org/10.3390/su12177076>
  144. Narayan, A., Hipel, K. W.: Long short term memory networks for short-term electric load forecasting. In: *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 2573–2578 (2017). <https://doi.org/10.1109/SMC.2017.8123012>
  145. Gers, F.: Long short-term memory in recurrent neural networks. Ph.D. thesis, École Polytechnique Fédérale de Lausanne (2001). <https://doi.org/10.5075/epfl-thesis-2366>
  146. Jaeger, H.: The “echo state” approach to analysing and training recurrent neural networks-with an erratum note. *Tech. Rep.* 34, Bonn, Germany: German National Research Center for Information Technology GMD Technical Report (2001)
  147. Maass, W., Natschläger, T., Markram, H.: Real-time computing without stable states: a new framework for neural computation based on perturbations. *Neural Comput.* **14**(11), 2531–2560 (2002). <https://doi.org/10.1162/089976602760407955>
  148. Schrauwen, B., Verstraeten, D., Campenhout, J. V. (ed.): An overview of reservoir computing: theory, applications and implementations. (ed.) In: *Proceedings of the 15th European Symposium on Artificial Neural Networks*, pp. 471–482 (2007)
  149. Lukoševičius, M., Jaeger, H.: Reservoir computing approaches to recurrent neural network training. *Comput. Sci. Rev.* **3**(3), 127–149 (2009). <https://doi.org/10.1016/j.cosrev.2009.03.005>
  150. Jaeger, H.: Echo state network. *Scholarpedia* **2**(9), 2330 (2007). <https://doi.org/10.4249/scholarpedia.2330>
  151. Tang, Y., Deng, J., Zang, C., Wu, Q.: Chaotic modeling of stream nitrate concentration and transportation via IFPA-ESN and turning point analyses. *Front. Environ. Sci.* (2022). <https://doi.org/10.3389/fenvs.2022.855694>
  152. Na, X., Ren, W., Xu, X.: Hierarchical delay-memory echo state network: a model designed for multi-step chaotic time series prediction. *Eng. Appl. Artif. Intell.* **102**, 104229 (2021). <https://doi.org/10.1016/j.engappai.2021.104229>
  153. Yao, X., Wang, Z., Zhang, H.: Prediction and identification of discrete-time dynamic nonlinear systems based on adaptive echo state network. *Neural Netw.* **113**, 11–19 (2019). <https://doi.org/10.1016/j.neunet.2019.01.003>
  154. Xu, M., Han, M., Qiu, T., Lin, H.: Hybrid regularized echo state network for multivariate chaotic time series prediction. *IEEE Trans. Cybern.* **49**(6), 2305–2315 (2019). <https://doi.org/10.1109/TCYB.2018.2825253>
  155. McDermott, P.L., Wikle, C.K.: Deep echo state networks with uncertainty quantification for spatio-temporal forecasting. *Environmetrics* (2018). <https://doi.org/10.1002/env.2553>
  156. Lian, L.: Wind speed prediction based on CEEMD-SE and multiple echo state network with Gauss-Markov fusion. *Rev. Sci. Instrum.* **93**(1), 015105 (2022). <https://doi.org/10.1063/5.0081086>
  157. Bai, Y., Liu, M.-D., Ding, L., Ma, Y.-J.: Double-layer staged training echo-state networks for wind speed prediction using variational mode decomposition. *Appl. Energy* **301**, 117461 (2021). <https://doi.org/10.1016/j.apenergy.2021.117461>
  158. Hu, H., Wang, L., Tao, R.: Wind speed forecasting based on variational mode decomposition and improved echo state network. *Renew. Energy* **164**, 729–751 (2021). <https://doi.org/10.1016/j.renene.2020.09.109>
  159. Tian, Z.: Approach for short-term wind power prediction via kernel principal component analysis and echo state network optimized by improved particle swarm optimization algorithm. *Trans. Institute Measurement Control* **43**(16), 3647–3662 (2021). <https://doi.org/10.1177/01423312211046421>
  160. López, E., Valle, C., Allende, H., Gil, E., Madsen, H.: Wind power forecasting based on echo state networks and long short-term memory. *Energies* (2018). <https://doi.org/10.3390/en11030526>
  161. Zhang, H., et al.: Self-organizing deep belief modular echo state network for time series prediction. *Knowl.-Based Syst.* **222**, 107007 (2021). <https://doi.org/10.1016/j.knosys.2021.107007>
  162. Li, Q., Wu, Z., Zhang, H.: Spatio-temporal modeling with enhanced flexibility and robustness of solar irradiance prediction: a chain-structure echo state network approach. *J. Clean. Prod.* **261**, 121151 (2020). <https://doi.org/10.1016/j.jclepro.2020.121151>
  163. Xu, M., Han, M., Lin, H.: Wavelet-denoising multiple echo state networks for multivariate time series prediction. *Information*

- Sci. **465**, 439–458 (2018). <https://doi.org/10.1016/j.ins.2018.07.015>
164. Sun, X., Li, T., Li, Q., Huang, Y., Li, Y.: Deep belief echo-state network and its application to time series prediction. *Knowl.-Based Syst.* **130**, 17–29 (2017). <https://doi.org/10.1016/j.knsys.2017.05.022>
  165. Li, Z., Tanaka, G.: Multi-reservoir echo state networks with sequence resampling for nonlinear time-series prediction. *Neurocomputing* **467**, 115–129 (2022). <https://doi.org/10.1016/j.neucom.2021.08.122>
  166. Wang, L., Su, Z., Qiao, J., Deng, F.: A pseudo-inverse decomposition-based self-organizing modular echo state network for time series prediction. *Appl. Soft Comput.* **116**, 108317 (2022). <https://doi.org/10.1016/j.asoc.2021.108317>
  167. Gao, R., Du, L., Duru, O., Yuen, K.F.: Time series forecasting based on echo state network and empirical wavelet transformation. *Appl. Soft Comput.* **102**, 107111 (2021). <https://doi.org/10.1016/j.asoc.2021.107111>
  168. Schwedersky, B.B., Flesch, R.C.C., Dangui, H.A.S.: Nonlinear MIMO system identification with echo-state networks. *J. Control Autom. Electr. Syst.* (2022). <https://doi.org/10.1007/s40313-021-00874-y>
  169. He, Y.-L., Tian, Y., Xu, Y., Zhu, Q.-X.: Novel soft sensor development using echo state network integrated with singular value decomposition: application to complex chemical processes. *Chemom. Intell. Lab. Sys.* **200**, 103981 (2020). <https://doi.org/10.1016/j.chemolab.2020.103981>
  170. Huang, R., Li, Z., Cao, B.: A soft sensor approach based on an echo state network optimized by improved genetic algorithm. *Sensors* (2020). <https://doi.org/10.3390/s20175000>
  171. Mansoor, M., Grimaccia, F., Leva, S., Mussetta, M.: Comparison of echo state network and feed-forward neural networks in electrical load forecasting for demand response programs. *Math. Comput. Simul.* **184**, 282–293 (2021). <https://doi.org/10.1016/j.matcom.2020.07.011>
  172. Yao, X., Wang, Z., Zhang, H.: A novel photovoltaic power forecasting model based on echo state network. *Neurocomputing* **325**, 182–189 (2019). <https://doi.org/10.1016/j.neucom.2018.10.022>
  173. Wen, S., et al.: Memristor-Based Echo State Network With Online Least Mean Square. *IEEE Trans. Syst. Man Cybern.: Syst.* **49**(9), 1787–1796 (2019). <https://doi.org/10.1109/TSMC.2018.2825021>
  174. Bala, A., Ismail, I., Ibrahim, R., Sait, S.M., Oliva, D.: An improved grasshopper optimization algorithm based echo state network for predicting faults in airplane engines. *IEEE Access* **8**, 159773–159789 (2020). <https://doi.org/10.1109/ACCESS.2020.3020356>
  175. Zhong, S., Xie, X., Lin, L., Wang, F.: Genetic algorithm optimized double-reservoir echo state network for multi-regime time series prediction. *Neurocomputing* **238**, 191–204 (2017). <https://doi.org/10.1016/j.neucom.2017.01.053>
  176. Racca, A., Magri, L.: Robust optimization and validation of echo state networks for learning chaotic dynamics. *Neural Netw.* **142**, 252–268 (2021). <https://doi.org/10.1016/j.neunet.2021.05.004>
  177. Gunning, D., et al.: XAI—explainable artificial intelligence. *Sci. Robotics* (2019). <https://doi.org/10.1126/scirobotics.aay7120>
  178. Huang, Y., Chen, D., Zhao, W., Mo, H.: Deep fuzzy system algorithms based on deep learning and input sharing for regression application. *Intern. J. Fuzzy Syst.* **23**(3), 727–742 (2021). <https://doi.org/10.1007/s40815-020-00998-4>
  179. Wang, L.-X., Mendel, J.: Generating fuzzy rules by learning from examples. *IEEE Trans. Syst. Man Cybern.* **22**(6), 1414–1427 (1992). <https://doi.org/10.1109/21.199466>
  180. Peng, W., Zhou, C., Li, C., Deng, X., Zhang, G.: Double-input rule modules stacked deep interval type-2 fuzzy model with application to time series forecasting. *Intern. J. Fuzzy Syst.* **23**(5), 1326–1346 (2021). <https://doi.org/10.1007/s40815-021-01087-w>
  181. Li, C., Zhou, C., Peng, W., Lv, Y., Luo, X.: Accurate prediction of short-term photovoltaic power generation via a novel double-input-rule-modules stacked deep fuzzy method. *Energy* **212**, 118700 (2020). <https://doi.org/10.1016/j.energy.2020.118700>
  182. Qasem, S.N., Mohammadzadeh, A.: A deep learned type-2 fuzzy neural network: Singular value decomposition approach. *Appl. Soft Comput.* **105**, 107244 (2021). <https://doi.org/10.1016/j.asoc.2021.107244>
  183. Luo, C., Tan, C., Wang, X., Zheng, Y.: An evolving recurrent interval type-2 intuitionistic fuzzy neural network for online learning and time series prediction. *Appl. Soft Comput.* **78**, 150–163 (2019). <https://doi.org/10.1016/j.asoc.2019.02.032>
  184. Bodyanskiy, Y., Vynokurova, O., Pliss, I., Peleshko, D., Rashkevych, Y.: Deep stacking convex neuro-fuzzy system and its on-line learning. *DepCoS-RELCOMEX 2017: Advances in Dependability Engineering of Complex Systems*, pp. 49–59 (2018). [https://doi.org/10.1007/978-3-319-59415-6\\_5](https://doi.org/10.1007/978-3-319-59415-6_5)
  185. Bodyanskiy, Y.V., Tyshchenko, O.K.: A hybrid cascade neuro-fuzzy network with pools of extended neo-fuzzy neurons and its deep learning. *Intern. J. Appl. Math. Comput. Sci.* **29**(3), 477–488 (2019). <https://doi.org/10.2478/amcs-2019-0035>
  186. Cao, Y., et al.: Deep learned recurrent type-3 fuzzy system: application for renewable energy modeling/prediction. *Energy Rep.* **7**, 8115–8127 (2021). <https://doi.org/10.1016/j.egyr.2021.07.004>
  187. Wang, G., Qiao, J.: An efficient self-organizing deep fuzzy neural network for nonlinear system modeling. *IEEE Trans. Fuzzy Syst.* (2021). <https://doi.org/10.1109/TFUZZ.2021.3077396>
  188. Wang, J., Peng, Z., Wang, X., Li, C., Wu, J.: Deep fuzzy cognitive maps for interpretable multivariate time series prediction. *IEEE Trans. Fuzzy Syst.* **29**(9), 2647–2660 (2021). <https://doi.org/10.1109/TFUZZ.2020.3005293>
  189. Park, S., Lee, S.J., Weiss, E., Motai, Y.: Intra- and inter-fractional variation prediction of lung tumors using fuzzy deep learning. *IEEE J. Trans. Eng. Health Med.* **4**, 1–12 (2016). <https://doi.org/10.1109/JTEHM.2016.2516005>
  190. Magdalena, L.: Semantic interpretability in hierarchical fuzzy systems: creating semantically decouplable hierarchies. *Information Sci.* **496**, 109–123 (2019). <https://doi.org/10.1016/j.ins.2019.05.016>
  191. Wang, G., Jia, Q.-S., Qiao, J., Bi, J., Liu, C.: A sparse deep belief network with efficient fuzzy learning framework. *Neural Netw.* **121**, 430–440 (2020). <https://doi.org/10.1016/j.neunet.2019.09.035>
  192. Wang, G., Qiao, J., Bi, J., Li, W., Zhou, M.: TL-GDBN: growing deep belief network with transfer learning. *IEEE Trans. Autom. Sci. Eng.* **16**(2), 874–885 (2019). <https://doi.org/10.1109/TASE.2018.2865663>
  193. Li, F., Qiao, J., Han, H., Yang, C.: A self-organizing cascade neural network with random weights for nonlinear system modeling. *Appl. Soft Comput.* **42**, 184–193 (2016). <https://doi.org/10.1016/j.asoc.2016.01.028>
  194. Han, H.-G., Zhang, H.-J., Qiao, J.-F.: Robust deep neural network using fuzzy denoising autoencoder. *Intern. J. Fuzzy Syst.* **22**(4), 1356–1375 (2020). <https://doi.org/10.1007/s40815-020-00845-6>
  195. George, S., Santra, A.K.: Fuzzy inspired deep belief network for the traffic flow prediction in intelligent transportation system using flow strength indicators. *Big Data* **8**(4), 291–307 (2020). <https://doi.org/10.1089/big.2019.0007>

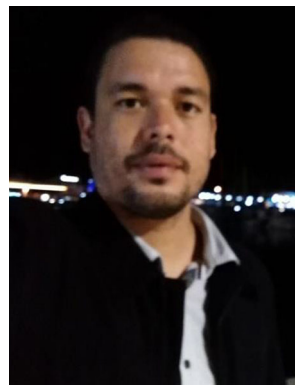
196. Borkar, G.M., Mahajan, A.R.: A secure and trust based on-demand multipath routing scheme for self-organized mobile ad-hoc networks. *Wireless Netw.* **23**(8), 2455–2472 (2016). <https://doi.org/10.1007/s11276-016-1287-y>
197. Chen, J., Yuan, W., Cao, J., Lv, H.: Traffic-flow prediction via granular computing and stacked autoencoder. *Granul. Comput.* **5**(4), 449–459 (2019). <https://doi.org/10.1007/s41066-019-00167-5>
198. Liao, C.-W., Wang, I.-C., Lin, K.-P., Lin, Y.-J.: A fuzzy seasonal long short-term memory network for wind power forecasting. *Mathematics* (2021). <https://doi.org/10.3390/math9111178>
199. Chang, P.-T.: Fuzzy seasonality forecasting. *Fuzzy Sets Syst.* **90**(1), 1–10 (1997). [https://doi.org/10.1016/S0165-0114\(96\)00138-8](https://doi.org/10.1016/S0165-0114(96)00138-8)
200. Imani, M., et al.: application of rough and fuzzy set theory for prediction of stochastic wind speed data using long short-term memory. *Atmosphere* (2021). <https://doi.org/10.3390/atmos12070924>
201. Pawlak, Z.: Rough sets. *Intern. J. Comput. Information Sci.* **11**(5), 341–356 (1982). <https://doi.org/10.1007/bf01001956>
202. Liu, L., Liu, F., Zheng, Y.: A novel ultra-short-term PV power forecasting method based on DBN-based Takagi-Sugeno fuzzy model. *Energies* (2021). <https://doi.org/10.3390/en14206447>
203. Chimatapu, R., Hagraas, H., Kern, M., Owusu, G.: Hybrid deep learning Type-2 fuzzy logic systems for explainable AI. In: 2020 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), pp. 1–6 (2020). <https://doi.org/10.1109/FUZZ48607.2020.9177817>
204. S.N., D., S., G., T.S., A.: Fuzzy echo state neural network with differential evolution framework for time series forecasting. In: 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), 1322–1327 (2018). <https://doi.org/10.1109/ICMLA.2018.00214>
205. Mahmoud, T.A., Elshenawy, L.M.: TSK fuzzy echo state neural network: a hybrid structure for black-box nonlinear systems identification. *Neural Comput. Appl.* (2022). <https://doi.org/10.1007/s00521-021-06838-2>
206. Wu, K., Liu, J., Liu, P., Yang, S.: Time series prediction using sparse autoencoder and high-order fuzzy cognitive maps. *IEEE Trans. Fuzzy Syst.* **28**(12), 3110–3121 (2020). <https://doi.org/10.1109/TFUZZ.2019.2956904>
207. Sevakula, R. K., Verma, N. K.: Fuzzy rule reduction using sparse auto-encoders. In: 2015 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), pp. 1–7 (2015). <https://doi.org/10.1109/FUZZ-IEEE.2015.7338118>
208. Ojha, V., Abraham, A., Snášel, V.: Heuristic design of fuzzy inference systems: a review of three decades of research. *Eng. Appl. Artif. Intell.* **85**, 845–864 (2019). <https://doi.org/10.1016/j.engappai.2019.08.010>



**Jorge S. S. Júnior** is a Ph.D. student in Electrical and Computer Engineering at University of Coimbra (2019-). He received the B.Sc. degree in Electrical Engineering (2014–2019) and Electrical Technician degree (2011–2013) from Federal Institute of Education, Science and Technology of Maranhão (IFMA, Imperatriz, Brazil). His current research interests include computational intelligence, fuzzy logic, deep fuzzy systems, explainable artificial intelligence, interpretable models and industrial processes. Jorge S.S. Júnior previously worked with manipulation and aerial robotics, intelligent control and multivariable modelling of complex systems.



**Jérôme Mendes** received the Ph.D. degree in electrical and computer engineering from the University of Coimbra, in 2014. He is currently a Senior Researcher at the Institute of Systems and Robotics, Coimbra. His research interests include computational intelligence with an emphasis on intelligent control (adaptive, autotuning, predictive, data-driven, evolving, and interpretable); intelligent identification (prediction, soft sensors, evolving design, and interpretable), and intelligent failure detection.



**Francisco Souza** received a Ph.D. in Electrical and Computer Engineering from the University of Coimbra in 2014. He works as a researcher at the Department of Analytical Chemistry & Chemometrics, where he works on modeling and optimization of industrial processes. Francisco's main interests are applied machine learning, industrial systems, and industry 4.0. He collaborated on research projects on industrial sustainability, process optimization, and data-driven modeling.



**Cristiano Prebida** is Assistant Professor in the department of electrical and computer engineering at the University of Coimbra, Portugal, where he is a member of the Institute of Systems and Robotics (ISR-UC). His main research interests are robotic perception, machine learning, Bayesian inference, autonomous vehicles, autonomous robots, agricultural robotics, and sensor fusion. C. Prebida has collaborated

related to autonomous driving, robotic perception, agricultural

robotics and applied machine learning. He works on multimodal and multisensory perception for robotics and autonomous systems applications, developing calibration strategies and probability-prediction approaches to increase robustness of deep models.